STOCHASTIC OPTIMAL CONTROL – A FORWARD AND BACKWARD SAMPLING APPROACH

A Thesis
Presented to
The Academic Faculty

by

Ioannis Exarchos

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Aerospace Engineering

Georgia Institute of Technology
December 2017

STOCHASTIC OPTIMAL CONTROL – A FORWARD AND BACKWARD SAMPLING APPROACH

Approved by:

Professor Panagiotis Tsiotras,
Committee Chair
School of Aerospace Engineering
Georgia Institute of Technology

Professor Wassim M. Haddad
School of Aerospace Engineering
Georgia Institute of Technology

Professor Panagiotis Tsiotras, Advisor
School of Aerospace Engineering
Georgia Institute of Technology

Professor Hao-min Zhou
School of Mathematics
Georgia Institute of Technology

Professor Evangelos A. Theodorou
School of Aerospace Engineering
Georgia Institute of Technology

Professor Ionel Popescu
School of Mathematics
Georgia Institute of Technology

Date Approved: 11/08/2017

"Life must be lived forwards, but can only be understood backwards"

– Søren Kierkegaard

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor, Dr. P. Tsiotras, for his continuous support and guidance during the many years of my Ph.D studies. It is with his encouragement that I was able not to merely specialize in a single topic, but to seek breadth of knowledge within my field of study along the way. I am also deeply grateful to my coadvisor Dr. Evangelos A. Theodorou, who introduced me to the topic of this dissertation, and sparked my interest in stochastic control and machine learning in general. Furthermore, I would like to express my sincere appreciation to Drs. W. M. Haddad, H. M. Zhou, and I. Popescu, members of my Ph.D Dissertation Committee, not only for their true interest in evaluating this dissertation, but also for generously sharing their wealth of knowledge in the classroom lectures that I attended.

The financial support I received from the A. S. Onassis Public Benefit Foundation was essential towards the unimpeded continuation of my studies, as well as the quality of my life during the past six years. For that, the Foundation has my eternal gratitude.

Finally, I would like to thank all of my friends, both in the United States and overseas, for their moral and emotional support during the difficult times of my graduate studies. Ashley, Jose, Michael, Nick, Oktay, Omid, Peter, Trevor, and the Greeks: thank you for this priceless gift; it would have been so much more difficult without you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Stochastic optimal control has seen significant recent development, motivated by its success in a plethora of engineering applications, such as autonomous systems, robotics, neuroscience, and financial engineering. Despite the many theoretical and algorithmic advancements that made such a success possible, several obstacles remain; most notable are (i) the mitigation of the curse of dimensionality inherent in optimal control problems, (ii) the design of efficient algorithms that allow for fast, online computation, and (iii) the expansion of the class of optimal control problems that can be addressed by algorithms in engineering practice.

Prior work on stochastic control theory and algorithms mitigates the complexity of the optimal control problem by sacrificing global optimality. Furthermore, several restrictive conditions are imposed, such as differentiability of the dynamics and cost functions, as well as certain assumptions involving control authority and stochasticity. Thus, state-of-the-art algorithms may only address special classes of systems. The goal of this dissertation is to establish a framework that goes beyond these limitations. The proposed stochastic control framework capitalizes on the innate relationship between certain nonlinear partial differential equations (PDEs) and forward and backward stochastic differential equations (FBSDEs), as demonstrated by a nonlinear version of the Feynman-Kac lemma. By means of this lemma, we are able to obtain a probabilistic representation of the solution to the nonlinear Hamilton-Jacobi-Bellman equation, expressed in form of a system of decoupled FBSDEs. This system of FBSDEs can then be solved numerically in lieu of the original PDE problem. We present a novel discretization scheme for FBSDEs, and enhance the resulting

algorithm with importance sampling, thereby constructing an iterative scheme that is capable of learning the optimal control without an initial guess, even in systems with highly nonlinear, underactuated dynamics.

The framework developed within this dissertation addresses several classes of stochastic optimal control, including $\mathcal{L}^2$, $\mathcal{L}^1$, game theoretic, and risk sensitive control, in both fixed-final-time as well as first-exit settings.

I

INTRODUCTION

1.1 Motivation and Previous Work

Stochastic optimal control lies within the foundation of mathematical control theory ever since its inception. Its usefulness has been proven in a plethora of engineering applications, such as autonomous systems, robotics, neuroscience, and financial engineering, among others. Specifically, in robotics and autonomous systems, stochastic control has become one of the most successful approaches for planning and learning, as demonstrated by its effectiveness in many applications, such as control of ground and aerial vehicles, articulated mechanisms and manipulators, and humanoid robots [108–110, 123, 127, 131]. In computational neuroscience and human motor control, stochastic optimal control theory is the primary framework used in the process of modeling the underlying computational principles of the neural control of movement [130, 132]. Furthermore, in financial engineering, stochastic optimal control provides the main computational and analytical framework, with widespread application in portfolio management and stock market trading [102, 121].

By and large, prior work on stochastic control theory and algorithms imposes restrictive conditions such as differentiability of the dynamics and cost functions, and furthermore requires certain assumptions involving control authority and stochasticity to be met. Thus, it may only address special classes of systems. The goal of this dissertation is to establish a framework that goes beyond these limitations. In particular, we propose a learning stochastic control framework which capitalizes on the innate relationship between certain nonlinear partial differential equations (PDEs) and forward and backward stochastic differential equations (FBSDEs), demonstrated

by a nonlinear version of the Feynman-Kac lemma. By means of this lemma, we are able to obtain a probabilistic representation of the solution to the nonlinear Hamilton-Jacobi-Bellman equation, expressed in form of a system of decoupled FBSDEs. This system of FBSDEs can then be simulated by employing linear regression techniques. We present a novel discretization scheme for FBSDEs, and enhance the resulting algorithm with importance samping, thereby constructing an iterative scheme that is capable of learning the optimal control without an initial guess, even in systems with highly nonlinear, underactuated dynamics. In addition, the proposed approach exhibits the following characteristics:

- Perform stochastic control and trajectory optimization without linearization of the dynamics and quadratic approximations of the cost functions.
- Find nonlinear feedback control policies that yield higher performance than their traditional trajectory optimization counterparts.
- Be based on sampling, scalable, and therefore directly applicable to high dimensional systems, and able to accommodate parallel computation.
- Expand the class of systems currently addressed by traditional stochastic optimal control methods.

The framework developed within this dissertation addresses several classes of stochastic optimal control, including $\mathcal{L}^2$, $\mathcal{L}^1$, game theoretic, and risk sensitive control, in both fixed-final-time and first-exit settings. In what follows, we review each of the aforementioned categories separately.

1.1.1 Stochastic $\mathcal{L}^2$-Optimal Control

The literature on stochastic optimal control has experienced a significant increase in attention during the last years. In most cases, the problem of obtaining an optimal control is associated with the solution of a generally nonlinear, second-order (in the

case of stochstic control) PDE, known as the Hamilton-Jacobi-Bellman (HJB) equation. A classification of different available methods can be achieved based on whether the solution of this PDE is sought for over the entire domain, or locally around a nominal system trajectory. In the first case, several attempts have been made to address the difficulty inherent in solving such nonlinear PDEs, as well as the curse of dimensionality, with various different methods and approaches. Such approaches include the use of the Galerkin method [7], level set methods [91], max-plus expansion of solutions [86], high-order Taylor series expansions [2], or semidefinite programming [72] for deterministic optimal control problems, while a stochastic setting is considered in [48, 53, 54]. With only but a few exceptions, most of these methods suffer from the curse of dimensionality. On the other hand, the latter category of local methods includes traditional approaches such as Stochastic Differential Dynamic Programing (S-DDP) [128, 133], which is based on linearization of the dynamics and a quadratic approximation of the value function around nominal trajectories, as well as sampling-based methods.

Sampling-based methods, within stochastic control, rely on a probabilistic representation of the solution to linear backward PDEs. This probabilistic representation is addressed by forward sampling of state trajectories via Stochastic Differential Equations (SDEs), and the numerical evaluation of expectations. Several results based on this framework appear in the literature under the names of Path Integral (PI) Control [58–60, 126, 128], Kullback-Leibler (KL) Control, or Linearly Solvable Optimal Control (LSOC) [34, 131]. These methods have become an exceedingly popular approach to solve nonlinear stochastic optimal control problems due to their ability to accommodate scalable iterative schemes. Their fundamental characteristic is that they rely on the exponential transformation of the value function. Under the exponential transformation, and by introducing certain restrictions between control authority, cost and stochasticity, there exists a direct relationship between the HJB PDE and the

backward Chapman-Kolmogorov PDE. The latter PDE, being linear, permits then the use of the linear Feynman-Kac lemma [61], which relates backward linear PDEs to forward SDEs. Thus, the corresponding optimal control problem can be solved using forward sampling. This approach has interesting implications, suggesting an information theoretic interpretation of stochastic optimal control, as well as further connections to the Legendre transformation in statistical mechanics [126, 129]. While forward sampling-based methods exhibit several advantages against traditional methods of stochastic control, such as the mild conditions on the differentiability of the cost and the stochastic dynamics, there are also some key disadvantages which pertain to the nature of the exponential transformation. In particular, the effect of the exponential transformation can be identified as the mapping of the value function $v(t, x)$, which has range $[0, \infty)$, to the desirability function $\psi(t, x)$, whose range is $(0, 1]$. This mapping leads to a drastic reduction in the ability to distinguish states with high cost (low desirability) from states with low cost (high desirability). This issue has been partially addressed with renormalization of the trajectory cost [128]. Finally, while the necessary constraint introduced between control authority and stochasticity can lead to symmetry breaking phenomena and delayed decision [58, 59], it is a rather restrictive assumption whenever applications to engineered systems are considered.

1.1.2 Stochastic $\mathcal{L}^1$-Optimal Control

By and large, the literature on optimal control deals with the minimization of a performance index which penalizes control *energy*, since the input appears in quadratic form as part of the running cost. Such problems are typically referred to as *minimum energy* problems in optimal control theory– they involve the minimization of the $\mathcal{L}^2$-norm of an otherwise unconstrained control signal. 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 etc.), 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 [29, 117], in which fuel is a limited resource. Indeed, in such applications, using the $\mathcal{L}^2$-norm results in significantly more propellant consumption as well as undesirable continuous thrusting. In some illustrative examples, this fuel penalty can be as high as 50% [111].

The notion of $\mathcal{L}^1$-optimal control is also tightly related to *Maximum Hands-Off control* [97, 98]. The distinguishing characteristic of a hands-off controller is its objective to retain a zero control input value over an extended time interval. In other words, the goal of "maximum hands-off" control is to accomplish a specific task while applying zero input for the longest time duration possible. Applications of this type of control range from the automotive industry (engine stop-start systems [33], hybrid vehicles [18]) to networked and embedded systems [57, 68]. The "hands-off" property is especially in a discrete context equivalent to *sparsity* of a signal, i.e., minimizing the total length of intervals over which the signal takes non-zero values. The relationship between $\mathcal{L}^1$-optimality and the "hands-off" property, or sparsity, is shown in [97, 98]. Specifically, if an $\mathcal{L}^1$-optimal control problem is *normal* (see [4], as well as Remark 6.1 in Chapter 6), then its optimal solution is also the optimal sparse, "hands-off" solution.

Despite the aforementioned advantages, investigation of $\mathcal{L}^1$-optimal control in the literature is not as widespread as $\mathcal{L}^2$, since it leads to significantly more complicated optimal control structures. These structures are usually a combination of *bang-off-bang* control (i.e, the control signal switches between its extrema and zero)

and *singular* control, in which the control input receives intermediate values. Moreover, the particular structure often depends on the specific initial condition or other parameter values, and neither existence, nor uniqueness of solutions, can always be guaranteed [4]. All these subtleties complicate the process of finding a solution, which partially explains the scarcity of $\mathcal{L}^1$-minimization results in the literature.

1.1.3 Differential Games and Risk Sensitive Control

The origin of game-theoretic control dates back to the work of Isaacs (1965) [55] on differential games for two strictly competitive players, which provided a framework for the treatment of such problems. Isaacs 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 by Isaacs under the assumptions of Lipschitz continuity of the dynamics and the payoff, as well as the assumption that both of them are *separable* in terms of the minimizing and maximizing controls.

Berkovitz [11] addressed differential games using standard variational techniques, a framework which was later adopted by Bryson, Ho, and Baron [52] to treat a special case of differential games, namely, games of pursuit and evasion. Pontryagin also addressed pursuit and evasion problems within the framework of differential games [104]. A treatment of the stochastic extension to differential games was first provided in [70]. Therein, the authors 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 [17, 24, 40, 51].

Despite the plethora of theoretical work in the area of differential games, the algorithmic part has received significantly less attention, due to the inherent difficulty of solving such problems. Apart from results addressing special cases of differential games (such as linear games with quadratic penalties, e.g. [32]), only a few numerical approaches have been suggested in the past, notably the Markov Chain approximation method [69, 120]; in general, however, these numerical procedures have found only limited applicability due to the "curse of dimensionality." Only recently, a specific class of minimax control trajectory optimization methods have been derived, all based on the foundations of *differential dynamic programming* (DDP) [92, 93, 124].

Game-theoretic or min-max extensions to optimal control are known to have a direct connection to robust and H^∞ nonlinear control theory, as well as to risk-sensitive optimal control [5, 25, 56]. The relationship to the latter category was first investigated by Jacobson in [56]. References [10,135] and [41] investigate risk-sensitive stochastic control in an LQG setting, and for nonlinear stochastic systems and infinite horizon control tasks, respectively. Ever since the fundamental work of [41, 56, 135], the topic of risk sensitivity has been studied extensively. In a risk-sensitive setting, the control objective is to minimize a performance index, which is expressed as a function of the mean and variance of a given state- and control-dependent cost. Therefore, the element of risk sensitivity arises from the minimization of the variance of that cost. An application of the Dynamic Programming principle on the risk-sensitive optimization criterion results in a backward PDE that is similar to the HJI PDE in which players pay an $\mathcal{L}^2$-type penalty for their control effort. Thus, risk-sensitive optimal control problems exhibit the same structure as that of a class of stochastic differential games [5].

1.2 Statement of Contributions

In this dissertation, we aim to develop a sampling-based control algorithm which capitalizes on the innate relationship between certain nonlinear PDEs and Forward and Backward SDEs, demonstrated by a *nonlinear* Feynman-Kac lemma. By means of this lemma, we obtain a probabilistic representation of the solution to the nonlinear HJB equation, expressed in the form of a system of decoupled FBSDEs. This system of FBSDEs can be solved by employing linear regression techniques. To enhance the efficiency of the proposed scheme when treating more complex nonlinear systems, we then derive an iterative algorithm based on Girsanov's theorem on the change of measure, which features importance sampling for the case of FBSDEs. The framework is capable of addressing several types of stochastic optimal control problems, such as $\mathcal{L}^2$, $\mathcal{L}^1$, risk-sensitive control, and differential games, considering both fixed final time and first exit settings. The contributions in this dissertation vis-à-vis prior work in the literature are as follows:

- With respect to the state-of-the-art on sampling-based methods for stochastic $\mathcal{L}^2$-optimal control: There is a significant difference between the proposed approach and the already existing sampling-based formulations (such as PI, KL, and LSOC). Specifically, our approach addresses directly the nonlinear PDE, while the latter make use of the exponential transformation, which under certain conditions yields a linear PDE problem, and then use forward sampling to address that linear problem. Thus, the herein proposed framework relaxes these restrictive conditions. Furthermore, while traditional sampling-based methods yield a solution only for the initial condition point (t, x) and must be applied in a receding horizon fashion, the solution obtained through the proposed method extends from the initial condition (t, x) to the terminal time T, covering the sampled state space area.

- With respect to stochastic $\mathcal{L}^1$-optimal control: It is shown that $\mathcal{L}^1$-optimal control problems of the form considered within this research correspond to a particular FBSDE problem, in light of the nonlinear Feynman-Kac lemma, which can then be solved in lieu of the original PDE problem. This work is the first to address $\mathcal{L}^1$-optimal problems in this context, and, to the best of our knowledge, the first to investigate stochastic $\mathcal{L}^1$-optimal control problems in continuous time.

- The class of problems addressed by the proposed framework is extended to treat cases in which the time horizon is not prespecified, as well as differential games.

- With respect to prior work on the nonlinear Feynman-Kac lemma, its applications to stochastic optimal control, and FBSDEs: The majority of prior work in this case appears mainly within the field of mathematical finance [19], and is typically limited to numerical schemes which are not scalable. Although some prior work exists addressing more complex generalizations of the class of problems we consider in this dissertation (see, for example, [22, 62, 63]), the results obtained therein have extremely limited practical applicability when engineering systems are concerned. This is because the preexisting numerical schemes are investigated with a focus on their theoretical properties, rather than their suitability for engineering applications. The fact that these schemes are unable to cope with the complexity of higher dimensional systems featuring nonlinear dynamics, as it is mostly the case in practical applications, is largely overlooked. As a result, most of the existing work is not accompanied by simulations, except for cases limited to simple, and mostly scalar, linear systems. In contrast, the applicability of the framework proposed in this dissertation is demonstrated on a four dimensional, highly nonlinear, unstable, underactuated system. This would be practically infeasible without three key elements, proposed herein for

the first time on FBSDEs and thus defining the novelty in this approach:

i) Restricting the class of problems to systems affine in controls with quadratic (or $\mathcal{L}^1$) control penalty. This eliminates the need to perform numerical optimization over the control input at each time step, and allows us to compute optimal policies by using sampling only.

ii) A modified FBSDE discretization scheme featuring only one regression per time step (instead of the $p+1$ per time step, where p is the dimensionality of noise, performed in state-of-the-art discretizations), and is shown to outperform the most established existing scheme in simulation accuracy in control applications.

iii) Most importantly: the iterative scheme utilizing Girsanov's theorem on the change of measure for FBSDEs. This step is absolutely critical if one wishes to apply any FBSDE algorithm for control of more complex systems, as it is practically infeasible to do so without importance sampling.

1.3 Outline

The remainder of this dissertation consists of the following chapters, the content of which is described as follows:

- Chapter 2 introduces notation and definitions used throughout this dissertation. Furthermore, it presents a brief review of the relevant theoretic background concerning probability theory, and forward and backward stochastic differential equations; in particular, definitions of the forward and backward processes, theorems concerning existence and uniqueness of solutions to systems of FBSDEs, the Markovian property of FBSDEs, and their connection to certain PDEs via a nonlinear Feynman-Kac type formula.

- In Chapter 3 we define the $\mathcal{L}^2$- type formulation of the stochastic optimal control problem. This specific class of stochastic optimal control allows for an explicit minimization of the Hamiltonian term within the Hamilton-Jacobi-Bellman (HJB) equation, thus simplifying the structure of the problem. We demonstrate that under a certain decomposability condition, the HJB equation lies within the class of PDEs that allow a probabilistic expression of their solution, in light of the nonlinear Feynman-Kac lemma, through FBSDEs. Thus, we can obtain the solution to the HJB equation by solving the associated system of FBSDEs.

- Chapter 4 is devoted to the investigation of numerical methods for the class of FBSDEs involved in this dissertation. In general, the procedure of obtaining a numerical solution for a system of FBSDEs consists of three elements: (i) a time discretization scheme for the forward process, (ii) a time discretization scheme for the backward process, and (iii) a numerical approximation scheme for the conditional expectation evaluation in each time step of the backward process. We provide a brief overview of the literature, introducing some of the most thoroughly studied time discretization and conditional expectation approximation schemes. We then propose a novel and efficient numerical scheme, suitable for the particular type of FBSDE systems considered in this dissertation, that greatly reduces the computational complexity in obtaining a solution, while exhibiting higher accuracy in simulations.

- Chapter 5 investigates the construction of an iterative scheme capable of addressing control problems that exhibit more complex, nonlinear dynamics. Specifically, we solve the optimal control problem iteratively by suitably modifying the drift of the forward process, thus directing the exploration of the state space

towards the given goal state, or any other state of interest, reachable by control. Furthermore, we discuss the scheme's convergence and error sources, and demonstrate its effectiveness in simulation.

- In Chapter 6 we turn our attention to stochastic $\mathcal{L}^1$-optimal control problems. We begin with a definition of the $\mathcal{L}^1$-type formulation, and show that this specific class of stochastic optimal control also allows, in a manner similar to its $\mathcal{L}^2$ counterpart, for an explicit minimization of the Hamiltonian term within the HJB equation. We then demonstrate that under the same decomposability condition as in Chapter 3, the HJB equation lies within the class of PDEs that allow a probabilistic expression of their solution via the nonlinear Feynman-Kac lemma. Thus, we can obtain the solution to the HJB equation by solving the associated system of FBSDEs. The chapter is concluded with simulations on different $\mathcal{L}^1$-optimal control problems.

- In Chapter 7, we demonstrate that framework developed in this dissertation can be employed in the solution of a variety of classes of stochastic differential game problems. Specifically, we show that the Hamilton-Jacobi-Isaacs PDEs, corresponding to $\mathcal{L}^2$ or $\mathcal{L}^1$ penalties for the players, assume simplified expressions under affine dynamics. Furthermore, an extension of the decomposability condition of Chapter 3 is enough to allow for a probabilistic representation of the solutions to these HJI PDEs via FBSDEs. Finally, we note that since the simplified HJI PDE that appears for the $\mathcal{L}^2$-case of stochastic differential games exhibits the same form as the HJB PDE of a risk-sensitive optimal control problem, the herein proposed scheme is applicable to this type of stochastic optimal control as well. The chapter is concluded with simulations.

- Chapter 8 is devoted to the extension the framework presented in this dissertation to address stochastic optimal control problems in which we do not specify

a priori a fixed time of termination, but rather, termination occurs when a particular state (or set of states) is reached. In the context of differential games, the boundary of such a set of states is called a *terminal surface*. Simulations illustrate the main idea in this chapter.

- In Chapter 9, we apply the proposed algorithm on a stochastic, first-exit, $\mathcal{L}^1$-optimal control problem, namely the *soft landing* problem in minimum-fuel powered descent guidance. The objective is to successfully land a spacecraft on a planet using the least amount of fuel, while concurrently ensuring that the landing speed is as low as possible, in order to minimize the risk of a harmful impact. The deterministic version of the problem allows for a simplified expression for the control input in terms of the switching time. Thus, we can compare the performance of the deterministic control law, applied both in an open loop and closed loop fashion, to that of the feedback control law obtained from the proposed framework, in the presence of a stochastic environment.

- Finally, Chapter 10 summarizes the key contributions of this dissertation and outlines future directions of research.

II

BACKGROUND AND PRELIMINARIES

In this chapter, we introduce notation and definitions used throughout this dissertation. We also review the relevant theoretic background concerning probability theory, stochastic calculus, and forward and backward stochastic differential equations. Specifically, we define the forward and backward processes, and review the theorems concerning existence and uniqueness of solutions to systems of FBSDEs, the Markovian property of FBSDEs, and their connection to certain PDEs via a nonlinear Feynman-Kac type formula.

2.1 Notation and Acronyms

The following list summarizes notation and acronyms used in this dissertation.

$\mathbb{R}$	the set of reals
$\mathbb{R}_+$	the set of nonnegative reals
$\mathbb{R}^n$	n-dimensional Euclidean space
$\mathbb{R}^{n\times m}$	all $n \times m$ real-valued matrices
I_n	the $n \times n$ identity matrix
$A^\top$	the transpose of a matrix A
$\mathrm{tr}A$	the trace of a matrix A
$\mathrm{sgn}(x)$	$\begin{cases} 1, \text{ if } x \geq 0 \\ -1, \text{ if } x < 0 \end{cases}$
C^k	the space of functions with continuous derivatives up to order k

$C^{1,2}(\mathbb{R} \times \mathbb{R}^n)$	the space of functions $f(t,x) : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ which are C^1 w.r.t. $t \in \mathbb{R}$ and C^2 w.r.t. $x \in \mathbb{R}^n$.
v_x, v_{xx}	the gradient and Hessian, respectively, of a function v
$\triangleq$	defined as
$\equiv$	identically equal to
$\approx$	approximately equal to
$\mathscr{F}_t$	filtration at time t
$\cup$, $\cap$	union, intersection
$\mathbb{E}[\cdot]$	mathematical expectation
$\mathbb{E}[\cdot\|\mathscr{F}_t]$	expectation conditioned on $\mathscr{F}_t$
$\mathcal{L}^p([0,T];X)$	the space of measurable functions $f : [0,T] \to X$ such that $\mathbb{E}[\int_0^T \|\|f(t)\|\|^p \mathrm{d}t] < \infty$, for $1 \leq p < \infty$.
$\mathcal{N}(\mu, \sigma^2)$	Gaussian (normal) distribution with mean μ and variance σ^2
W_t	a standard Brownian motion process
FSDE	the forward stochastic differential equation (forward process)
BSDE	the backward stochastic differential equation (backward process)
FBSDE	a system of forward and backward stochastic differential equations
HJB	the Hamilton-Jacobi-Bellman equation
HJI	the Hamilton-Jacobi-Isaacs equation
SLP	the Soft Landing Problem

2.2 Probability and Stochastic Processes

This section summarizes the most important mathematical concepts used throughout this dissertation. More details on these concepts can be found in references [61, 71, 100].

Definition 2.1. *(σ-algebra, Measurable Space):* Let Ω be a set. A *σ-algebra* $\mathscr{F}$ on Ω is a collection of subsets of Ω that contains the empty set, the set Ω itself, and is closed under complement and countable union of its members. The pair $(\Omega, \mathscr{F})$ is called a *measurable space*.

A probability space is a measurable space equipped with a probability measure:

Definition 2.2. *(Probability Measure, Probability Space):* Let $\{A_i\}_{i=1}^{\infty} \subset \mathscr{F}$ be any collection of events. A *probability measure* $\mathbb{P}$ on a measurable space $(\Omega, \mathscr{F})$ is a function $\mathbb{P} : \mathscr{F} \to [0, 1]$ satisfying the following axioms: (i). $\mathbb{P}(\emptyset) = 0$, (ii). $\mathbb{P}(\Omega) = 1$ and (iii). if the events A_i are disjoint (i.e. $A_i \cap A_j = \emptyset$ for $i \neq j$) then the probability of their union equals to the sum of their probabilities. The triple $(\Omega, \mathscr{F}, \mathbb{P})$ is called a *probability space*. Furthermore, $(\Omega, \mathscr{F}, \mathbb{P})$ is called a *complete probability space* if $\mathscr{F}$ contains all subsets G of Ω with $\mathbb{P}$- outer measure zero.

If $(\Omega, \mathscr{F}, \mathbb{P})$ is a given probability space, then a function $X : \Omega \to \mathbb{R}^n$ is called *$\mathscr{F}$-measurable* if its pre-image belongs to $\mathscr{F}$, i.e., $X^{-1}(U) \triangleq \{\omega \in \Omega : X(\omega) \in U\} \in \mathscr{F}$.

Definition 2.3. *(Random Variable):* Given $(\Omega, \mathscr{F}, \mathbb{P})$, a *random variable* X is an $\mathscr{F}$-measurable function $X : \Omega \to \mathbb{R}^n$.

Every random variable induces a probability measure (or *distribution*) μ_X on $\mathbb{R}^n$, defined by $\mu_X(B) = \mathbb{P}(X^{-1}(B))$. The *mathematical expectation* of X is then defined as

$$\mathbb{E}[X] \triangleq \int_{\Omega} X(\omega) \mathrm{d}\mathbb{P}(\omega) = \int_{\mathbb{R}^n} x \mathrm{d}\mu_X(x).$$

More generally, if $f : \mathbb{R}^n \to \mathbb{R}$ is measurable, we define

$$\mathbb{E}[f(X)] \triangleq \int_\Omega f(X(\omega))\mathrm{d}\mathbb{P}(\omega) = \int_{\mathbb{R}^n} f(x)\mathrm{d}\mu_X(x).$$

Two subsets $A, B \in \mathscr{F}$ are called *independent* if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. If two random variables $X, Y : \Omega \to \mathbb{R}$ are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. The concept of *conditional expectation* will also be used extensively in this dissertation, and is defined as follows:

Definition 2.4. *(Conditional Expectation):* Given $(\Omega, \mathscr{F}, \mathbb{P})$, a random variable $X : \Omega \to \mathbb{R}^n$ such that $\mathbb{E}[|X|] < \infty$, and $\mathcal{H} \subset \mathscr{F}$ a σ-algebra, then the *conditional expectation* of X given $\mathcal{H}$, denoted $\mathbb{E}[X|\mathcal{H}]$ is a function from Ω to $\mathbb{R}^n$ such that (i). $\mathbb{E}[X|\mathcal{H}]$ is $\mathcal{H}$-measurable and (ii). $\int_H \mathbb{E}[X|\mathcal{H}]\mathrm{d}\mathbb{P} = \int_H X\mathrm{d}\mathbb{P}$ for all $H \in \mathcal{H}$.

We now proceed to the definition of stochastic processes:

Definition 2.5. *(Stochastic Process):* A *stochastic process* is a parameterized collection of random variables $\{X_t\}_{t\in\mathcal{T}}$, defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, and assuming values in $\mathbb{R}^n$.

The parameter space $\mathcal{T}$ is usually the semi-infinite interval $[0, \infty)$. In this dissertation, however, it will usually consist of the compact set $[0, T]$ for some constant $T > 0$. For each fixed t, we have a random variable $\omega \to X_t(\omega)$, $\omega \in \Omega$, while on the other hand, fixing a certain ω we can consider the function $t \to X_t(\omega)$, $t \in \mathcal{T}$, which is called a *sample path* or *realization*. In this sense, t can be seen as "time" and each ω can be seen as a "particle", or "experiment". Note that the notation $X_t(\omega)$, X_t, or $X(t, \omega)$ are used interchangeably. An important class of stochastic processes are the *square-integrable* processes, defined as follows:

Definition 2.6. *(Square-integrability):* A stochastic process X_t is called *square-integrable* if $\mathbb{E}\big[\int_\tau^T |X_t|^2\mathrm{d}t\big] < \infty$ for any $T > \tau$.

Definition 2.7. *(Filtration, Adapted Process):* A *filtration* on $(\Omega, \mathscr{F}, \mathbb{P})$ is a family $\{\mathscr{F}_t\}_{t\geq 0}$ of σ-algebras $\mathscr{F}_t$ such that $\mathscr{F}_s \subset \mathscr{F}_t$ whenever $0 \leq s < t$, i.e, $\{\mathscr{F}_t\}$ is increasing. Then, a process $\{X_t\}_{t\geq 0}$ is called $\mathscr{F}_t$*-adapted* if for each $t \geq 0$ the function $\omega \to X(t, \omega)$ is $\mathscr{F}_t$-measurable.

Note that the terms *adapted* and *progressively measurable* are sometimes used interchangeably as well. For a stochastic process, the notation $(\Omega, \mathscr{F}, \mathbb{P})$ is substituted by $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P})$, or simply $(\Omega, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P})$. A large class of stochastic processes are the so-called *martingales*:

Definition 2.8. *(Martingale):* An n-dimensional stochastic process $\{M_t\}_{t\geq 0}$ on $(\Omega, \mathscr{F}, \mathbb{P})$ is called a *martingale* with respect to a filtration $\{\mathscr{F}_t\}_{t\geq 0}$ and probability measure $\mathbb{P}$ if (i). M_t is $\mathscr{F}_t$-measurable for all t, (ii). $\mathbb{E}[|M_t|] < \infty$ for all t and (iii). $\mathbb{E}[M_s|M_t] = M_t$ for any $s \geq t$.

Perhaps the most famous special case of a martingale is the *Brownian motion*, also known as the *Wiener process*:

Definition 2.9. *(Standard Brownian Motion):* A *standard Brownian motion* process (or *Wiener process*) is a family $\{W_t\}_{t\geq 0}$ of real-valued random variables defined on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$ that satisfies:

- $W_0 = 0$ almost surely.
- If $0 = t_0 < t_1 < t_2 < \ldots < t_n$, then the random variables $W_{t_k} - W_{t_{k-1}}$ for $k = 1, 2, \ldots, n$ are independent (i.e., W_t has independent increments).
- For each $s, t \geq 0$, the random variable $W_{t+s} - W_t$ is normally distributed with mean zero and variance s.
- For almost all $\omega \in \Omega$, the function $W_t = W_t(\omega)$ is everywhere continuous in t.

The above definition covers the one-dimensional process, but can be extended to multiple dimensions in a straightforward manner; for a p-dimensional standard Brownian motion process, one has to merely stack p *independent* standard one-dimensional Brownian motion processes in a p-dimensional vector. The differential $\mathrm{d}W_t$ of a standard Brownian motion stems from the limit $\mathrm{d}W_t = \lim_{\Delta t \to \mathrm{d}t}(W_{t+\Delta t} - W_t)$, and thus in light of the definition of Brownian motion, we establish that $\mathrm{d}W_t \sim \mathcal{N}(0, \mathrm{d}t)$. An immediate consequence is the following important property:

$$\mathbb{E}[(\mathrm{d}W_t)^2] = \mathrm{d}t.$$

Notice that the ratio $\mathrm{d}W_t/\mathrm{d}t$ follows the distribution $\mathcal{N}(0, 1/\mathrm{d}t)$, and therefore has infinite variance as $\mathrm{d}t \to 0$. In engineering, the process $v(t) = \mathrm{d}W_t/\mathrm{d}t$ is referred to as *white noise*. Based on the Brownian motion differential, we proceed to the definition of the *Itô integral*.

Definition 2.10. *(Itô Integral):* Let W_t be a standard Brownian motion and let F_t be any measurable, square-integrable, $\mathscr{F}_t$-adapted process. The *Itô integral* of F_t against W_t up to time t is a stochastic process G_t denoted by

$$G_t = \int_0^t F_\tau \mathrm{d}W_\tau.$$

The construction of the above integral is formally established using simple functions, see [61,100] for details. Note that G_t is in fact a martingale, and its expectation is equal to zero. We next define the concept of *absolute continuity* of measures.

Definition 2.11. *(Absolute Continuity):* Let $(\Omega, \mathscr{F}, \mathscr{F}_{t t \geq 0}, \mathbb{P})$ be a complete filtered probability space, fix some $T > 0$, and let $\mathbb{Q}$ be another probability measure on $\mathscr{F}_T$. Then $\mathbb{Q}$ is *absolutely continuous* with respect to $\mathbb{P}|_{\mathscr{F}_T}$ (the restriction of $\mathbb{P}$ to $\mathscr{F}_T$) if $\mathbb{P}(H) = 0$ implies $\mathbb{Q}(H) = 0$ for all $H \in \mathscr{F}_T$.

The above condition occurs if and only if there exist an $\mathscr{F}_T$-measurable random variable $M_T(\omega) \geq 0$ such that $\mathrm{d}\mathbb{Q}(\omega) = M_T(\omega)\mathrm{d}\mathbb{P}(\omega)$ on $\mathscr{F}_T$, in which case we may write

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = M_T, \qquad \text{on } \mathscr{F}_T.$$

The above ratio is called the *Radon-Nikodym derivative*. The following lemma demonstrates that the restrictions of absolutely continuous measures are also absolutely continuous, and the process of Radon-Nikodym derivatives is a martingale:

Lemma 2.1. *(Process of Radon-Nikodym Derivatives): Suppose that $\mathbb{Q}$ is absolutely continuous with respect to $\mathbb{P}|_{\mathscr{F}_T}$, with $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = M_T$ on $\mathscr{F}_T$. Then the restrictions $\mathbb{Q}|_{\mathscr{F}_t}$ and $\mathbb{P}|_{\mathscr{F}_t}$ are also absolutely continuous for all $t \in [0,T]$, and the process of Radon-Nikodym derivatives defined as*

$$M_t \triangleq \frac{\mathrm{d}(\mathbb{Q}|_{\mathscr{F}_t})}{\mathrm{d}(\mathbb{P}|_{\mathscr{F}_t})}, \qquad t \in [0,T],$$

is a martingale with respect to $\mathscr{F}_t$ and $\mathbb{P}$.

This lemma concludes the review on probability and general stochastic processes. In the following section, we shall focus on a specific class of stochastic processes called *Itô diffusions*, or *Itô stochastic differential equations*.

2.3 Forward Stochastic Differential Equations

Throughout the rest of this dissertation, we shall assume $(\Omega, \mathscr{F}, \{\mathscr{F}_s\}_{s\geq 0}, \mathbb{P})$ to be a complete filtered probability space on which a p-dimensional standard Brownian motion W_s is defined, such that $\{\mathscr{F}_s\}_{s\geq 0}$ is the natural filtration of W_s augmented by all $\mathbb{P}$-null sets.

2.3.1 The Forward Process

As a forward process we shall define the square-integrable, $\{\mathscr{F}_s\}_{s\geq 0}$-adapted (also called *progressively measurable*) process $X(\cdot)$[1], which, for any given $(t,x)\in[0,T]\times\mathbb{R}^n$, satisfies the Itô stochastic differential equation (SDE)

$$\begin{cases} \mathrm{d}X_s = b(s,X_s)\mathrm{d}s + \Sigma(s,X_s)\mathrm{d}W_s, & s\in[t,T], \\ X_t = x. \end{cases} \tag{1}$$

The solution to this SDE, denoted as $X_s^{t,x}$, wherein (t,x) are the initial condition[2] parameters, is given in integral form as

$$X_s^{t,x} = x + \int_t^s b(\tau,X_\tau)\mathrm{d}\tau + \int_t^s \Sigma(\tau,X_\tau)\mathrm{d}W_\tau, \qquad s\in[t,T], \tag{2}$$

with τ being a dummy variable of integration. Here, the functions $b:[0,T]\times\mathbb{R}^n\to\mathbb{R}^n$, $\Sigma:[0,T]\times\mathbb{R}^n\to\mathbb{R}^{n\times p}$ are assumed to be deterministic, that is, they do not depend explicitly on $\omega\in\Omega$. The forward process (1) is also called the *state process* in the FBSDE literature.

2.3.2 Existence and Uniqueness of Solutions to FSDEs

We begin by stating the existence and uniqueness theorem for the case of forward SDEs. Specifically, existence and uniqueness is guaranteed in the presence of *global* Lipschitz continuity (uniformly in t), and a linear growth condition on b and Σ [61, 100]:

Theorem 2.1. *(Existence and Uniqueness of Solutions to SDEs): Let $T>0$ and*

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

[2] Throughout this dissertation, all initial or terminal condition equalities for random processes, such as $X_t = x$, are to be understood in the *almost sure* sense.

$b(\cdot)$, $\Sigma(\cdot)$ be measurable functions satisfying

$$\|b(s,x) - b(s,y)\| + \|\Sigma(s,x) - \Sigma(s,y)\| \leq C\|x-y\|, \qquad s \in [t,T], \quad x,y \in \mathbb{R}^n, \tag{3}$$

for some constant C and

$$\|b(s,x)\| + \|\Sigma(s,x)\| \leq D(1+\|x\|), \qquad (s,x) \in [t,T] \times \mathbb{R}^n, \tag{4}$$

for some constant D. Then the SDE (1) has a unique, square-integrable and adapted solution X_s.

A few useful remarks:

- Local Lipschitz continuity is enough to guarantee uniqueness of solutions [61], however it does not provide guarantees against a finite escape time.
- In some texts (e.g. see [138]), the linear growth condition is replaced by an integrability condition, namely $\|b(\cdot,0)\| + \|\Sigma(\cdot,0)\| \in \mathcal{L}^2([0,T])$.
- Reference [3] proves existence and uniqueness of solutions in controlled diffusions under the relaxed condition of *local* Lipschitz continuity. Somewhat less restrictive conditions also appear in [42, 84], see also [85]; these impose a *local* Lipschitz continuity along with a monotonicity condition.

2.3.3 Girsanov's Theorem on the Change of Measure

We conclude this section by presenting the Girsanov theorem, a fundamental result in the general theory of stochastic analysis. Essentially, the theorem states that one may change the drift coefficient of an Itô SDE without radically changing its law; in fact, the law of the modified process will be absolutely continuous with respect to the law of the original process, and one can compute the Radon-Nikodym derivative explicitly (see also Definition 2.11 and Lemma 2.1).

Theorem 2.2. *(Girsanov's Theorem): Let $X_s^{t,x} \in \mathbb{R}^n$ be the solution to the Itô SDE (1), and $\tilde{X}_s^{t,x}$ be the solution to the process defined by*

$$\begin{cases} \mathrm{d}\tilde{X}_s = [b(s,\tilde{X}_s) + \Sigma(s,\tilde{X}_s)K_s]\mathrm{d}s + \Sigma(s,\tilde{X}_s)\mathrm{d}W_s, & s \in [t,T], \\ \tilde{X}_t = x, \end{cases} \tag{5}$$

wherein K_s is any measurable, square-integrable and adapted process, and all functions satisfy the standard conditions for existence and uniqueness of solutions. Let $\mathrm{d}\mathbb{Q}(\omega) = M(s,\omega;T)\mathrm{d}\mathbb{P}(\omega)$, where

$$M_s \triangleq \exp\left(-\int_t^s K_\tau^\top \mathrm{d}W_\tau - \frac{1}{2}\int_t^s |K_\tau|^2 \mathrm{d}\tau \right), \qquad s \in [t,T],$$

and define

$$\tilde{W}_s \triangleq \int_t^s K_\tau \mathrm{d}\tau + W_s, \qquad s \in [t,T].$$

Then $\mathbb{Q}$ is a probability measure on $\mathscr{F}_T$, the process $\tilde{W}_s$ is a Brownian motion with respect to $\mathbb{Q}$, and we may write

$$\begin{cases} \mathrm{d}\tilde{X}_s = b(s,\tilde{X}_s)\mathrm{d}s + \Sigma(s,\tilde{X}_s)\mathrm{d}\tilde{W}_s, & s \in [t,T], \\ \tilde{X}_t = x, \end{cases} \tag{6}$$

Therefore, the $\mathbb{Q}$-law of $\tilde{X}_s^{t,x}$ is the same as the $\mathbb{P}$-law of $X_s^{t,x}$ for all $s \in [t,T]$.

More details on Girsanov's theorem can be found in references [61, 100].

2.4 FBSDE Theory

Systems of forward and backward stochastic differential equations consist of a forward process, such as the one defined in Section 2.3.1, along with a backward process. We define the the backward process in what follows.

2.4.1 The Backward Process

In contrast to the forward process, the backward process is a square-integrable, $\{\mathscr{F}_s\}_{s\geq 0}$-adapted pair $(Y(\cdot), Z(\cdot))$ defined via a BSDE satisfying a *terminal condition*:

$$\begin{cases} \mathrm{d}Y_s = -h(s, X_s^{t,x}, Y_s, Z_s)\mathrm{d}s + Z_s^\top \mathrm{d}W_s \qquad s \in [t,T], \\ Y_T = g(X_T), \end{cases} \tag{7}$$

Here, the component Z is essentially the derivative of Y with respect to W_s, and thus is uniquely determined by Y (and W_s) [141]. The solution is implicitly defined by the initial condition parameters (t,x) of the FSDE since it obeys the terminal condition $g(X_T^{t,x})$, and thus we will similarly use the notation $Y_s^{t,x}$ and $Z_s^{t,x}$. The integral form of (7) is

$$Y_s^{t,x} = g(X_T^{t,x}) + \int_s^T h(\tau, X_\tau^{t,x}, Y_\tau, Z_\tau)\mathrm{d}\tau - \int_s^T Z_\tau^\top \mathrm{d}W_\tau, \qquad s \in [t,T]. \tag{8}$$

The functions $h : [0,T] \times \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^p \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ are assumed to be deterministic, that is, they do not depend explicitly on $\omega \in \Omega$. The function $h(\cdot)$ is called *generator* or *driver*.

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 require new definitions such as the *backward Itô integral* or, more generally, the so-called *anticipating stochastic calculus* (see relevant discussion in Chapter 1 of [83]). In this work we will restrict the analysis to adapted solutions. It turns out that a terminal value problem involving BSDEs admits an adapted solution if we back-propagate the conditional expectation of the process, that is, if we set $Y_s \triangleq \mathbb{E}[Y_s|\mathscr{F}_s]$. 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.

2.4.2 Existence and Uniqueness of Solutions to FBSDEs

To guarantee existence and uniqueness of a solution (X, Y, Z) in FBSDEs, an additional Lipschitz continuity assumption of the generator as well as a growth condition on both the generator and the terminal function must be imposed [35, 83]:

Theorem 2.3. *(Existence and Uniqueness of Solutions to FBSDEs): In addition to the assumptions of Theorem 2.1, let $h(\cdot)$ and $g(\cdot)$ be measurable functions such that*

$$|h(s,x,y_1,z_1) - h(s,x,y_2,z_2)| \leq C(|y_1 - y_2| + \|z_1 - z_2\|), \tag{9}$$
$$s \in [t,T], \quad (x,y,z) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^p,$$

for some constant C and

$$|h(s,x,y,z)| + |g(x)| \leq D(1 + \|x\|^q), \qquad (s,x,y,z) \in [t,T] \times \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^p, \tag{10}$$

for some constant D and real $q \geq 1/2$. Then the system of FBSDEs (1),(7) has a unique, square-integrable and adapted solution $(X(\cdot), Y(\cdot), Z(\cdot))$.

We note that, while the existence and uniqueness of solutions to BSDEs have been initially investigated for the case of drivers satisfying Lipschitz conditions for the variables y and z as stated above, the literature has since then seen substantial development. Indeed, several papers extend these results to drivers that are only continuous and satisfy linear growth [75], or superlinear in y and quadratic in z [74]. The case of quadratic growth in z has also been analyzed in [67, 125]. More results can also be found in references [21, 79]. See also Chapter 7 in [141].

2.4.3 The Markovian Property

The class of FBSDEs investigated in this work satisfy the distinguishing characteristic that the forward SDE does not depend on Y_s or Z_s. Thus, the resulting system of FBSDEs is said to be *decoupled*. If, in addition, the functions b, Σ, h and g are deterministic, then the adapted solution (Y, Z) exhibits the *Markovian* property; namely, it can be written as deterministic functions of solely time and the state process. Using an induction argument, the following theorem is proven [35]:

Theorem 2.4. *(The Markovian Property): There exist two deterministic measurable 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 the BSDE (7) is*

$$Y_s^{t,x} = v(s, X_s^{t,x}), \qquad Z_s^{t,x} = \Sigma^\top(s, X_s^{t,x}) d(s, X_s^{t,x}), \qquad s \in [t,T]. \tag{11}$$

Furthermore, if b, Σ, h *and* g *are continuously differentiable with respect to* (x, y, z) *with uniformly bounded derivatives, then for* $s \in [t,T]$, $x \in \mathbb{R}^n$,

$$Z_s^{t,x} = \Sigma^\top(s, X_s^{t,x}) \partial_x v(s, X_s^{t,x}), \qquad s \in [t,T]. \tag{12}$$

The Markovian property established by the above theorem will be proven to be of paramount importance in the process of obtaining numerical schemes to solve systems of FBSDEs. Specifically, it implies that the conditional expectations present in any backward scheme can be viewed as functions of time and the state process only. Locating these functions is of course an infinite dimensional problem, but one may still obtain a satisfactory approximation by considering the projection on a finite dimensional subspace of functions. This topic will be investigated in greater detail during the review on numerical methods in Section 4.3.

2.4.4 Connections to PDEs

There is an innate relation between stochastic differential equations and second-order partial differential equations of parabolic or elliptic type. Specifically, solutions to a certain class of nonlinear partial differential equations (PDEs) can be represented by solutions to FBSDEs, in the same spirit as demonstrated by the famous Feynman-Kac formulas [61, 118] for linear PDEs and forward SDEs. Although several results exist featuring slightly different conditions and restrictions [35, 83, 101, 138, 141], in this work we shall present two equivalence theorems. The first one links a PDE to a system of FBSDEs, and is taken from [138], while the second, establishing the converse, appears in [35].

Theorem 2.5. *(Nonlinear Feynman-Kac): Consider the Cauchy problem*

$$\begin{cases} v_t + \frac{1}{2}\mathrm{tr}(v_{xx}\Sigma(t,x)\Sigma^\top(t,x)) + v_x^\top b(t,x) + h(t,x,v,\Sigma^\top(t,x)v_x) = 0, \\ (t,x) \in [0,T) \times \mathbb{R}^n, \quad v(T,x) = g(x), \quad x \in \mathbb{R}^n, \end{cases} \tag{13}$$

wherein the functions Σ, b, h and g satisfy mild regularity conditions (see Remark 2.1). Then (13) admits a unique (viscosity) solution $v : [0,T] \times \mathbb{R}^n \to \mathbb{R}$, which has the following probabilistic representation:

$$v(t,x) = Y_t^{t,x}, \qquad \forall (t,x) \in [0,T] \times \mathbb{R}^n, \tag{14}$$

wherein $(X(\cdot), Y(\cdot), Z(\cdot))$ is the unique adapted solution of the FBSDE system (1),(7). Furthermore,

$$(Y_s^{t,x}, Z_s^{t,x}) = \Big(v(s, X_s^{t,x}),\ \Sigma^\top(s, X_s^{t,x}) v_x(s, X_s^{t,x}) \Big), \qquad s \in [t,T], \tag{15}$$

and if (13) admits a classical solution, then (14) provides that classical solution.

Remark 2.1. *Concerning the regularity conditions of Theorem 2, [138] 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 with respect to (t, x). However, the nonlinear Feynman-Kac lemma has been recently extended to cases in which the driver is only continuous, and satisfies quadratic growth in z; see References [20, 27, 67, 74]. See also Theorem 7.3.6 in [141].*

Remark 2.2. *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 nondegenerate PDE*

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

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

Several extensions to the nonlinear Feynman-Kac lemma appear in the literature to treat more general cases of PDEs. See for example [14, 103] for fully nonlinear PDEs, or [62–64] for a treatment on the Hamilton-Jacobi-Bellman PDE. More general PDEs are also treated via second-order BSDEs (2BSDEs) [22, 49, 105, 106, 119, 144].

The second theorem is a converse to Theorem 2.5, proven for the special case in which Y is one dimensional [35]:

Theorem 2.6. *(Nonlinear Feynman-Kac Converse): Suppose that the FBSDE solution Y is one-dimensional and that h and g are uniformly continuous with respect to x. Then the function v defined by $v(t,x) = Y_t^{t,x}$ is a viscosity solution of the PDE (13).*

We note that the viscosity solution of Theorem 2.6 can also be proven to be unique under more restrictive conditions on the generator function [35].

III

STOCHASTIC OPTIMAL CONTROL – $\mathcal{L}^2$ FORMULATION

In this chapter, we define the $\mathcal{L}^2$- type formulation of the stochastic optimal control problem. This specific class of stochastic optimal control allows for an explicit minimization of the Hamiltonian term within the Hamilton-Jacobi-Bellman (HJB) equation, thus greatly simplifying the structure of the problem. We shall demonstrate that under a certain decomposability condition, the HJB equation exhibits the same form as the Cauchy problem (13) of Theorem 2.5 in Section 2.4.4. Thus, we can obtain the solution to the HJB equation by solving the associated system of FBSDEs. The discussion on the numerical solution procedures of FBSDE systems is further postponed until Chapter 4.

3.1 Problem Statement

On the filtered probability space $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P})$, consider the problem of minimizing the expected cost defined by the cost functional

$$J(u(\cdot);\tau,x_\tau) = \mathbb{E}\big[g(x(T)) + \int_\tau^T q(t,x(t)) + \frac{1}{2}u^\top(t)Ru(t)\mathrm{d}t\big], \tag{16}$$

associated with the stochastic controlled system, which is represented by the Itô stochastic differential equation (SDE)

$$\begin{cases} \mathrm{d}x(t) = f(t,x(t))\mathrm{d}t + G(t,x(t))u(t)\mathrm{d}t + \Sigma(t,x(t))\mathrm{d}W_t, & t\in[\tau,T] \\ x(\tau) = x_\tau, \end{cases} \tag{17}$$

with $T > \tau \geq 0$, wherein T is a fixed time of termination[1], $x \in \mathbb{R}^n$ is the state vector, $u \in \mathbb{R}^\nu$ is the control vector, and $\mathrm{d}W_t$ are increments of a p-dimensional standard Brownian motion. The functions $g : \mathbb{R}^n \to \mathbb{R}$, $q : [0,T] \times \mathbb{R}^n \to \mathbb{R}$, $f : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$, $G : [0,T] \times \mathbb{R}^n \to \mathbb{R}^{n\times\nu}$, and $\Sigma : [0,T] \times \mathbb{R}^n \to \mathbb{R}^{n\times p}$ are deterministic, that is, they do not depend explicitly on $\omega \in \Omega$. We assume that all standard technical conditions [138] which pertain to the filtered probability space and the regularity of functions are met, in order to guarantee existence and uniqueness of solutions to (17), and a well defined cost functional (16). These conditions include the following:

i) The functions g, q, f, G and Σ are continuous w.r.t. time t (in case there is explicit dependence), Lipschitz (uniformly in t) with respect to the state variables, and satisfy a standard growth condition over the domain of interest (see existence and uniqueness of solutions to SDEs, Section 2.3.2). This guarantees that the SDE solution does not have a finite escape time, similar to the case of ordinary differential equations.

ii) $R \in \mathcal{P}^\nu$, where $\mathcal{P}^\nu$ denotes the set of all $(\nu \times \nu)$ positive definite real symmetric matrices.

iii) The control process $u : [0,T] \times \Omega \to U$, with U being a compact subset of $\mathbb{R}^\nu$, is square-integrable and $\{\mathscr{F}_t\}_{t\geq 0}$-adapted[2]. The latter essentially translates into the control input being non-anticipating, i.e., relying only on past and present information. We denote the set of all admissible U-valued functions as $\mathscr{U}[\tau, T]$.

[1] Optimal control problems in which the duration is not fixed a priori will be addressed in Chapter 8.

[2] see Definitions 2.6, 2.7 in Section 2.2.

For any given initial condition (τ, x_τ), we wish to minimize (16) under all admissible functions $u(\cdot) \in \mathscr{U}[\tau, T]$. We define the value function V as

$$\begin{cases} V(\tau, x_0) = \inf\limits_{u(\cdot)\in\mathscr{U}[\tau,T]} J(\tau, x_\tau; u(\cdot)), & (\tau, x_\tau) \in [0, T) \times \mathbb{R}^n, \\ V(T, x) = g(x), & x \in \mathbb{R}^n. \end{cases} \tag{18}$$

By applying the stochastic version of Bellman's principle of optimality, it is shown [39, 138] that if the value function is in $C^{1,2}([0, T] \times \mathbb{R}^n)$, then it is a solution to the following terminal value problem of a nonlinear second order partial differential equation, known as the Hamilton-Jacobi-Bellman equation:

$$\begin{cases} v_t + \inf\limits_{u\in U} H(t, x, u, v_x, v_{xx}) = 0, & (t, x) \in [0, T) \times \mathbb{R}^n, \\ v(T, x) = g(x), & x \in \mathbb{R}^n. \end{cases} \tag{19}$$

where v_x and v_{xx} denote the gradient and the Hessian of v, respectively, and the Hamiltonian H is defined as

$$H(t, x, u, p, P) \triangleq \frac{1}{2}\mathrm{tr}(P\Sigma(t, x)\Sigma^\top(t, x)) + p^\top(f(t, x) + G(t, x)u) + q(t, x) + \frac{1}{2}u^\top Ru, \tag{20}$$

$$\forall (t, x, u, p, P) \in [0, T] \times \mathbb{R}^n \times U \times \mathbb{R}^n \times \mathcal{S}^n,$$

where $\mathcal{S}^n$ denotes the set of all $(n \times n)$ non-negative definite real symmetric matrices. Note that this result can be extended to include cases where the value function does not satisfy the smoothness condition. Then, if one also considers viscosity solutions of (19), the value function is proven to be a viscosity solution of (19). Furthermore, the viscosity solution is equal to the classical solution, if a classical solution exists. For the chosen form of the cost integrand at hand, and assuming that the optimal control lies in the interior of U, we may carry out the infimum operation by taking

the gradient of the Hamiltonian with respect to u and setting it equal to zero, thus obtaining

$$\frac{\partial H}{\partial u} = 0 \quad \text{or} \quad Ru + G^\top(t,x)v_x(t,x) = 0, \tag{21}$$

and therefore the optimal control is given by

$$u^*(t,x) = -R^{-1}G^\top(t,x)v_x(t,x), \qquad (t,x) \in [0,T] \times \mathbb{R}^n. \tag{22}$$

Inserting the above expression back into the original HJB equation and suppressing function arguments for notational brevity, we obtain the equivalent characterization

$$\begin{cases} v_t + \frac{1}{2}\mathrm{tr}(v_{xx}\Sigma\Sigma^\top) + v_x^\top f + q - \frac{1}{2}v_x^\top G R^{-1} G^\top v_x = 0, \quad (t,x) \in [0,T) \times \mathbb{R}^n, \\ v(T,x) = g(x), \quad x \in \mathbb{R}^n. \end{cases} \tag{23}$$

3.2 A Feynman-Kac type Representation

A comparison of equations (23) and (13) indicates that the nonlinear Feynman-Kac representation can be applied to the HJB equation given by (23) under a certain decomposability condition, stated in the following assumption:

Assumption 3.1. *There exists a matrix-valued function* $\Gamma : [0,T] \times \mathbb{R}^n \to \mathbb{R}^{p \times \nu}$ *such that* $G(t,x) = \Sigma(t,x)\Gamma(t,x)$ *for all* $(t,x) \in [0,T] \times \mathbb{R}^n$.

This assumption implies that the range of G must be a subset of the range of Σ, thus excluding the case of a channel containing control input but no noise, although the converse is allowed. This is a fundamental difference between the proposed approach and already existing sampling-based methods for stochastic control relying on the linear Feynman-Kac lemma: in the latter category, it is additionally required that no channel may exist in which there is noise but no control input, and the choice of the design parameter R (i.e., the control cost penalty) is restricted by the stochasticity characteristics of the system. Thus, the proposed approach imposes significantly

less restrictive conditions. Under Assumption 3.1, the HJB equation given by (23) can be rewritten as

$$\begin{cases} v_t + \frac{1}{2}\mathrm{tr}(v_{xx}\Sigma\Sigma^\top) + v_x^\top f + q - \frac{1}{2}v_x^\top \Sigma\Gamma R^{-1}\Gamma^\top\Sigma^\top v_x = 0, \quad (t,x) \in [0,T) \times \mathbb{R}^n, \\ v(T,x) = g(x), \quad x \in \mathbb{R}^n, \end{cases} \tag{24}$$

in which function arguments have been again suppressed, and which now satisfies the format of (13) with

$$b(t,x) \equiv f(t,x), \tag{25}$$

and

$$h(t,x,z) \equiv q(t,x) - \frac{1}{2}z^\top\Gamma(t,x)R^{-1}\Gamma^\top(t,x)z. \tag{26}$$

We may thus obtain the (viscosity) solution of (24) by simulating the system of FBSDEs given by (1) and (7) using the definitions (25) and (26). Notice that (1) corresponds in this case to the uncontrolled ($u = 0$) system dynamics. Having established the equivalence of the HJB PDE problem solution to that of a system of FBSDEs, we will now investigate how FBSDEs can be solved numerically, in the following chapter.

IV

NUMERICAL SOLUTIONS TO FBSDES

This chapter is devoted to the investigation of numerical methods for the class of FBSDEs involved in this dissertation. In general, the procedure of obtaining a numerical solution for a system of FBSDEs consists of three elements: (a). a time discretization scheme for the forward process, (b). a time discretization scheme for the backward process, and (c). a numerical approximation scheme for the conditional expectation evaluation in each time step of the backward process. We provide a brief overview of the literature, introducing some of the most thoroughly studied time discretization and conditional expectation approximation schemes. We then propose a novel and efficient numerical scheme, suitable for the particular type of FBSDE systems considered in this dissertation, that greatly reduces the computational complexity in obtaining a solution, while exhibiting higher accuracy in simulations.

4.1 PDE vs. FBSDE Algorithms

A first distinction between algorithms can be observed on the basis of whether they target directly FBSDEs or PDEs. Indeed, as discussed in Section 2.4.4, there is an equivalence between FBSDE problems and certain PDE problems. Thus, one branch of numerical algorithms for FBSDEs does not directly solve these FBSDEs, but instead focuses on obtaining numerical solutions to their associated PDEs (see for example [30, 36, 80, 82, 90]). In general, these algorithms have limited practical applicability due to low performance in cases where the coefficients are not smooth and/or in high-dimensional problems, owing to their bad scalability. However, for low-dimensional cases involving smooth coefficients, they are very efficient and hard to compete against. Nevertheless, in what follows we will concentrate on algorithms

that deal directly with FBSDE problems.

There are two major components typically present in numerical approximations of FBSDEs. The first component consists of selecting a time discretization of the FBSDE, which essentially involves the derivation of an appropriate propagation rule (a scheme) on a selected time grid. Two schemes are needed, namely one for the forward process and one for the backward process respectively. Due to the nature of backward SDEs, the corresponding backward scheme will necessarily involve conditional expectations. In general, these conditional expectations cannot be evaluated in closed form, and thus we arrive at the second component common in all FBSDE algorithmic procedures, namely the application of a suitable numerical approximation to estimate these conditional expectations.

4.2 Time Discretization

We begin by selecting a time grid $\{t = t_0 < \ldots < t_N = T\}$ for the interval $[t, T]$, and denote by $\Delta t_i \triangleq t_{i+1} - t_i$ the i-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-th Brownian motion increment[1]. For notational brevity, we also denote $X_i \triangleq X_{t_i}$. The simplest discretized scheme for the forward process is the Euler scheme, which is also called *Euler-Maruyama* scheme [43, 66]:

$$\begin{cases} X_{i+1} \approx X_i + b(t_i, X_i)\Delta t_i + \Sigma(t_i, X_i)\Delta W_i, & i = 0, \ldots, N-1, \\ X_0 = x. \end{cases} \tag{27}$$

Several alternative, higher order schemes exist that can be selected in lieu of the Euler scheme. The most common are the Milstein scheme as well as various Taylor schemes of different order. These build on top of the basic Euler scheme by adding correction terms. Furthermore, schemes of even higher order can be obtained using Itô-Taylor

[1]Here, ΔW_i would be simulated as $\sqrt{\Delta t_i}\xi_i$, where $\xi_i \sim \mathcal{N}(0, I)$.

approximations. Multi-step as well as implicit schemes also exist, but their application in the literature seems to be rare. A detailed analysis for all aforementioned schemes can be found in [66]. Finally, it is important to note that for some processes, such as the geometric Brownian motion for example, X_t can be obtained analytically, and thus can be sampled perfectly (i.e., without numerical error) on the selected grid.

There are several ways to discretize the backward process, leading to both explicit and implicit schemes. As a short survey, we shall first derive the simplest and most commonly used scheme, and furthermore briefly present some alternative choices. To this end, we further introduce the notation $Y_i = Y_{t_i}$ and $Z_i = Z_{t_i}$. Then, recalling that adapted BSDE solutions impose $Y_s \triangleq \mathbb{E}[Y_s|\mathscr{F}_s]$ and $Z_s \triangleq \mathbb{E}[Z_s|\mathscr{F}_s]$ (i.e., a backpropagation of the conditional expectations), we approximate equation (7) by

$$Y_i \approx Y_{i+1} + h(t_i, X_i, Y_i, Z_i)\Delta t_i - Z_i^\top \Delta W_i. \tag{28}$$

Multiplying with a Brownian increment ΔW_i and taking the conditional expectation yields

$$\begin{aligned} 0 &\approx \mathbb{E}\big[\Delta W_i(Y_{i+1} + h(t_i, X_i, Y_i, Z_i)\Delta t_i) - \Delta W_i \Delta W_i^\top Z_i|\mathscr{F}_{t_i}\big] \\ &\approx \mathbb{E}[\Delta W_i Y_{i+1}|\mathscr{F}_{t_i}] - \Delta t_i Z_i, \end{aligned} \tag{29}$$

which suggests that Z_i can be approximated as

$$Z_i \approx \frac{1}{\Delta t_i}\mathbb{E}[\Delta W_i Y_{i+1}|\mathscr{F}_{t_i}]. \tag{30}$$

Then, in order to obtain an approximation of Y_i, we apply the conditional expectation on (28) resulting in

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

By choosing to evaluate $h(\cdot)$ at Y_{i+1} instead of Y_i, the scheme can be made explicit without influencing the convergence rate [13]. The explicit backward scheme is thus summarized as

$$\mathcal{S}_1 \begin{cases} \text{Initialize: } Y_N \approx g(X_N), \\ Z_i \approx \dfrac{1}{\Delta t_i}\mathbb{E}[\Delta W_i Y_{i+1}|\mathscr{F}_{t_i}], \\ Y_i \approx \mathbb{E}[Y_{i+1} + h(t_i, X_i, Y_{i+1}, Z_i)\Delta t_i|\mathscr{F}_{t_i}], \end{cases} \tag{32}$$

iterated for $i = N-1, \ldots, 0$. Scheme $\mathcal{S}_1$, which is of order $1/2$, is by far the most well-established scheme in the literature [19, 141]. It was initially proposed independently by both [15, 140], wherein a detailed convergence analysis can be found (see also [14] and the error analysis in [44]). Extensions to the case of jump-diffusions can be found in [13, 73]. Concerning the implicit version of $\mathcal{S}_1$, in which $h(\cdot)$ is evaluated at Y_i instead of Y_{i+1}, equation (31) can be solved iteratively within each time step (a so-called inner iteration) [46]. A variation of the implicit version of this scheme involving importance sampling as a means of reducing the variation of the conditional expectation approximation has been proposed by [94].

By virtue of the tower property of conditional expectations[2], scheme $\mathcal{S}_1$ can be written equivalently as [47]

$$\mathcal{S}_2 \begin{cases} Z_i \approx \dfrac{1}{\Delta t_i}\mathbb{E}\left[\Delta W_i\left(g(X_T) + \displaystyle\sum_{k=i+1}^{N-1} h_k(t_k, X_k, Y_{k+1}, Z_k)\Delta t_i\right)|\mathscr{F}_{t_i}\right], \\ Y_i \approx \mathbb{E}\left[g(X_T) + \displaystyle\sum_{k=i}^{N-1} h_k(t_k, X_k, Y_{k+1}, Z_k)\Delta t_i|\mathscr{F}_{t_i}\right], \end{cases} \tag{33}$$

iterated for $i = N-1, \ldots, 0$. It is important to note that, although the schemes $\mathcal{S}_1$ and $\mathcal{S}_2$ are mathematically identical, their numerical properties in practice are very different. Indeed, when conditional expectations are approximated numerically (see

[2]For a random variable Y_s which is $\mathscr{F}_s$-measurable and $\mathscr{F}_s \subset \mathscr{F}_T$, the tower property reads $\mathbb{E}[Y_s|\mathscr{F}_s] = \mathbb{E}\Big[\mathbb{E}[Y_s|\mathscr{F}_T]\big|\mathscr{F}_s\Big]$.

following section), then the two schemes cease to be identical and the latter scheme exhibits smaller propagation of errors. For an implicit version of $\mathcal{S}_2$, which evaluates $h(\cdot)$ at Y_k instead of Y_{k+1} and leads to the application of a Picard type iterative procedure (outer iteration) see [8, 45, 46]. Again however, no improvement in the convergence rate can be found compared to the explicit version [47].

Higher order discretization schemes for FBSDEs are available in the literature and are based on the trapezoidal (Crank-Nicolson) rule [76, 142, 143], but do so at the expense of introducing more conditional expectations that have to be evaluated. Indeed, regardless of how the backward process is discretized, in all cases the schemes involve calculating such conditional expectations. For FBSDEs within the particular class considered in this dissertation, and by virtue of the Markovian property of solutions presented in Section 2.4, all expectations in schemes $\mathcal{S}_1$ and $\mathcal{S}_2$, which are conditioned on $\mathscr{F}_{t_i}$, can be replaced with expectations conditioned on X_i. This is a critical step towards the development of an implementable scheme that can be used in practice. In general however, these conditional expectations still cannot be obtained in closed form, and thus need to be approximated numerically. There are several ways in which these approximations can be performed, giving rise to different algorithms.

4.3 Conditional Expectation Approximation Methods

In this section we will review several numerical methods employed to approximate the conditional expectations that arise in the backward process discretization step as described in the previous section. Indeed, there are several different techniques appearing in the literature including

- Approximation of the driving Brownian motion by a scaled random walk, and calculation of the conditional expectations using a tree structure [16, 81]. This method is suitable for low-dimensional problems.

- Quantization methods for reflected BSDEs [6] and coupled FBSDEs [26], which

present a probabilistic approach in which a random variable is replaced by its projection on a finite grid.

- Gauss-Hermite Quadrature [145], Cubature methods [23], and sparse grid methods [139]. These methods rely on approximating the integral of the expectation on a specific number of grid points.

- The Fourier Cosine method [112, 113]. Given the terminal condition $g(\cdot)$, the Fourier Cosine method is initialized by expanding the solution at the terminal condition into Fourier cosine series, wherein the integration is performed over suitably truncated grid. Then, the series coefficients are back propagated until the initial condition is reached. Being a grid-based method, the Fourier Cosine method is suitable for low-dimensional problems due to its bad scalability. Also, it may suffer from the Gibbs phenomenon, in which case the use of spectral filters for smoothing is required.

- Monte Carlo based methods, which include nonparametric kernel estimators, Malliavin Monte Carlo [12, 15], and Least Squares Monte Carlo [8, 9, 28, 44, 46, 47, 73, 94], with the latter being arguably the most established method for FBSDE applications so far. Monte Carlo methods are especially promising due to their good scalability properties. We shall examine these methods in more detail in what follows.

4.4 Monte Carlo Based Methods for Conditional Expectation Approximation

The main advantage of Monte Carlo based methods for conditional expectation approximation is that, in theory, the convergence rate does not depend a priori on the dimension of the problem, thus rendering them robust to the curse of dimensionality. In practice however, the performance of these estimators in terms of variance and

convergence rate usually does depend on the complexity and dimension of the problem, and therefore the above statement needs to be tempered [14]. Still, although not completely immune, these methods remain the most promising approach so far to address high dimensional problems.

Monte Carlo methods for conditional expectation approximation address the general problem of numerically estimating conditional expectations of the form $\mathbb{E}[Y|X]$ for square integrable random variables X and Y, if one is able to sample M independent copies of pairs (X, Y). They are based on the principle that the conditional expectation of a random variable can be modeled as a function of the variable on which it is conditioned on, that is, $\mathbb{E}[Y|X] = \phi^*(X)$, where ϕ^* solves 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 $\mathbb{E}[|\phi(X)|^2] < \infty$. Thus, the goal is to infer the mapping ϕ^* given only a finite amount of sample data, a classic *regression* problem within the field of machine learning. In theory, any approach developed within the machine learning framework can be employed to solve this problem. Several practical limitations arise however when this framework is to be applied specifically to FBSDEs. Indeed, recall that most of the schemes presented in Section 4.2 require at least two conditional expectation approximations *per time step*. Thus, a useful approximation should be relatively fast in order to keep the total running time of the algorithm reasonable, but also accurate enough to avoid accumulation of numerical errors during back propagation.

4.4.1 Nonparametric Kernel Estimators

In this approach, the conditional expectation is written as [95]

$$\phi^*(x) = \mathbb{E}[Y|X = x] = \int yp(y|x)\mathrm{d}y = \frac{\int yp(x,y)\mathrm{d}y}{\int p(x,y)\mathrm{d}y}. \tag{35}$$

Having M sample pairs of (X^j, Y^j), $j = 1, \ldots, M$, kernel density estimation suggests

$$p(x,y) \approx \frac{1}{M}\sum_{j=1}^{M} \kappa_h(x - X^j)\kappa_h(y - Y^J), \tag{36}$$

wherein $\kappa_h(\cdot)$ represents a chosen kernel function, parameterized by h. Substituting this expression in equation (35) and using the properties of smoothing kernels leads to the conditional expectation approximation

$$\phi^*(x) \approx \frac{\sum_{j=1}^{M} \kappa_h(x - X^j)Y^j}{\sum_{j=1}^{M} \kappa_h(x - X^j)}. \tag{37}$$

This method is called kernel regression, kernel smoothing, or the Nadaraya-Watson model [96, 134]. Several kernel choices exist, such as Gaussian, RBF for higher dimensional inputs, Epanechnikov, tri-cube etc. See [95] for a detailed presentation. A similar result using indicator functions can be found in [99].

We note that, although this method has been applied to approximate conditional expectations, it has not been employed in the context of FBSDEs so far. This is probably due to its non parametric nature, in the sense that all data generated at each time step need to be retained for inference, which renders its use rather cumbersome.

4.4.2 The Malliavin Monte Carlo Method

The Malliavin Monte Carlo Method for approximating conditional expectations was introduced in [12]. It uses the Malliavin integration by parts formula to estimate conditional expectations of the form $\mathbb{E}[Y|X = x] \triangleq \phi^*(x)$, which is given as a ratio

of two statistics, in a way similar to the one used in the kernel estimators presented in the previous paragraph. Given $M = NK$ (N being the number of time steps, K being a positive integer) independent copies of (X^j, Y^j), the conditional expectation is expressed as

$$\mathbb{E}[Y|X = x] \approx \frac{\sum_{j=1}^{M} Y^j H_x(X^j) S^j}{\sum_{j=1}^{M} H_x(X^j) S^j}. \tag{38}$$

Here, H_x is the Heaviside function, defined as $H_x(y) = \prod_{i=1}^{n} 1_{x_i \leq y_i}$, with i being a particular dimension of x, and S^j are independent copies of a random variable whose precise definition depends on the particular application. In the context of FBSDEs, the reader is referred to Section 6 of [15] for more details. We note the following important remarks:

- As in kernel estimation of Section 4.4.1, the regression estimator is the ratio of two statistics, which is not guaranteed to be integrable. This difficulty is alleviated in [15] by introducing a truncation procedure along the above backward simulation scheme.

- By suitably modifying the numerator of (38), one can obtain an equivalent expression which exhibits lower variance (see Remark 3.3 in [12]).

- In the case of FBSDEs, applying this method requires strict regularity conditions on the forward process. Indeed, [15] assumes that $\Sigma(t, x)$ is invertible for all (t, x), and that b, Σ, and Σ^{-1} are in C_b^∞ (i.e., infinitely many times continuously differentiable, bounded functions).

4.4.3 The Least Squares Monte Carlo Method

The Least Squares Monte Carlo (LSMC) method for approximating conditional expectations is arguably the most established method in FBSDE applications literature [8, 9, 28, 44, 46, 47, 73, 94], and has been studied extensively within the FBSDE framework. Initially introduced in the field of financial mathematics by Longstaff

and Schwartz in 2001 [78], the method suggests a finite-dimensional approximation of problem (34) by decomposing $\phi^*(\cdot) \approx \sum_{k=1}^{K} \varphi_k(\cdot)\alpha_k^* = \varphi(\cdot)\alpha^*$, with $\varphi(\cdot)$ being a row vector of K predetermined basis functions and α a column vector of constants, thus solving

$$\alpha^* = \arg\min_{\alpha \in \mathbb{R}^K} \mathbb{E}[|\varphi(X)\alpha - Y|^2], \tag{39}$$

with k being the dimension of the basis. This problem is then simplified to a linear least-squares problem if one substitutes the expectation operator with its empirical estimator [50], thus obtaining

$$\alpha^* = \arg\min_{\alpha \in \mathbb{R}^K} \frac{1}{M} \sum_{j=1}^{M} |\varphi(X^j)\alpha - Y^j|^2, \tag{40}$$

wherein (X^j, Y^j), $j = 1, \ldots, M$ are independent copies of (X, Y). Introducing the notation

$$\Phi(X) = \begin{bmatrix} \varphi(X^1) \\ \vdots \\ \varphi(X^M) \end{bmatrix} \in \mathbb{R}^{M \times K}, \tag{41}$$

the solution to this least-squares problem can be obtained by directly solving the normal equation, i.e.,

$$\alpha^* = \left(\Phi^\top(X)\Phi(X)\right)^{-1} \Phi^\top(X) \begin{bmatrix} Y^1 \\ \vdots \\ Y^M \end{bmatrix}, \tag{42}$$

or by performing gradient descent. The LSMC estimator for the conditional expectation assumes then the form $\mathbb{E}[Y|X = x] \triangleq \phi^*(x) \approx \varphi(x)\alpha^*$. This procedure is incorporated within the Schemes $\mathcal{S}_1$ and $\mathcal{S}_2$ to substitute each conditional expectation quantity, for each time step. Of course, the basis functions can differ both for different conditional expectations as well as for different time steps. Using LSMC on

FBSDEs was first suggested by [46], which also contains an analysis on the different error sources. The same method has been applied on the explicit scheme $\mathcal{S}_1$ [73], as well as on both the explicit [47] and implicit [8] version of $\mathcal{S}_2$. In general, combining one of the above schemes together with the LSMC method introduces errors. Due to the nature of back propagation, the errors accumulate as the algorithm is iterated backwards in time. This explains why $\mathcal{S}_1$ and $\mathcal{S}_2$, for example, while being mathematically identical, give rise to different numerical error propagation when LSMC is applied to them, with $\mathcal{S}_2$ exhibiting better performance [8, 47]. It also motivates the development of *martingale basis functions* [9]. In [9], the authors suggest splitting the second expectation in $\mathcal{S}_1$ into two terms, thus having three conditional expectations in total. Then, by choosing a particular set of basis functions to approximate the conditional expectation operator, one can compute the first two conditional expectation approximations in closed form rather than using linear regression, based on their values at the previous time step. Thus, linear regression is used only to estimate the third term, namely $\mathbb{E}[h(t_i, X_i, Y_{i+1}, Z_i)\Delta t_i | X_i]$, which is also the term contributing the least amount of simulation error.

4.5 A Novel, Efficient Numerical Scheme for FBSDEs

It is noteworthy to mention that all schemes presented in Section 4.2 are generic, in the sense that they can be applied to *any* decoupled system of FBSDEs. This is due to the Markovian property of decoupled FBSDEs, presented in Section 2.4.3, which stipulates that the solution $\{Y_s^{t,x}, Z_s^{t,x}\}_{s\in[t,T]}$ is given by deterministic functions $v(\cdot)$ and $d(\cdot)$ per equation (11). However, the FBSDEs that arise through the nonlinear Feynman-Kac representation of solutions to the HJB equation, as in the case at hand, exhibit an additional smoothness property. Indeed, by virtue of equation (15), the Z-process in (7) corresponds to the term $\Sigma^\top(s, X_s^{t,x})v_x(s, X_s^{t,x})$, that is, $d(\cdot) \equiv v_x(\cdot)$.

Therefore, we can write

$$Z_i = \mathbb{E}[Z_i|\mathscr{F}_{t_i}] = \mathbb{E}[\Sigma^\top(t_i, X_i)\nabla_x v(t_i, X_i)|X_i] = \Sigma^\top(t_i, X_i)\nabla_x v(t_i, X_i). \tag{43}$$

Choosing to evaluate $h(\cdot)$ in the approximation (31) at the right, namely as

$$Y_i = \mathbb{E}[Y_i|\mathscr{F}_{t_i}] \approx \mathbb{E}[Y_{i+1} + h(t_{i+1}, X_{i+1}, Y_{i+1}, Z_{i+1})\Delta t_i|X_i], \tag{44}$$

and initializing the scheme with

$$Y_T = g(X_T), \qquad Z_T = \Sigma(T, X_T)^\top \nabla_x g(X_T), \tag{45}$$

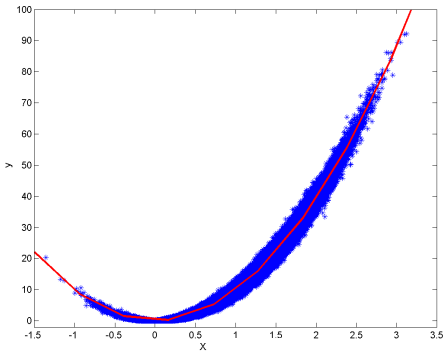
for a $g(\cdot)$ which is differentiable almost everywhere, we first perform linear regression to estimate the conditional expectation of Y as a function of x at the time step t_i using the LSMC method, and then obtain the approximation of the conditional expectation of Z by taking the gradient with respect to x on $\mathbb{E}[Y_i|X_i = x] \approx \varphi(x)\alpha_i^*$ and scaling it with Σ, i.e.,

$$Z_i \approx \Sigma(t_i, X_i)^\top \nabla_x \varphi(X_i)\alpha_i^*. \tag{46}$$

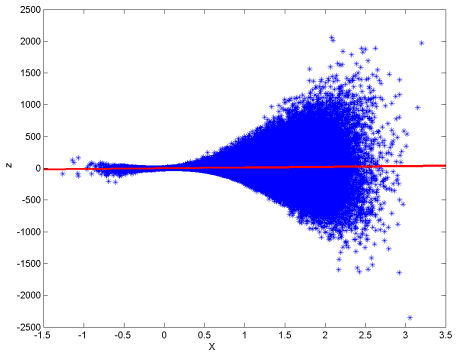
Note that this approach requires the basis functions $\varphi(\cdot)$ of our choice to be differentiable almost everywhere, so that $\nabla_x\varphi(x)$ is available in analytical form for almost any x. Combining this scheme with the LSMC method yields an algorithm which is summarized as

$$\begin{cases} \text{Initialize} : Y_T = g(X_T), \quad Z_T = \Sigma(T, X_T)^\top \nabla_x g(X_T), \\ \alpha_i^* = \arg\min\limits_{\alpha} \dfrac{1}{M}\left\|\Phi(X_i)\alpha - \left(Y_{i+1} + \Delta t_i h(t_{i+1}, X_{i+1}, Y_{i+1}, Z_{i+1})\right)\right\|^2, \\ Y_i = \Phi(X_i)\alpha_i^*, \qquad Z_i^m = \Sigma(t_i, X_i^m)^\top \nabla_x \varphi(X_i^m)\alpha_i^*, \qquad m = 1, \ldots, M, \end{cases} \tag{47}$$

where the matrix Φ is defined in (41). Again, the minimizer of equation (47) can be obtained by directly solving the normal equation (42), or by performing gradient descent.



(a) Regression on the Y data: The blue dots represent the data for the given time instant while the red curve denotes the fitted function representing the conditional expectation of Y as a function of x.



(b) Similar to (a), for the Z- regression.

Figure 1: Plots of the data set available for approximating the conditional expectation through regression, generated during the solution of a scalar linear problem, for a given time step. Notice that the estimation of Z_i through regression is very sensitive due to the nature of the data.

There are two significant advantages of this scheme as opposed to, e.g., scheme $\mathcal{S}_1$ of Section 4.2. The first one is that the proposed scheme reduces the number of computations by performing only one regression per time step, instead of the $p+1$ per time step, where p is the dimensionality of noise, required in the generic scheme. The second, even more important advantage lies in the fact that the gradient estimation per (30) is very sensitive to the number of available samples due to the nature of the data (see Figure 1), and has an increasing variance as the time steps become finer. Indeed, the worst error contribution in the generic scheme stems from estimating $Z_i \approx \frac{1}{\Delta t_i}\mathbb{E}[\Delta W_i Y_{i+1}|X_i]$, since the variance blows up as Δt_i becomes finer due to the presence of the term $\Delta W_i/\Delta t_i$ [73]. Thus, there is a significant random fluctuation

in the coefficients $\alpha_z(t)$ of the Z-regression, which decreases rather slowly[3] as the number of samples is increased, for fixed Δt_i. The modified scheme does not suffer from this phenomenon. A comparison of the ability of the two schemes to recover the coefficients given in closed form for the case of a linear-quadratic regulator (LQR) problem is given in the following section. Specifically, the coefficient comparison is depicted in Figure 3, which clearly demonstrates the superiority of the proposed scheme in recovering the gradient. A more detailed analysis on the various error sources of the framework shall be postponed until Section 5.3.

4.6 Simulation Comparison

Testing the algorithm on a linear system allows for an evaluation the performance through direct comparison with the closed form LQR solution. It also highlights the superiority of the proposed scheme, compared to the scheme $\mathcal{S}_1$ presented in Section 4.2, whenever the solution is expected to satisfy smoothness conditions. We simulate the algorithm for $f(t,x) = 0.2x = c_1 x$, $G(t,x) = \Sigma(t,x) = 0.5 = c_2$, $q(t,x) = 0$, $R = 2$, $x(0) = 1$, $T = 1$ and $g(x(T)) = 10x^2(T) = c_3 x^2(T)$, thus penalizing deviation from the origin at the time of termination, T. This problem admits a closed form solution [122] for the optimal control u^*, which is given by

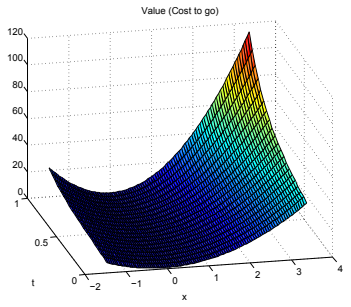
$$u^*(t,x) = -\frac{c_2}{R} P(t) x, \tag{48}$$

where $P(t)$ is the solution to the ordinary differential equation

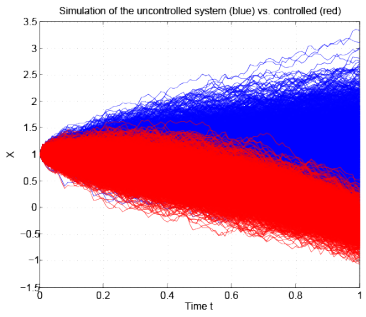
$$\begin{aligned} \dot{P}(t) &= -2c_1 P(t) + \frac{c_2^2}{R} P^2(t), \\ P(T) &= 2c_3. \end{aligned}$$

[3] In general, the error convergence rate in Monte Carlo methods is inversely proportional to the square root of the number of realizations, $\sqrt{M}$ [137]

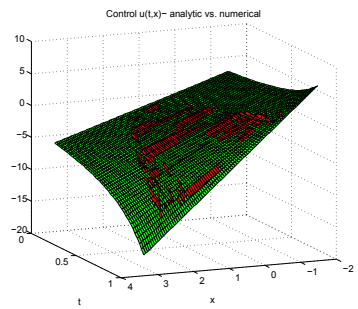
For the purposes of comparison with the closed form solution, the set of basis functions for Y was selected to be $[1\ x^2]^\top$, and $[x]$ for Z, whenever a regression for Z was employed. Figure 2(a) shows the Value function generated by the algorithm, Figure 2(b) depicts several uncontrolled and optimally controlled trajectories, while Figure 2(c) illustrates a comparison between the closed form control solution and the numerical obtained by the algorithm. Concerning the algorithm's precision, using ten thousand trajectories and a time grid of $\Delta t = 0.01$, the relative difference between the numerical and analytic value for $v(0, x_0)$ is only 0.42%. Finally, Figure 3 presents a comparison between the ability of the generic and the proposed scheme to recover the theoretical coefficients for Z, given a variety of sample sizes. It is evident that the estimation of the gradient of Y by means of separate regressions, as done in scheme $\mathcal{S}_1$, is very inefficient both in terms of computational cost (requiring $p+1$ regressions per time step instead of just one), as well as accuracy.



(a) The Value function obtained by the algorithm.

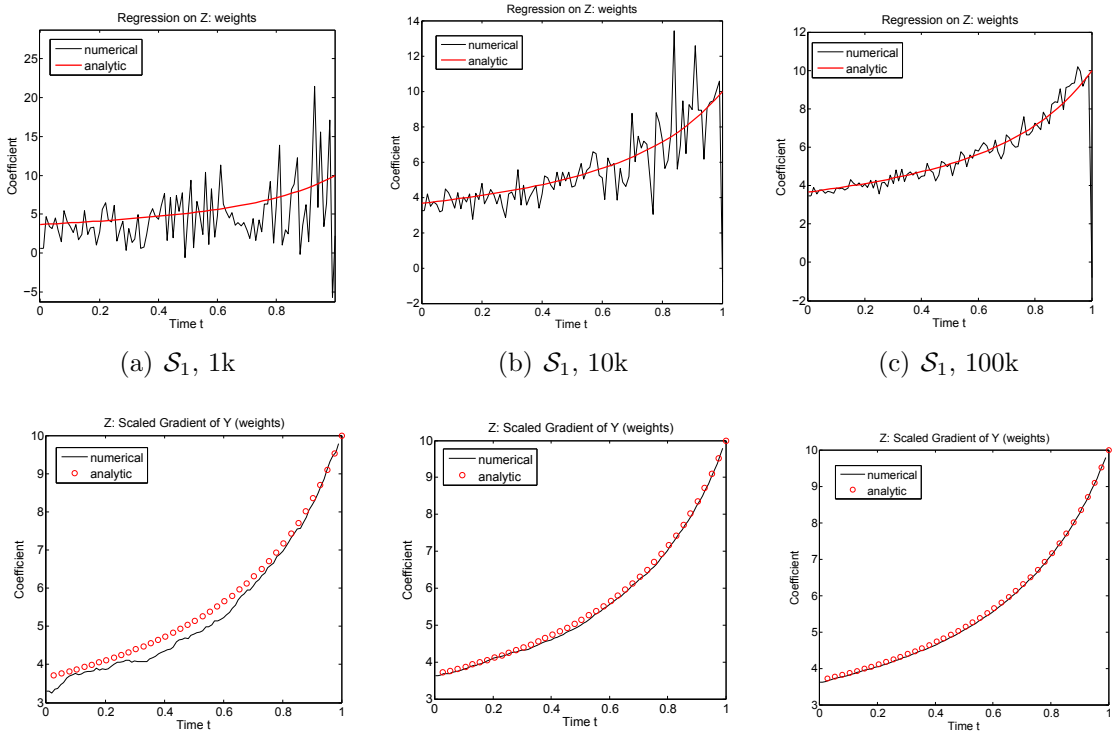


(b) Simulation of uncontrolled trajectories (blue) and optimally controlled trajectories (red) of the system.



(c) A comparison between the closed form solution for the optimal control given by equation (48) (green) and the algorithm solution (red).

Figure 2: Simulation for a scalar linear system: the value function, system trajectories and control comparison.



(a) $\mathcal{S}_1$, 1k (b) $\mathcal{S}_1$, 10k (c) $\mathcal{S}_1$, 100k

(d) Proposed, 1k (e) Proposed, 10k (f) Proposed, 100k

Figure 3: Comparison between the basis function coefficients for $Z(t,x)$ obtained numerically (black) and by the closed form theoretical solution (red). Top row – (a), (b), (c): $\mathcal{S}_1$ scheme; obtained by employing direct regression for Z, using 1k, 10k, and 100k sample trajectories respectively. Bottom row – (d), (e), (f): Proposed scheme; obtained via the scaled gradient of Y, without extra regression, for 1k, 10k, 100k sample trajectories respectively.

V

ITERATIVE METHODS AND IMPORTANCE SAMPLING

The proposed framework, as it has been presented so far, is limited in its ability to provide approximations to the value function to only those areas of the state space that are reachable by the unforced dynamics (Eq. (1)). Nevertheless, there are several cases of systems in which the goal state practically cannot be reached by the uncontrolled system dynamics (consider, for example, a forward unstable system such as the inverted pendulum). Furthermore, even in the case in which the target state is indeed reached by unforced trajectories, as the dimensionality of the state space increases, the density of sample trajectories along any given path from the initial state to the target state reduces quickly, thus increasing the demand for available samples. In this chapter, we seek to address these issues by proposing a modification to the drift term of the sampled trajectories. Specifically, by changing the drift, we can direct the exploration of the state space towards the given goal state, or any other state of interest, reachable by control. As will be shortly demonstrated, such a scheme can be constructed through Girsanov's theorem on the change of measure. While the application of Girsanov's theorem on FBSDEs is not an entirely new concept, as it was first used to facilitate variance reduction [94], and as a means to establish a connection between the nonlinear Cameron-Martin formula and FBSDEs [77]. In this dissertation however, it shall be applied to construct an iterative scheme capable of addressing control problems that exhibit more complex, nonlinear dynamics.

The present chapter is organized as follows: we first establish the equivalence between the original system of FBSDEs and one of modified drift. We then discuss the practical implementation of this result in the process of designing an iterative

scheme that is capable of recovering the optimal solution in more complex, nonlinear systems, where a single run of the algorithm is insufficient to produce good results. Section 5.3 is devoted to the analysis of convergence and the various error sources of the scheme. The chapter is concluded with simulations on the stochastic $\mathcal{L}^2$-optimal control of an inverted pendulum and a cart-pole system.

5.1 Modifying the Drift through Girsanov's Theorem

We now state and prove the main theorem in this chapter, which states that one may alter the drift of the forward process if this modification is appropriately compensated for in the backward process:

Theorem 5.1. *(Change of Measure for FBSDEs): Let $(X_s^{t,x}, Y_s^{t,x}, Z_s^{t,x})$ be the solution of the FBSDE system (1), (7), and let $K_s : [0,T] \times \Omega \to \mathbb{R}^p$ be any $\mathscr{F}_s$-adapted, bounded, and square integrable process. Now, consider the forward process with drift dynamics modified by the process K_s*

$$\begin{cases} \mathrm{d}\tilde{X}_s = [b(s,\tilde{X}_s) + \Sigma(s,\tilde{X}_s)K_s]\mathrm{d}s + \Sigma(s,\tilde{X}_s)\mathrm{d}W_s, & s \in [t,T] \\ \tilde{X}_t = x. \end{cases} \tag{49}$$

along with the compensated BSDE

$$\begin{cases} \mathrm{d}\tilde{Y}_s = [-h(s,\tilde{X}_s,\tilde{Y}_s,\tilde{Z}_s) + \tilde{Z}_s^\top K_s]\mathrm{d}s + \tilde{Z}_s^\top \mathrm{d}W_s, & s \in [t,T], \\ \tilde{Y}_T = g(\tilde{X}_T), \end{cases} \tag{50}$$

and denote its solution by $(\tilde{X}_s^{t,x}, \tilde{Y}_s^{t,x}, \tilde{Z}_s^{t,x})$. Then $(x, Y_t^{t,x}, Z_t^{t,x}) = (x, \tilde{Y}_t^{t,x}, \tilde{Z}_t^{t,x})$ almost surely. Furthermore, if

$$(Y_s^{t,x}, Z_s^{t,x}) = \Big(v(s,X_s^{t,x}),\ \Sigma^\top(s,X_s^{t,x})v_x(s,X_s^{t,x})\Big), \qquad s \in [t,T], \tag{51}$$

and

$$(\tilde{Y}_s^{t,x}, \tilde{Z}_s^{t,x}) = \Big(\tilde{v}(s, \tilde{X}_s^{t,x}),\ \Sigma^\top(s, \tilde{X}_s^{t,x})\tilde{v}_x(s, \tilde{X}_s^{t,x})\Big), \qquad s \in [t, T], \tag{52}$$

with v, $\bar{v}$ being solutions to Cauchy problems satisfying the format of (13), then $v(\cdot) \equiv \tilde{v}(\cdot)$ almost everywhere.

Proof. The first statement of Theorem 5.1 claims that the solutions of the two systems of FBSDEs coincide at the initial condition (t, x). To prove this, we define a new measure $\mathbb{Q}$ with $\mathrm{d}\mathbb{Q}(\omega) = M(t, \omega; T)\mathrm{d}\mathbb{P}(\omega)$, where

$$M_s \triangleq \exp\Big(-\int_t^s K_\tau^\top \mathrm{d}W_\tau - \frac{1}{2}\int_t^s |K_\tau|^2 \mathrm{d}\tau \Big), \qquad s \in [t, T], \tag{53}$$

is the process of Radon-Nikodym derivatives $\mathrm{d}\mathbb{Q}^{(s)}/\mathrm{d}\mathbb{P}^{(s)}$ with $\mathbb{Q}^{(s)}$ and $\mathbb{P}^{(s)}$ being the restrictions of $\mathbb{Q}$ and $\mathbb{P}$ to $\mathscr{F}_s$, respectively. Then, by Girsanov's Theorem (Theorem 2.2, see also [61,100]), M_s is a $\mathbb{P}$-martingale, the $\mathbb{P}$-law of (X, Y, Z) is the same as the $\mathbb{Q}$-law of $(\tilde{X}, \tilde{Y}, \tilde{Z})$, and

$$\tilde{W}_s \triangleq \int_t^s K_\tau \mathrm{d}\tau + W_s, \qquad s \in [t, T],$$

is a Brownian motion under $\mathbb{Q}$. Now, defining the $\mathbb{Q}$-Brownian increment $\mathrm{d}\tilde{W}_s = K_s\mathrm{d}t + \mathrm{d}W_s$, it becomes evident that equations (49) and (50) are simply copies of the dynamics of equations (1) and (7), if one substitutes $\mathrm{d}W_s$ in the latter with $\mathrm{d}\tilde{W}_s$:

$$\begin{cases} \mathrm{d}\tilde{X}_s = b(s, \tilde{X}_s)\mathrm{d}s + \Sigma(s, \tilde{X}_s)\mathrm{d}\tilde{W}_s, \qquad s \in [t, T] \\ \tilde{X}_t = x. \end{cases}$$

$$\begin{cases} \mathrm{d}\tilde{Y}_s = -h(s, \tilde{X}_s, \tilde{Y}_s, \tilde{Z}_s)\mathrm{d}s + \tilde{Z}_s^\top \mathrm{d}\tilde{W}_s, \qquad s \in [t, T], \\ \tilde{Y}_T = g(\tilde{X}_T). \end{cases}$$

Since at the time of initialization, t, M_t is by construction equal to one with probability one (in both $\mathbb{P}$ and $\mathbb{Q}$-measure), the measures $\mathbb{P}$ and $\mathbb{Q}$ restricted to $\mathscr{F}_t$ are equal, and therefore the pairs (Y_t, Z_t) and $(\tilde{Y}_t, \tilde{Z}_t)$ are equal in expectation as well. This proves that the value function at the initial condition (t, x) is independent of the drift term modification.

The second statement of Theorem 5.1 claims that if each of the two FBSDE systems are associated with the solution of a Cauchy problem respectively[1], then the solutions of these Cauchy problems match. This fact is easily established if one examines the associated PDEs. Indeed, the FBSDE problem defined by (49) and (50) corresponds to the PDE problem

$$\begin{cases} v_t + \frac{1}{2}\mathrm{tr}(v_{xx}\Sigma\Sigma^\top) + v_x^\top(b + \Sigma K) + h(t, x, v, \Sigma^\top v_x) - v_x^\top \Sigma K = 0, \\ (t, x) \in [0, T) \times \mathbb{R}^n, \quad v(T, x) = g(x), \end{cases} \tag{54}$$

which of course is identical to the PDE problem (13), as we have merely added and subtracted the term $v_x^\top \Sigma K$. Thus, although the FBSDEs are different, they are associated with the same PDE problem. □

Returning to the original problem formulation and recalling the definition of $\Gamma(\cdot)$ in Assumption 3.1, we may apply any nominal control $\bar{u}$ to the state dynamics in order to obtain the modified drift system, which exhibits the form

$$\mathrm{d}x(t) = [f(t, x(t)) + \Sigma(t, x(t))\Gamma(t, x(t))\bar{u}(t)]\mathrm{d}t + \Sigma(t, x(t))\mathrm{d}W_t. \tag{55}$$

Thus, the controlled system trajectories are sampled from the forward process (49) with

$$K_s = \Gamma(s, X_s)\bar{u}(s), \qquad s \in [t, T], \tag{56}$$

[1]Note that the second part of Theorem 5.1 assumes a link between PDEs and FBSDEs –given by a converse Feynman-Kac lemma [35]– which is not necessary in the first part.

while $b(s, X_s) \equiv f(s, X_s)$ as per (25). Notice that the nominal control $\bar{u}$[2] may be any open or closed-loop control, a random control, or even a control calculated by a previous run of the algorithm. In the latter case, one obtains a more refined solution, thus arriving at an iterative scheme. This will be discussed in greater detail in the following section.

To implement importance sampling, we return to the discrete representation of Section 4.5, and define $K_i = K_{t_i}$. The forward process can again be sampled using the Euler-Maruyama scheme. For the backward process, there are two equivalent ways in which one can incorporate importance sampling in an algorithm, but the most straightforward way is to simply define

$$\tilde{h}(s, x, y, z, k) \triangleq h(s, x, y, z) - z^\top k, \tag{57}$$

and utilize the proposed scheme using $\tilde{h}$ instead of h.

5.2 Incorporating Importance Sampling and Sample Trajectory Blending

If no initial guess for the control input is available, the algorithm can be initialized using sample trajectories generated by zero or random control inputs. In the latter case, the goal is to amplify the exploration of the state space, whenever the noise level is too low to result in adequate exploration. If an initial guess for control exists, it may speed up the iterative scheme and improve the accuracy, but it is otherwise not an absolute requirement. A full iteration of the algorithm will then provide an approximation of the value function based on the chosen basis functions, which is accurate for that particular area of the state space that was visited by the sampled trajectories. In the next iteration, sample trajectories are generated using the control

[2]Relation (56) is valid under the extra condition that the control input is bounded, a mild restriction for engineering purposes.

law (22), which is based on the value function approximation of the previous iteration. Notice however, that the sampled trajectories of these two subsequent iterations differ significantly– one was generated with zero (or random) control, whereas the other was generated using the optimal control resulting from the first iteration. Thus, different areas of the state space are visited during the sampling stages of those two iterations. Since the value function approximation is accurate only in the area visited by the initial trajectories, by evaluating the control law along the newly generated samples, we are essentially performing extrapolation. Depending on the problem, this extrapolation may or may not be accurate. If it is accurate, then a very small –if any– change will be observed in the recovered basis function coefficients after the second run has been concluded. In general however, the observed change will be significant. This is due to the discrepancy between the areas visited during the sampling stage of the algorithm, and the areas that are visited when the control law is evaluated. Intuitively, convergence of the algorithm occurs when the sample trajectory areas and controlled trajectory areas are sufficiently close or coincide.

While in many cases, solving the problem in an iterative fashion by applying a previously calculated control law, leads to a smooth convergence to the optimal trajectory and cost, there are instances in which the transition between successive controlled trajectories oscillates significantly, thus preventing convergence (see the simulations section of Chapter 6, for example). The underlying cause seems to be the algorithm's sensitivity to changes in the control law between iterations. Essentially, the control input changes too drastically between iterations, similar in nature to a gradient descent algorithm taking a step size which is too large. Mitigation of this phenomenon can be accomplished, however, through the blending of the sample trajectories used by the algorithm. Specifically, instead of generating all sample trajectories for the next iteration using solely the obtained control law, we may sample only a short percentage of the total number of them (typically 2-5%). Thus, the new

pool of sampled trajectories consists mainly (95-98%) of the same trajectories as in the previous iteration, while only a few are new, resulting from the newly obtained control law. Furthermore, we may choose the old trajectories to correspond to lowest cost realizations, thereby discarding the least favorable ones in favor of new realizations generated using a new control law. Essentially, instead of completely renewing the pool of trajectory samples and respective control inputs in each iteration, we create pools of favorable samples and controls, that remain largely the same, discarding bad trajectory-control couples in favor of newly sampled ones. This results in pools that possibly combine trajectories/controls of several previous iterations, provided they are good enough. Defining the ratio $\gamma \triangleq \frac{M^{\text{old}}}{M} \in [0, 1)$, i.e., the percentage of trajectory samples of the previous iteration present in the next iteration, the complete procedure, featuring importance sampling and trajectory blending, is summarized in the Algorithm 1 table. Note that one can also terminate this algorithm when the

Algorithm 1 NFK-FBSDE Algorithm with Importance Sampling and Sample Trajectory Blending

Input: Initial condition x_0, initial control input $\bar{u}$ if available (otherwise zero), terminal time T, number of Monte Carlo samples M, blending ratio $\gamma \in [0, 1)$, number of iterations N_{it}.

Output: Basis function coefficients for the value function, α_i.

procedure NFK_FBSDE(x_0, $\bar{u}$, T, M, γ, N, N_{it})
    Assign M control inputs $\bar{u}$ using either initial, zero, or random values, to generate a collection $\mathscr{U}_c$.
    Sample a collection $\mathscr{X}$ of M state trajectories by applying discretization (27) on equation (49), using the control sequences of $\mathscr{U}_c$;
    for $1 : N_{\text{it}}$ **do**
        Using $\mathscr{X}$ and $\mathscr{U}_c$, repeat the backward scheme (47) for $N - 1$ time steps, using $\tilde{h}$ of equation (57) to obtain α_i for each time step $i = 0, \ldots, N - 1$;
        Sample $(1-\gamma)M$ new trajectories per discretization (27) using the control law (22), and evaluate the cost (16);
        Discard $(1 - \gamma)M$ trajectories of $\mathscr{X}$ and controls $\mathscr{U}_c$ that correspond to high cost and add the newly sampled ones;
    end for
    return α_i.
end procedure

evaluation of the cost in successive iterations does not exhibit significant change, in lieu of a predetermined number of iterations N_{it}.

5.3 Scheme Convergence

The proposed algorithm consists of two components in which numerical approximation is performed, thus raising the question of convergence guarantees. Specifically, we identify the following components:

- The time discretization schemes (Section 4.2). Concerning this component, convergence of the schemes presented in Section 4.2 is established in its respective literature [15, 47, 140]. Proving the convergence of the proposed scheme is more involved, since the error in this case is no longer independent of the error arising from the numerical approximation of conditional expectations. A proof of convergence of the proposed scheme was constructed, but it was not complete before the dissertation submission deadline, and thus its publication will be postponed until a future date.

- The LSMC method of approximating conditional expectations (Section 4.4.3). Here, we may identify two sub-components:

  i) The use of a finite number of basis functions for the conditional expectation in (34). Convergence is straightforward: assuming that the unknown function ϕ lies within a space that can be spanned by a (possibly infinite) set of basis functions of our choice, the projection error vanishes as their number tends to infinity, thereby spanning the entire space in which ϕ lies. The rate of convergence, however, is difficult to analyze [141].

  ii) The use of a finite number of samples in the empirical estimator for the expectation in (40). The empirical estimator converges as the Monte Carlo samples tend to infinity, by virtue of the Law of Large Numbers. In general,

the convergence rate in Monte Carlo methods is proportional to the square root of the number of realizations, $\sqrt{M}$, by the Central Limit Theorem [137].

It is important to note that, in contrast to the aforementioned two components, the following two components do *not* require convergence analysis:

- The PDE-FBSDE problem equivalence, illustrated by the nonlinear Feynman-Kac lemma (Section 2.4.4), during which no approximation is performed.
- The importance sampling component, which is based on Girsanov's theorem (Section 5.1). Again, no convergence analysis is necessary because the two expressions are mathematically equivalent. No approximation step is performed. The different numerical properties arise only because of the finite number of samples that are used, and vanish as the number of samples tends to infinity.

Unfortunately, obtaining error bounds in the FBSDE literature has been proven to be a difficult task. We can identify three sources of error in the algorithm:

- The *time discretization error*, which is introduced as soon as a time discretization scheme is applied to the continuous forward and backward processes (1) and (7). For the $\mathcal{S}_1$-scheme, this error decreases at a rate $\sqrt{N}$, where N is the number of (equidistant) time steps [73].
- The *projection error*, which results from projecting the unknown, exact solution of the infinite dimensional problem (34) to a finite set of basis functions, in order to obtain the finite dimensional approximation (39). As noted in previous literature, this error is hard to quantify except on some special cases [46].
- The *simulation error*, which is incurred by substituting the expectation operator with its empirical estimator in equations (39)-(40) and using only a

finite number M of Monte Carlo samples for the purposes of linear regression. Concerning the $\mathcal{S}_1$ scheme, the worst contribution stems from estimating $Z_i \approx \frac{1}{\Delta t_i}\mathbb{E}[\Delta W_i Y_{i+1}|X_i]$, since the variance blows up as Δt_i becomes finer due to the presence of the term $\Delta W_i/\Delta t_i$.

5.4 Simulation Results

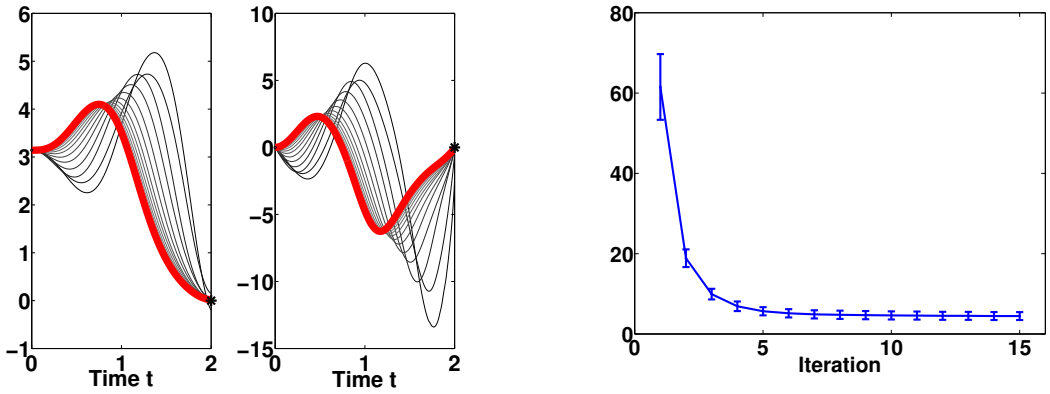
Simulations on nonlinear systems were performed to demonstrate that the nonlinearity in the dynamics is handled efficiently, and furthermore illustrate the importance of the iterative nature of the scheme when dealing with more complicated problems.

5.4.1 The Inverted Pendulum

The equations of motion for the inverted pendulum are given by

$$m\ell^2\ddot{\theta} + b\dot{\theta} - mg\ell\sin\theta = u, \tag{58}$$

and stochasticity enters the system in form of perturbations in the torque u. For the purposes of this simulation, two thousand trajectories were generated on a time grid of 0.005 with time horizon T=2. The system noise covariance was set to 0.1. No initial guess for the control input was necessary. For the basis of the Value function approximation, modified Chebyshev polynomials [65] up to second order have been selected. The scheme was repeated for 15 iterations, without any use of trajectory blending ($\gamma = 0$). The algorithm successfully learned the optimal control to invert and stabilize the pendulum. Figure 4(a) depicts the mean of the controlled trajectories for each algorithm iteration (gray scale). We observe that a balancing yet suboptimal trajectory is obtained at the very first iteration of the algorithm, while subsequent iterations further improve it until convergence. The trajectories after the final iteration are shown in red. Finally, Figure 4(b) depicts the convergence of the cost mean and standard deviation as the iterative scheme progresses.



(a) Trajectory mean for the position (left) and velocity (right) of the controlled system for each iteration (gray scale) and after the final iteration (red). The black dots represent the target states.

(b) Cost mean ± 3 standard deviations per iteration.

Figure 4: Mean optimal state trajectories and cost per iteration for the inverted pendulum.

5.4.2 The Cart-Pole System

To assess the efficiency of the proposed scheme in underactuated systems, we simulated the algorithm on a cart-pole system (see Figure 5). The equations of motion are given by

$$\ddot{x} = \frac{1}{m_c + m_p \sin^2\theta}\Big(u - m_p \sin\theta(\ell\dot{\theta}^2 + g\cos\theta)\Big), \tag{59}$$

$$\ddot{\theta} = \frac{1}{\ell(m_c + m_p \sin^2\theta)}\Big(u\cos\theta - m_p\ell\dot{\theta}^2\cos\theta\sin\theta \quad + (m_c + m_p)g\sin\theta\Big), \tag{60}$$

and stochasticity enters the system in form of perturbations in u. For the purposes of simulation, five thousand trajectories were generated on a time grid of 0.005 with time horizon T=3. The system noise covariance was set to 1. Again, no initial guess for the control input was necessary. For the basis of the value function approximation, modified Chebyshev polynomials [65] up to second order have been selected. The scheme was repeated for 35 iterations, without any use of trajectory blending ($\gamma = 0$). Figure 6(a) depicts the mean of the controlled trajectories for each algorithm iteration

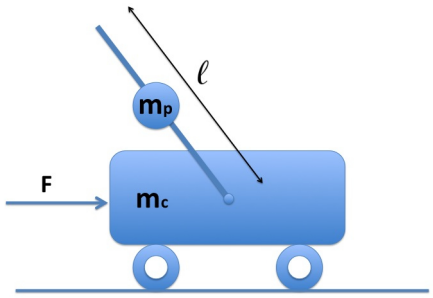
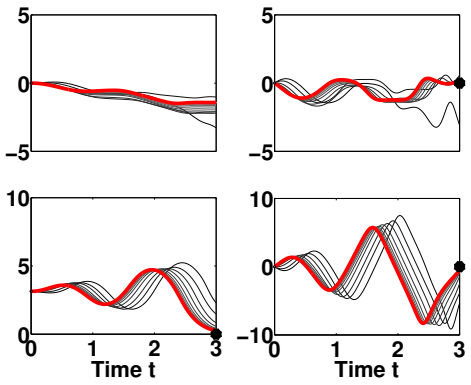
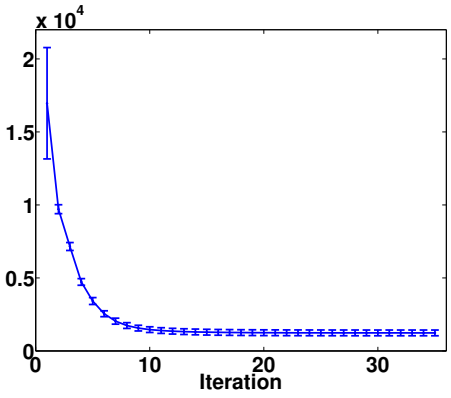


Figure 5: Cart pole: m_c denoted the mass of the cart, m_p denotes the mass of the pole and ℓ is the length of the pole.

(gray scale). The trajectories after the final iteration are shown in red. Finally, Figure 6(b) depicts the convergence of the cost mean and standard deviation as the iterative scheme progresses.



(a) Top left– cart position, top right– cart velocity, bottom left– pole position, bottom right – pole velocity. Trajectory mean of the controlled system for each iteration (gray scale) and after the final iteration (red). The black dots represent the target states.



(b) Cost mean ± 3 standard deviations per iteration.

Figure 6: Mean optimal state trajectories and cost per iteration in the cart-pole system.

VI

THE STOCHASTIC $\mathcal{L}^1$-OPTIMAL CONTROL PROBLEM

In this chapter, we turn our attention to stochastic $\mathcal{L}^1$-optimal control problems. We begin with a definition of the $\mathcal{L}^1$- type formulation, and show that this specific class of stochastic optimal control problems allows for an explicit minimization of the Hamiltonian term within the Hamilton-Jacobi-Bellman (HJB) equation, thus greatly simplifying the structure of the problem. We then demonstrate that under the same decomposability condition, namely Assumption 3.1 of Section 3.2, the HJB equation exhibits the same form as the Cauchy problem (13) of Theorem 2.5 in Section 2.4.4. Thus, we may obtain the solution to the HJB equation by solving the associated system of FBSDEs. The chapter is concluded with simulations that validate the numerical algorithm by applying it on a well-known minimum fuel problem, which offers an analytic solution. Furthermore, we demonstrate the superiority of the proposed stochastic control law against deterministic control laws, whenever they are applied in the presence of stochastic disturbances. The algorithm's ability to handle nonlinear dynamics, on the other hand, is also demonstrated by an application on the inverted pendulum system.

6.1 Problem Statement

On the filtered probability space $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P})$, consider the problem of minimizing the expected cost defined by the cost functional

$$J(\tau, x_\tau; u(\cdot)) = \mathbb{E}\bigg[g(x(T)) + \int_\tau^T q(t, x(t)) + p(t, x(t))^\top |u(t)| \, \mathrm{d}t\bigg], \tag{61}$$

associated with the stochastic controlled system, which is represented by the Itô stochastic differential equation (SDE)

$$\begin{cases} \mathrm{d}x(t) = f(t, x(t))\mathrm{d}t + G(t, x(t))u(t)\mathrm{d}t + \Sigma(t, x(t))\mathrm{d}W_t, & t \in [\tau, T], \\ x(\tau) = x_\tau. \end{cases} \tag{62}$$

In the above, T is a fixed time of termination[1], $x \in \mathbb{R}^n$ is the state vector, $\mathrm{d}W_t$ are increments of a p-dimensional standard Brownian motion, and $u \in U \subset \mathbb{R}^\nu$ is the control vector, where $U = [-u_1^{\min}, u_1^{\max}] \times [-u_2^{\min}, u_2^{\max}] \times \cdots \times [-u_\nu^{\min}, u_\nu^{\max}]$, with $u_i^{\min} \geq 0$, $u_i^{\max} > 0$. Note that the assumption about the signs of $u_i^{\min}$ and $u_i^{\max}$ is without loss of generality. The same analysis can be performed for any $u_i^{\min} < u_i^{\max}$ regardless of their sign. Furthermore, $|\cdot|$ denotes the element-wise absolute value, $p : [0, T] \times \mathbb{R}^n \to \mathbb{R}_+^\nu$ is a (possibly time/state dependent) vector of nonnegative weights, and $q : [0, T] \times \mathbb{R}^n \to \mathbb{R}$ is the state-depended part of the running cost. If the "fuel consumption" penalty is to be applied on all control channels equally, independently of time or state, then p reduces to a constant vector of ones. Finally, all aforementioned functions, as well as $g : \mathbb{R}^n \to \mathbb{R}$, $f : [0, T] \times \mathbb{R}^n \to \mathbb{R}^n$, $G : [0, T] \times \mathbb{R}^n \to \mathbb{R}^{n \times \nu}$, and $\Sigma : [0, T] \times \mathbb{R}^n \to \mathbb{R}^{n \times p}$, are deterministic, in the sense that they do not depend explicitly on $\omega \in \Omega$. We assume that all standard technical conditions [138] which pertain to the filtered probability space and the regularity of functions are met, in order to guarantee existence and uniqueness of solutions to (62), and a well defined cost functional (61). These include the following:

i) The functions g, q, p, f, G and Σ are continuous with respect to time t (in case there is explicit time dependence), Lipschitz (uniformly in t) with respect to the state variables, and satisfy a standard growth condition over the domain of interest (see existence and uniqueness of solutions to SDEs, Section 2.3.2).

[1] Optimal control problems in which the duration is not fixed a priori will be addressed in Chapter 8.

ii) The control process $u : [\tau, T] \times \Omega \to U \subset \mathbb{R}^{\nu}$ is square-integrable and $\{\mathscr{F}_t\}_{t\geq 0}$-adapted, which essentially translates into the control input being non-anticipating, i.e., relying only on past and present information. We denote the set of all admissible U-valued functions as $\mathscr{U}[\tau, T]$.

For any given initial condition (τ, x_τ), we wish to minimize (61) under all admissible functions $u(\cdot) \in \mathscr{U}[\tau, T]$. We define the value function V as

$$\begin{cases} V(\tau, x_\tau) = \inf\limits_{u(\cdot)\in\mathscr{U}[\tau,T]} J(\tau, x_\tau; u(\cdot)), \quad (\tau, x_\tau) \in [0, T) \times \mathbb{R}^n, \\ V(T, x) = g(x), \qquad x \in \mathbb{R}^n. \end{cases} \tag{63}$$

By applying the stochastic version of Bellman's principle of optimality, it is shown [39, 138] that if the value function is in $C^{1,2}([0, T] \times \mathbb{R}^n)$, then it is a solution to the following terminal value problem of a nonlinear second order partial differential equation, known as the Hamilton-Jacobi-Bellman (HJB) equation, which –omitting function arguments for brevity– assumes for the problem at hand the following form

$$\begin{cases} v_t + \inf\limits_{u\in U} \left\{ \frac{1}{2}\mathrm{tr}(v_{xx}\Sigma\Sigma^\top) + v_x^\top f + \big(v_x^\top G + p^\top \mathrm{D}(\mathrm{sgn}(u))\big)u + q \right\} = 0, \\ (t, x) \in [0, T) \times \mathbb{R}^n, \quad v(T, x) = g(x), \quad x \in \mathbb{R}^n, \end{cases} \tag{64}$$

wherein v_x and v_{xx} denote the gradient and the Hessian of v, respectively, $\mathrm{D}(x) \in \mathbb{R}^{n\times n}$ denotes the diagonal matrix with the components of $x \in \mathbb{R}^n$ in its diagonal, and $\mathrm{sgn}(\cdot)$ denotes the signum function. Note that this result can be extended to include cases where the value function does not satisfy the smoothness condition. Then, if one also considers viscosity solutions of (64), the value function is proven to be a viscosity solution of (64). Furthermore, the viscosity solution is equal to the classical solution, if a classical solution exists. For the chosen forms of cost integrand and dynamics at hand, we may carry out the infimum operation over u 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, $\text{sgn}(u_i) = +1$. Then, if $(v_x^\top G)_i + (p^\top)_i > 0$, the Hamiltonian is minimized for $u_i = -u_i^{\min} \leq 0$, which leads to a contradiction. On the other hand, if $(v_x^\top G)_i + (p^\top)_i < 0$, the Hamiltonian is minimized for $u_i = u_i^{\max} > 0$, which is consistent with the hypothesis.

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

The optimal control law thus obtained is given by

$$u_i^* = \begin{cases} u_i^{\max}, & (v_x^\top G)_i < -(p^\top)_i \\ -u_i^{\min}, & (v_x^\top G)_i > (p^\top)_i, \\ 0, & -(p^\top)_i < (v_x^\top G)_i < (p^\top)_i, \end{cases} \qquad i = 1, \cdots, \nu, \tag{65}$$

namely, the optimal control law turns out to be *bang-bang* control. This covers the particular case when $u_i^{\min} = 0$ as well, in which case the corresponding condition is $-(p^\top)_i < (v_x^\top G)_i$.

Remark 6.1. *Notice that in the control law given by (65), we do not assign a value for u^* whenever $(v_x^\top G)_i = -(p^\top)_i$ or $(v_x^\top G)_i = (p^\top)_i$, because in those two cases the control input is not uniquely defined. In fact, any value in $[0, u_i^{\max}]$ and $[-u_i^{\min}, 0]$ respectively attains the same infimum value in (64). A problem in which either one of these equalities is satisfied over a nontrivial time interval is a singular fuel-optimal problem [4]. In what follows, we shall assume that the minimum fuel problem is normal, in the sense that the aforementioned equalities are not satisfied over a nontrivial time interval, $\mathbb{P}$- almost surely.*

Substituting the control law given by (65), the HJB equation 64 assumes the equivalent form

$$\begin{cases} v_t + \frac{1}{2}\mathrm{tr}(v_{xx}\Sigma\Sigma^\top) + v_x^\top f + q + \sum_{i=1}^{\nu} \min\left\{ \left(v_x^\top G + p^\top\right)_i u_i^{\max},\ 0, -\left(v_x^\top G - p^\top\right)_i u_i^{\min} \right\} = 0, \\ (t,x) \in [0,T) \times \mathbb{R}^n, \qquad v(T,x) = g(x), \quad x \in \mathbb{R}^n. \end{cases} \tag{66}$$

6.2 A Feynman-Kac type Representation

A comparison of equations (66) and (13) indicates that the nonlinear Feynman-Kac representation can be applied to the HJB equation given by (66) under Assumption 3.1, in which case (66) satisfies the format of (13) with

$$b(t,x) \equiv f(t,x) \tag{67}$$

and

$$h(t,x,z) \equiv q + \sum_{i=1}^{\nu} \min\left\{ \left(z^\top \Gamma + p^\top\right)_i u_i^{\max},\ 0,\ -\left(z^\top \Gamma - p^\top\right)_i u_i^{\min} \right\}. \tag{68}$$

We may thus obtain the (viscosity) solution of (66) by simulating the system of FBSDEs given by (1) and (7) using the definitions (67) and (68).

6.3 Simulation Results

The aim of the simulations presented in this section is twofold. First, the proposed algorithm is validated by means of an application to a linear problem for which an open loop control law is available in closed-form for the deterministic setting of that problem. This is the double integrator problem in Section 6.3.1. We demonstrate that the algorithm is able to recover the optimal control sequence, using only importance sampling. For this problem, sample trajectory blending is not necessary. Furthermore, the obtained stochastic feedback control law is shown to outperform

both the deterministic open loop as well as the deterministic closed-loop control law in the presence of noise. Finally, in Section 6.3.2, the ability of the algorithm to handle nonlinear dynamics, as well as the significance of the sample trajectory blending technique, are demonstrated through simulations on an inverted pendulum system.

6.3.1 The Double Integrator

To validate the proposed algorithm on stochastic $\mathcal{L}^1$-optimal control problems, we tested it on the fuel-optimal control problem of a stochastic double integrator plant. The deterministic case offers a closed form solution; see [4], Ch. 8-6. Specifically, the deterministic problem reads: Given the system equations

$$\dot{x}_1(t) = x_2(t) \tag{69}$$

$$\dot{x}_2(t) = u(t), \qquad |u(t)| \leq 1, \tag{70}$$

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

$$J = \int_0^T |u(t)| \, \mathrm{d}t, \tag{71}$$

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

$$T \geq x_{20} + \sqrt{4x_{10} + 2x_{20}^2}, \tag{72}$$

in which case the existence of a unique solution is guaranteed. The corresponding fuel-optimal control sequence is $\{-1, 0, +1\}$, in which the control switching times t_1

and t_2 are

$$t_1 = 0.5\left(T + x_{20} - \sqrt{(T - x_{20})^2 - 4x_{10} - 2x_{20}^2}\right), \tag{73}$$

$$t_2 = 0.5\left(T + x_{20} + \sqrt{(T - x_{20})^2 - 4x_{10} - 2x_{20}^2}\right), \tag{74}$$

that is,

$$u^*(t) = \begin{cases} -1, & t \in [0, t_1), \\ 0, & t \in [t_1, t_2), \\ 1, & t \in [t_2, T]. \end{cases} \tag{75}$$

A stochastic counterpart of this problem is obtained if the system equations are modeled in the following form

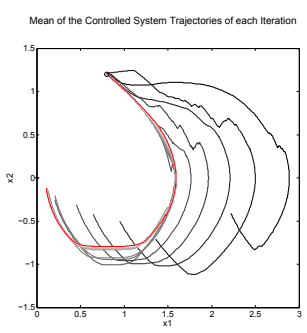
$$\mathrm{d}x_1(t) = x_2(t)\ \mathrm{d}t, \tag{76}$$

$$\mathrm{d}x_2(t) = u(t)\ \mathrm{d}t + \sigma\ \mathrm{d}w(t), \qquad |u(t)| \leq 1, \tag{77}$$

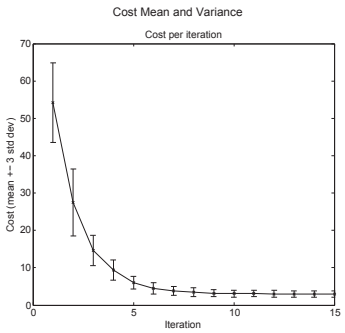
i.e., modeling stochasticity in form of perturbations in the control input u. An alternative stochastic counterpart could feature noise in the first channel as well. Terminal state conditions are not meaningful in a stochastic setting, since whenever the system dynamics are modeled by controlled diffusions, the probability of hitting a particular point in state space *exactly* is zero. Therefore, instead of the final condition $(x_1(T), x_2(T)) = (0, 0)$, we introduce a "soft" constraint in the cost function by adding a terminal cost:

$$J = \mathbb{E}\left[C(x_1^2(T) + x_2^2(T)) + \int_0^T |u(t)|\ \mathrm{d}t\right], \tag{78}$$

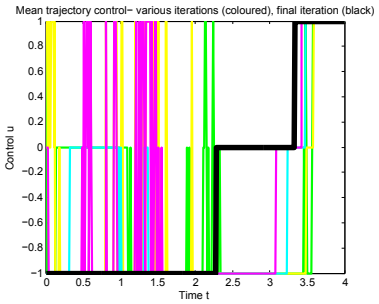
where C is a large enough constant, thus penalizing deviation from the origin at the time of termination.



(a) The mean of the controlled system trajectories of each iteration (grayscale) and after the final iteration (red).



(b) Cost mean ± 3 standard deviations per iteration.

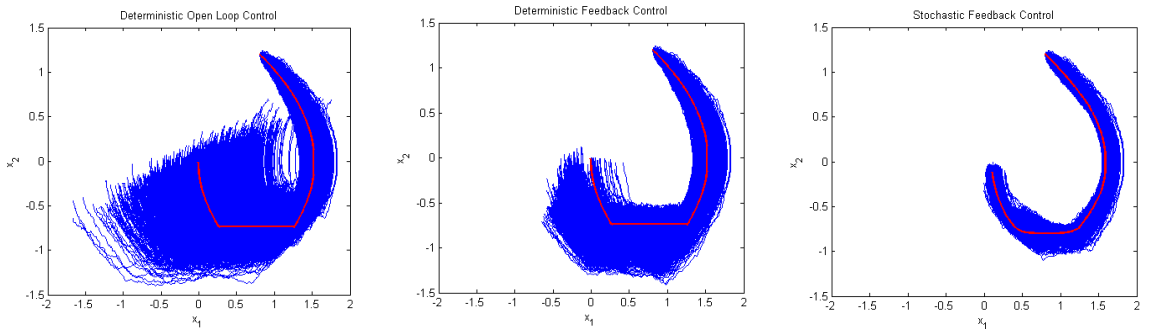


(c) The control input for the mean system trajectory for each iteration (coloured) and after the final iteration (black). We see that the optimal control sequence $\{-1, 0, +1\}$ is finally recovered.

Figure 7: Double integrator plant, phase, cost, and control sequence.

For the purposes of simulation, two thousand trajectories were generated on a time grid of $\Delta t = 0.01$, with $\sigma = 0.1$, $T = 4$ and $(x_{10}, x_{20}) = (0.8, 1.2)$. For the basis of the value function approximation, modified Chebyshev polynomials [65] up to second order have been selected. The proposed algorithm was run for 15 iterations, using solely importance sampling. The use of sample trajectory blending was not necessary for the convergence of the algorithm ($\gamma = 0$ in Algorithm 1). Figure 7.(a) 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 7.(b) depicts the cost mean ± 3 standard deviations per iterations of the algorithm. Lastly, Figure 7.(c) shows the corresponding controls for these mean trajectories in various colors, each color representing an algorithm iteration. The control that corresponds to the final algorithm iteration is marked in black and illustrates that the optimal control sequence $\{-1, 0, +1\}$ is indeed finally recovered.

We now compare the performance of the proposed stochastic control law against the deterministic control law (75), if both laws are applied in a system influenced by



(a) Deterministic, open loop.

(b) Deterministic, closed loop (by recalculating the control at each time step).

(c) Stochastic feedback control resulting from the proposed algorithm.

Figure 8: Comparison between the deterministic control law (75), applied in open loop (a) and closed loop (b) fashion, as well as the stochastic feedback control resulting from the proposed algorithm.

noise. Specifically, for the same noise profile, we used the three following approaches:

- application of the deterministic control law (75), calculated once (at the initial condition) and applied in an open-loop fashion (D-OL),
- the same control law, applied in a feedback fashion, in which for each time instant and state (t_i, x_i) of the sampled trajectories, the controls are recalculated[2] using the current state as initial condition and $T - t_i$ as a new fixed final time (D-CL),
- the proposed stochastic feedback control law, obtained by our algorithm (S-CL).

The results of each approach are depicted in Figure 8(a), (b), and (c), respectively. As expected, in D-OL, noise results in large variation between trajectories, many of which fail to reach the goal state. Performance is improved in the case of D-CL, as the deterministic controls are recalculated at each iteration, however the improvement is rather minor. This is because in D-CL, even though the control law is applied in a feedback fashion, it does not account for the noise, and thus the resulting trajectories

[2]Note that the control law in (75) is valid for initial conditions in the upper right quadrant. See [4] for more details.

are allowed to drift to areas of the state space for which a new fixed final time $T - t_i$ no longer guarantees existence of a solution that leads to the goal state. The S-CL law obtained by the proposed algorithm does not seem to suffer from this phenomenon. A comparison of the cost mean and variance of these three approaches is shown in Figure 9. Specifically, D-OL, D-CL and S-CL result in a cost mean of 5.36, 4.75 and 2.97 respectively, and a cost variance of 14.00, 2.49 and 0.07 respectively. Note that here we evaluate the cost given by equation (78) for all approaches. In the deterministic setting, and in presence of the fixed final state conditions, the two costs (71) and (78) are equivalent.

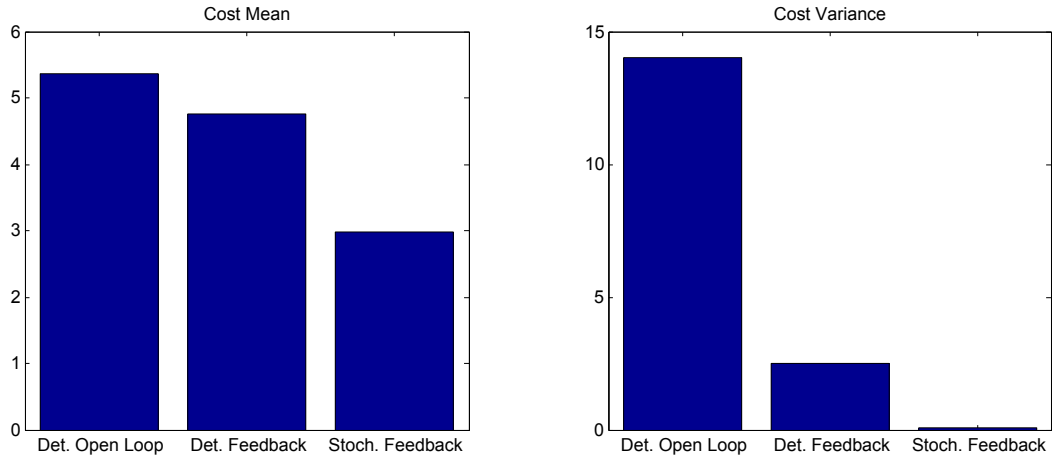


Figure 9: Cost comparison between the deterministic open loop bang-bang control law (75) used in open loop, in closed loop, and the stochastic feedback bang-bang control of the proposed algorithm. Cost mean (left) and variance (right).

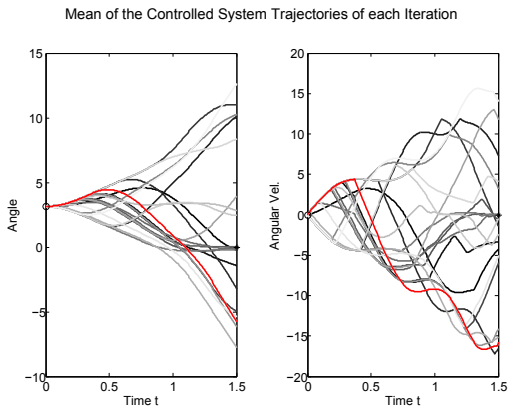
6.3.2 The Inverted Pendulum

The equations of motion for the inverted pendulum are given by

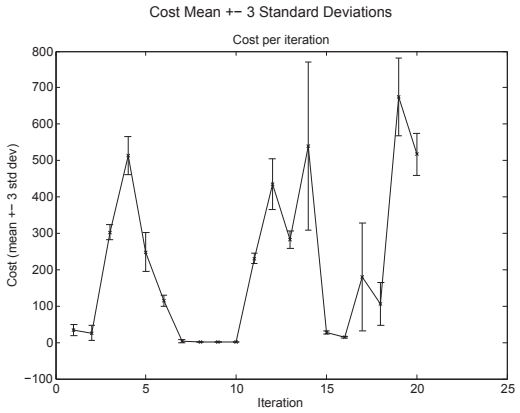
$$m\ell^2\ddot{\theta} + b\dot{\theta} - mg\ell \sin\theta = u, \tag{79}$$

and stochasticity enters the system in form of perturbations in the torque u. The constraint $u^{\max} = u^{\min} = mg\ell$ makes this problem nontrivial, since the controller

is forced to generate enough momentum by swinging back and forth to successfully invert the pendulum. For the purposes of this simulation, two thousand trajectories were generated on a time grid of $\Delta t = 0.005$ with time horizon $T = 1.5$. The blending ratio was set to $\gamma = 0.98$, meaning that in each iteration, 2% of the least favorable sample trajectories and associated control inputs are discarded from the pools in favor of newly sampled ones. The system noise covariance was set to 0.1. No initial guess for the control input was necessary. For the basis of the value function approximation, modified Chebyshev polynomials [65] up to second order have been selected. Figure 10 depicts an unsuccessful attempt of the algorithm to invert and stabilize the pendulum, in the absence of sample trajectory blending ($\gamma = 0$). We observe that the mean of the controlled system trajectories fluctuates between iterations (Figure 10.a), and thus no convergence to an optimal trajectory is achieved. The same behavior is reflected in the cost (Figure 10.b). In contrast, for $\gamma = 0.98$, the task is achieved after approximately 55 iterations with minor improvements thereafter, as shown in Figure 11. These results highlight the importance of sample trajectory blending as a technique to smoothen changes in the optimal control between successive algorithm iterations.

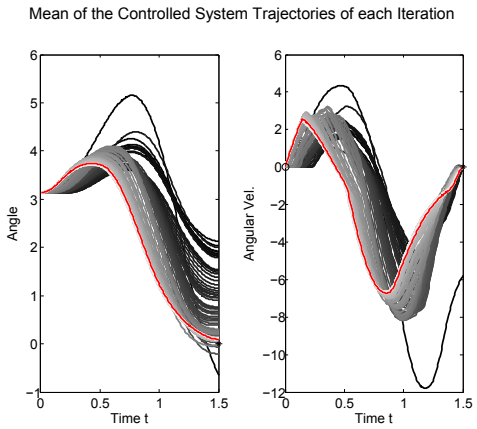


(a) The mean of the controlled system trajectories of each iteration (grayscale) and after the final iteration (red).

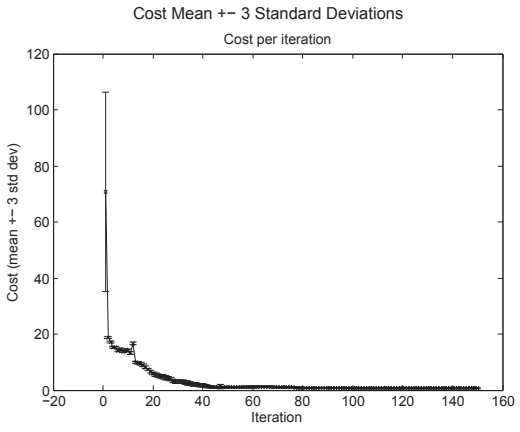


(b) Cost mean ± 3 standard deviations per iteration.

Figure 10: The inverted pendulum system: Inability of the algorithm to converge in the absence of sample trajectory blending.



(a) The mean of the controlled system trajectories of each iteration (grayscale) and after the final iteration (red).



(b) Cost mean ± 3 standard deviations per iteration.

Figure 11: The inverted pendulum system: The algorithm converges for a blending ratio of $\gamma = 0.98$

VII

STOCHASTIC DIFFERENTIAL GAMES AND RISK-SENSITIVE CONTROL

The aim of this chapter is to demonstrate that framework developed in this dissertation can be employed in the solution of a variety of classes of stochastic differential game problems as well. Specifically, we show that the Hamilton-Jacobi-Isaacs PDEs, corresponding to $\mathcal{L}^2$- or $\mathcal{L}^1$-type of control penalties for the players, assume simplified expressions under affine dynamics. Furthermore, an extension of the decomposability condition of Chapter 3 is enough to allow for a probabilistic representation of the solutions to these HJI PDEs via FBSDEs. Finally, we note that since the simplified HJI PDE that appears for the $\mathcal{L}^2$-case of stochastic differential games exhibits the same form as the HJB PDE of a risk-sensitive optimal control problem, the herein proposed scheme is applicable to this type of stochastic optimal control as well. The chapter is concluded with simulations.

7.1 Game Formulation

On the filtered probability space $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P})$, consider a differential game in which the expected game payoff is defined by the functional

$$P(u(\cdot), v(\cdot);\ \tau, x_\tau) = \mathbb{E}\big[g(x(T)) + \int_\tau^T q(t, x(t)) + L^u(u(t)) - L^v(v(t))\mathrm{d}t\big], \qquad (80)$$

where $T > \tau \geq 0$, T is a fixed time of termination[1], and $x \in \mathbb{R}^n$ represents the game state vector. The minimizing player seeks to minimize the payoff by controlling the

[1]Games in which the duration is not fixed a priori but instead involve a terminal surface will be addressed in Chapter 8.

vector $u \in \mathcal{U} \subset \mathbb{R}^\nu$, while the maximizing player seeks to maximize the payoff 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:

$$\begin{aligned} \mathcal{L}^2: &\qquad L(s) = \frac{1}{2} s^\top R s, \\ \mathcal{L}^1: &\qquad L(s) = p^\top |s|, \end{aligned}$$

where R is a positive definite real-valued matrix, p a vector of positive weights and $|\cdot|$ represents 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)

$$\begin{cases} \mathrm{d}x(t) = f(t, x(t))\mathrm{d}t + G(t, x(t))u(t)\mathrm{d}t + B(t, x(t))v(t)\mathrm{d}t + \Sigma(t, x(t))\mathrm{d}W_t, \\ t \in [\tau, T], \qquad x(\tau) = x_\tau, \end{cases} \tag{81}$$

in which $\mathrm{d}W_t$ are standard Brownian motion increments. We assume that all standard technical conditions which pertain to the filtered probability space and the regularity of functions are met, in order to guarantee existence, uniqueness of solutions to (81), and a well defined payoff (80) (see Section 2.3.2) Furthermore, the square-integrable 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 $\{\mathscr{F}_t\}_{t \geq 0}$-adapted, which essentially translates into the control inputs being non-anticipating, i.e., relying only on past and present information. If the control penalty for the maximizing or minimizing player is of the $\mathcal{L}^2$-type, then $\mathcal{U}$ and/or $\mathcal{V}$ may be open 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 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.

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. Formally, for any given initial condition (τ, x_τ), we investigate the game of conflicting control actions u, v that minimize (80) under all admissible non-anticipating strategies assigned to $u(\cdot)$, while maximizing it over all admissible non-anticipating strategies assigned to $v(\cdot)$. The structure of this problem, due to the form of the dynamics and cost at hand, satisfies the Isaacs condition[2] [39, 55, 107], and the payoff is a saddlepoint solution to the following terminal value problem of a second order partial differential equation, known as the Hamilton-Jacobi-Isaacs (HJI) equation

$$\begin{cases} V_t + \inf\limits_{u \in \mathcal{U}} \sup\limits_{v \in \mathcal{V}} \left\{ \frac{1}{2} \mathrm{tr}(V_{xx} \Sigma \Sigma^\top) + V_x^\top (f + Gu + Bv) + q + L^u(u) - L^v(v) \right\} = 0, \\ (t, x) \in [0, T) \times \mathbb{R}^n, \qquad V(T, x) = g(x), \quad x \in \mathbb{R}^n. \end{cases} \tag{82}$$

In the above, function arguments have been suppressed for notational compactness,

[2] The Isaacs condition renders the viscosity solutions of the upper and lower value functions equal (see [40]), thus making the order of maximization/minimization inconsequential.

and V_x and V_{xx} denote the gradient and the Hessian of V, respectively. The term inside 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.

7.1.1 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 (82) 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

$$u^*(t,x) = -R_u^{-1}G^\top(t,x)V_x(t,x), \tag{83}$$

$$v^*(t,x) = R_v^{-1}B^\top(t,x)V_x(t,x). \tag{84}$$

Inserting the above expression back into the HJI equation (82) and suppressing function arguments for notational brevity, we obtain the equivalent characterization

$$\begin{cases} V_t + \frac{1}{2}\mathrm{tr}(V_{xx}\Sigma\Sigma^\top) + V_x^\top f + q - \frac{1}{2}V_x^\top\left(GR_u^{-1}G^\top - BR_v^{-1}B^\top\right)V_x = 0, \\ (t,x) \in [0,T) \times \mathbb{R}^n, \quad V(T,x) = g(x), \quad x \in \mathbb{R}^n. \end{cases} \tag{85}$$

7.1.2 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, the HJI equation (82) can be written as

$$\begin{cases} V_t + \inf\limits_{u\in\mathcal{U}} \sup\limits_{v\in\mathcal{V}} \Big\{ \frac{1}{2}\mathrm{tr}(V_{xx}\Sigma\Sigma^\top) + V_x^\top f + \big(V_x^\top G + p_u^\top \mathrm{D}(\mathrm{sgn}(u))\big)u \\ \qquad\qquad + \big(V_x^\top B - p_v^\top \mathrm{D}(\mathrm{sgn}(v))\big)v + q_0 \Big\} = 0, \quad (t,x) \in [0,T) \times \mathbb{R}^n, \\ V(T,x) = g(x), \quad x \in \mathbb{R}^n, \end{cases} \tag{86}$$

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 by performing the same analysis as in Chapter 6, to obtain the optimal control law for the minimizing player:

$$u_i^* = \begin{cases} u_i^{\max}, & (V_x^\top G)_i < -(p_u^\top)_i \\ -u_i^{\min}, & (V_x^\top G)_i > (p_u^\top)_i, \\ 0, & -(p_u^\top)_i < (V_x^\top G)_i < (p_u^\top)_i, \end{cases} \qquad i = 1, \cdots, \nu, \tag{87}$$

while the optimal control law for the maximizing player reads:

$$v_i^* = \begin{cases} v_i^{\max}, & (V_x^\top B)_i > (p_v^\top)_i \\ -v_i^{\min}, & (V_x^\top B)_i < -(p_v^\top)_i, \\ 0, & -(p_v^\top)_i < (V_x^\top B)_i < (p_v^\top)_i. \end{cases} \qquad i = 1, \cdots, \mu, \tag{88}$$

Remark 7.1. *We note again that, as in Chapter 6, the control laws given by (87)-(88) 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 (86). A problem in which either one of these equalities is satisfied over a nontrivial time interval is a singular fuel-optimal problem [4]. In this work, we shall assume that the problem is normal, in the sense that the aforementioned equalities are not satisfied over a nontrivial time interval.

We may insert the optimal control laws (87)-(88) back into the HJI equation (86), to obtain the equivalent expression

$$\begin{cases} V_t + \frac{1}{2}\mathrm{tr}(V_{xx}\Sigma\Sigma^\top) + V_x^\top f + q \\ \qquad\qquad + \sum_{i=1}^{\nu} \min\left\{ \left(V_x^\top G + p_u^\top\right)_i u_i^{\max},\ 0, -\left(V_x^\top G - p_u^\top\right)_i u_i^{\min} \right\} \\ \qquad\qquad + \sum_{i=1}^{\mu} \max\left\{ \left(V_x^\top B - p_v^\top\right)_i v_i^{\max},\ 0, -\left(V_x^\top B + p_v^\top\right)_i v_i^{\min} \right\} = 0, \\ (t,x) \in [0,T) \times \mathbb{R}^n, \qquad V(T,x) = g(x), \quad x \in \mathbb{R}^n, \end{cases} \tag{89}$$

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

7.1.3 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 the control laws (83) and (88) for the minimizing and maximizing player, respectively, while the HJI equation would assume the form

$$\begin{cases} V_t + \frac{1}{2}\mathrm{tr}(V_{xx}\Sigma\Sigma^\top) + V_x^\top f + q - \frac{1}{2}V_x^\top G R_u^{-1} G^\top V_x \\ \qquad + \sum_{i=1}^{\mu} \max\left\{ \left(V_x^\top B - p_v^\top\right)_i v_i^{\max},\ 0, -\left(V_x^\top B + p_v^\top\right)_i v_i^{\min} \right\} = 0, \qquad (90) \\ (t,x) \in [0,T) \times \mathbb{R}^n, \qquad V(T,x) = g(x), \quad x \in \mathbb{R}^n. \end{cases}$$

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

7.2 A Feynman-Kac type Representation

By comparing the PDEs in Sections 7.1.1, 7.1.2 and 7.1.3 with the Cauchy problem (13), we may conclude that the nonlinear Feynman-Kac representation can be applied to each HJI equation of these sections under an extension of the decomposability condition of Assumption 3.1 in Section 3.2, stated as follows:

Assumption 7.1. *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* $(t,x) \in [0,T] \times \mathbb{R}^n$.

Similarly to Assumption 3.1, Assumption 7.1 implies that the ranges of G and L must be a subset of the range of Σ, and thus excludes the case of a channel containing control input but no noise, although the converse is allowed. Under Assumption 7.1, the HJI equations in Sections 7.1.1, 7.1.2 and 7.1.3 (equations (85), (89) and (90), respectively) satisfy the Cauchy problem (13) standard form. We readily obtain the following SDE coefficients $b(\cdot)$ and $h(\cdot)$:

$$b(t,x) \equiv f(t,x) \qquad (91)$$

and

$$\text{Case I}: \quad h(t,x,z) \equiv q - \frac{1}{2} z^\top \left(\Gamma R_u^{-1} \Gamma^\top - \Lambda R_v^{-1} \Lambda^\top \right) z, \tag{92}$$

$$\begin{aligned} \text{Case II}: \quad h(t,x,z) \equiv q &+ \sum_{i=1}^{\nu} \min \left\{ \left(z^\top \Gamma + p_u^\top \right)_i u_i^{\max},\ 0,\ -\left(z^\top \Gamma - p_u^\top \right)_i u_i^{\min} \right\} \\ &+ \sum_{i=1}^{\mu} \max \left\{ \left(z^\top \Lambda - p_v^\top \right)_i v_i^{\max},\ 0, -\left(z^\top \Lambda + p_v^\top \right)_i v_i^{\min} \right\}, \end{aligned} \tag{93}$$

$$\begin{aligned} \text{Case III}: \quad h(t,x,z) \equiv q &- \frac{1}{2} z^\top \Gamma R_u^{-1} \Gamma^\top z \\ &+ \sum_{i=1}^{\mu} \max \left\{ \left(z^\top \Lambda - p_v^\top \right)_i v_i^{\max},\ 0, -\left(z^\top \Lambda + p_v^\top \right)_i v_i^{\min} \right\}. \end{aligned} \tag{94}$$

The (viscosity) solution of PDEs (85), (89) or (90) are thus obtained by simulating the FBSDE systems given by (1) and (7) per definitions (91) and (92), (93) or (94), respectively. Notice that (1) corresponds again to the uncontrolled ($u = 0$, $v = 0$) system dynamics. We conclude this section by noting that the resulting FBSDE problem can be solved iteratively using the importance sampling algorithm of Section 5.2.

7.3 Connection to Risk-Sensitive Control

The connection between dynamic games and risk-sensitive stochastic control is well-documented in the literature [5, 25, 56]. Specifically, the optimal controller of a stochastic control problem with exponentiated integral cost (a so-called risk-sensitive problem) turns out to be identical to the minimizing player's unique minimax controller in a stochastic differential game setting. Indeed, consider the problem of minimizing the expected cost given by

$$J(u(\cdot); \tau, x_\tau) = \epsilon \ln \mathbb{E} \left\{ \exp \frac{1}{\epsilon} \left[g(x(T)) + \int_\tau^T q(t, x(t)) + \frac{1}{2} u(t)^\top R u(t)\ \mathrm{d}t \right] \right\}, \tag{95}$$

where ϵ is a small positive number. The state dynamics are described by the Itô SDE

$$\begin{cases} \mathrm{d}x(t) = f(t,x(t))\mathrm{d}t + G(t,x(t))u(t)\mathrm{d}t + \sqrt{\dfrac{\epsilon}{2\gamma^2}}\tilde{\Sigma}(t,x(t))\mathrm{d}W_t, \qquad t \in [\tau, T], \\ x(\tau) = x_\tau. \end{cases} \tag{96}$$

In this setting, the name "risk-sensitive" arises because of the nature of cost (95): indeed, performing a Taylor series expansion, one obtains

$$J = \mathbb{E}[J_0] + \frac{1}{2\epsilon}\mathrm{var}[J_0] + \ldots,$$

in which J_0 denotes the quantity inside the brackets $[\cdot]$ in (95). Thus, this selection of cost penalizes both the mean and the variance of J_0. Suppressing function arguments for notational compactness, the associated Hamilton-Jacobi-Bellman PDE for this problem is [5]

$$\begin{cases} V_t + \inf\limits_{u\in\mathcal{U}}\left\{\dfrac{\epsilon}{4\gamma^2}\mathrm{tr}(V_{xx}\tilde{\Sigma}\tilde{\Sigma}^\top) + V_x^\top(f+Gu) + q + \dfrac{1}{2}u^\top R u + \dfrac{1}{4\gamma^2}V_x^\top\tilde{\Sigma}\tilde{\Sigma}^\top V_x\right\} = 0, \\ (t,x) \in [0,T)\times\mathbb{R}^n, \qquad V(T,x) = g(x), \quad x\in\mathbb{R}^n. \end{cases} \tag{97}$$

The infimum operation can be performed explicitly, and yields the optimal control $u^*(t,x) = -R^{-1}G^\top(t,x)V_x(t,x)$. Setting $\Sigma = \sqrt{\epsilon/2\gamma^2}\tilde{\Sigma}$ and substituting the optimal control in the PDE (97) we readily obtain the equivalent characterization

$$\begin{cases} V_t + \dfrac{1}{2}\mathrm{tr}(V_{xx}\Sigma\Sigma^\top) + V_x^\top f + q - \dfrac{1}{2}V_x^\top\left(GR^{-1}G^\top - \dfrac{1}{\epsilon}\Sigma\Sigma^\top\right)V_x = 0, \\ (t,x)\in[0,T)\times\mathbb{R}^n, \quad V(T,x) = g(x), \quad x\in\mathbb{R}^n. \end{cases} \tag{98}$$

The above equation is merely a special case of equation (82) obtained for the game-theoretic version, if one substitutes $R_v = (1/\epsilon)I$ and $B = \Sigma$. Notice that this special case of B automatically satisfies Assumption 7.1 with Λ being the identity matrix.

Thus, imposing the same decomposability condition on G, the solution to the risk-sensitive stochastic optimal control problem can be obtained by simulating the system of FBSDEs given by (1) and (7) using the definitions (91) and (92).

7.4 Simulations

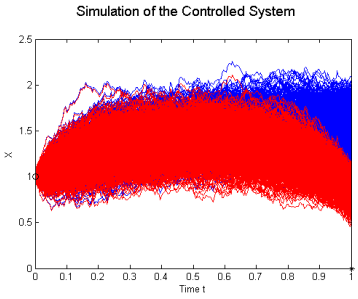
As a proof of concept, we simulated the algorithm for two different cases of differential games: a scalar system with nonlinear drift dynamics, and a game based on the single integrator problem in the Simulations section of Chapter 6.

7.4.1 A Scalar Example

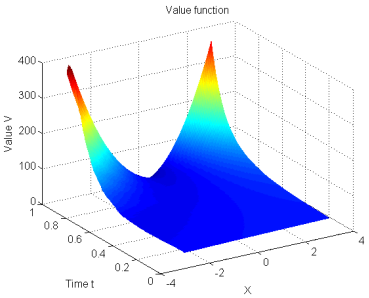
We consider the scalar system with nonlinear drift $\mathrm{d}x = (4\cos x + u + 0.5xv)\mathrm{d}t + 0.5\mathrm{d}w$, setting $q(t,x) = 0$, $R_u = 2$, $R_v = 5$, $x(0) = 1$, $T = 1$ and $g(x_T) = 40x_T^2$, thus penalizing deviation from the origin at the time of termination, T. Two thousand trajectories were generated on a time grid of $\Delta t = 0.005$, while the set of basis functions for Y was selected to be $[1\ x\ x^2]^\top$. The results are depicted in Fig. 12. From the shape of the value function in Fig. 12(b) it is seen that the value is relatively flat at the beginning since there is no state-dependent running cost and becomes progressively quadratic at the final time owing to the boundary condition $V(T, x_T) = 40x_T^2$. Note, however, that Fig. 12(b) shows the value function over a rectangular grid. In fact, we have an accurate estimate of the value function only over the area of the state space visited by the sampled (open-loop) trajectories. In that sense, the areas not visited by the system are extrapolated based on the basis functions chosen to represent V.

7.4.2 A Single Integrator Game with Mixed Types of Penalties

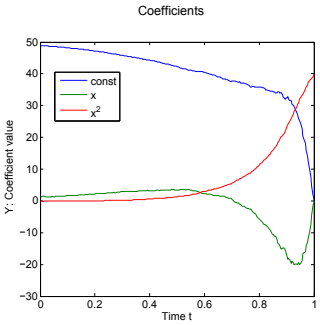
We consider a stochastic differential game based on the single integrator optimal control problem of the Simulations section of Chapter 6. Here, the minimizing player has a control restricted to $|u(t)| \leq 1$ and pays an $\mathcal{L}^1$ penalty, while the maximizing



(a) Trajectories of the uncontrolled ($u = v = 0$) system (blue), and the optimally controlled system (red).



(b) The Value function.



(c) Coefficients of the basis functions used to decompose the Value function.

Figure 12: Simulation for a scalar nonlinear differential game: controlled and uncontrolled system trajectories, the value function, and the coefficients of its decomposition.

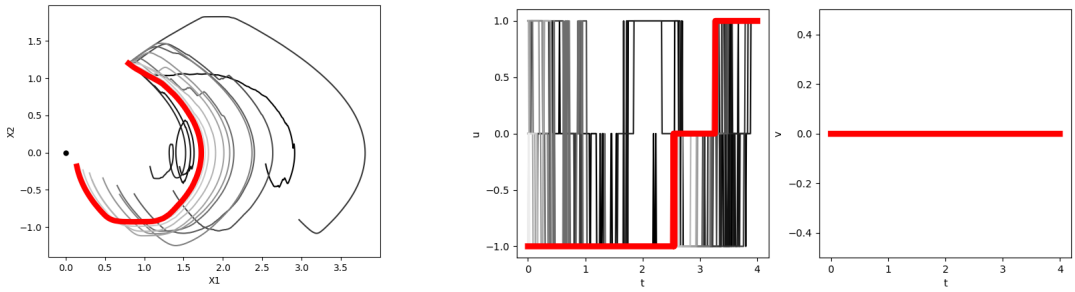
player has no control constraints and pays an $\mathcal{L}^2$ penalty. The dynamics are given by

$$\mathrm{d}x_1(t) = x_2(t)\,\mathrm{d}t, \quad \mathrm{d}x_2(t) = (u(t) + \beta\, v(t))\,\mathrm{d}t + \sigma\,\mathrm{d}w(t), \quad |u(t)| \leq 1, \tag{99}$$

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:

$$P = \mathbb{E}\left[10\,(x_1^2(T) + x_2^2(T)) + \int_0^T |u(t)| - 2.5\,v^2(t)\,\mathrm{d}t\right]. \tag{100}$$

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 executed 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 of the single integrator optimal control problem in the Simulations section of Chapter 6. Indeed, as shown in Figure 13.(b), this optimal control sequence $\{-1, 0, +1\}$ is recovered. Increasing

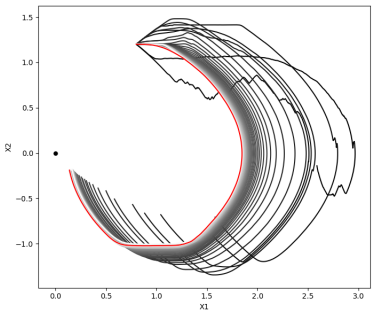


(a) Mean system trajectories.

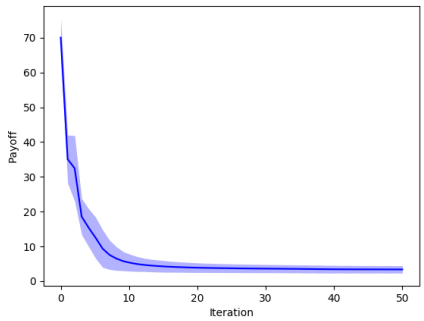
(b) Control.

Figure 13: 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 (coloured) and after the final iteration (black). We see that the optimal minimizing control sequence $\{-1, 0, +1\}$ is finally recovered.

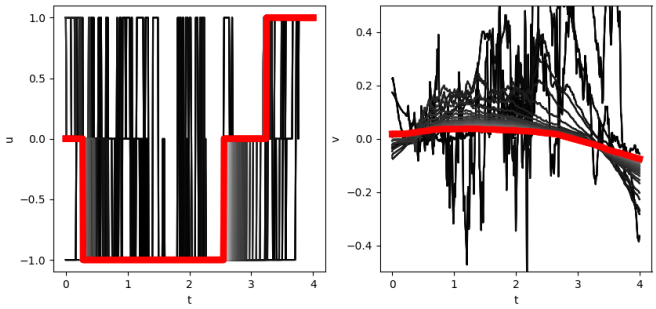
the value of β to 0.1, Figure 14.(a) depicts the mean of the controlled trajectories in phase-plane after each iteration of the algorithm (gray scale). The trajectory that corresponds to the final iteration is marked in red. Figure 14.(b) depicts the payoff mean $\pm$ 3 standard deviations per iteration of the algorithm. Interestingly enough, the optimal minimizing control now differs, as shown in Figure 14.(c).



(a) Mean system trajectories.



(b) Cost



(c) Control.

Figure 14: 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 (coloured) and after the final iteration (black). We see that the optimal minimizing control sequence has now changed.

VIII

FIRST EXIT FORMULATIONS

All optimal control problems formulated thus far require the selection of an appropriate value for the time of termination. In several situations however, choosing such a value can be challenging and may introduce unnecessary restrictions in the class of problem solutions. This can be illustrated by means of the following example: consider the case in which a system is expected to accomplish a specific task (e.g., inverting a pendulum, achieving a desired configuration for a robotic arm etc.). The presence of a fixed final time implies the constraint that this task needs to be accomplished at *exactly* time T, thereby effectively rejecting control solutions that accomplish it at a different time, sooner or later. Unless timing is a critical factor in control design, the particular time instant in which the task is completed is often unimportant. In other words, the class of control solutions within which we seek the optimum is restricted to those solutions that satisfy the terminal time requirement, and may greatly influence the resulting cost, without being an important control design factor. This implies that there is an additional dimension with respect to which the performance index can be optimized, namely the time of termination, a fact which has been addressed in classic optimal control theory by considering free final time formulations.

Similarly, in the context of differential games, the game formulations presented in Chapter 7 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. The set of states signaling game termination are called a *terminal surface* in the literature of differential games.

In a stochastic setting, free final time problems without cost discounting can be troublesome due to the absence of boundedness guarantees. Furthermore, since the presented approach is a sampling-based method, allowing the process to continue without imposing an upper bound on its duration may yield trajectory samples that have a very large – or even possibly infinite – time duration, and thus cannot be simulated (see relevant discussion in [37]). However, we may formulate a *first exit* problem with time upper bound, in which the process terminates as soon as the task has been achieved, *or* a specified maximum time duration has passed, whichever event occurs first. Thus, this formulation enables optimization with respect to time as well. Intuitively, if the optimal mean final time is finite and the upper bound is large enough, this aforementioned formulation should yield that optimal final time (in expectation). This extension also allows us to address differential games involving a terminal surface.

8.1 Problem Statement

Let $\mathcal{G}$ be the domain of the drift-diffusion process within the state space, and let $\partial\mathcal{G} \in C^1$ be its boundary, the crossing of which signals early process termination, i.e., $\partial\mathcal{G}$ separates the target set from the rest of the domain (in the case of differential games, $\partial\mathcal{G}$ represents the terminal surface). Given $(\Omega, \mathscr{F}, \{\mathscr{F}_s\}_{s\geq 0}, \mathbb{P})$, we may define the cost

$$J(u(\cdot); x_0, T) = \mathbb{E}\big[\Psi(\mathcal{T}, x(\mathcal{T})) + \int_0^{\mathcal{T}} L(t, x(t), u(t))\mathrm{d}t\big], \tag{101}$$

in which L is either an $\mathcal{L}^2$ or $\mathcal{L}^1$-type running cost, as described in Chapters 3 or 6, respectively (see equations (16) and (61)), and $\mathcal{T}$ and $\Psi(\cdot)$ are defined as follows:

$$\mathcal{T} \triangleq \min\{\tau_{\text{exit}}, T\}, \quad \text{with} \quad \tau_{\text{exit}} \triangleq \inf\{s \in [0, T] \ : x(s) \in \partial\mathcal{G}\}, \tag{102}$$

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

$$\Psi(t,x) \triangleq \begin{cases} g(x), & (t,x) \in \{T\} \times \mathcal{G}, \\ \psi(t,x), & (t,x) \in [0,T) \times \partial\mathcal{G}. \end{cases} \tag{103}$$

Here, $g(\cdot)$ is the usual fixed final time terminal cost, while $\psi(\cdot)$ is a function assigning a terminal cost for time instants $t < T$, whenever the trajectories hit the target set before the maximum time of termination has elapsed. Following a similar procedure as in Sections 3.2 or 6.2, and under Assumption 3.1, the resulting HJB PDE is [39]

$$\begin{cases} v_t + \frac{1}{2}\text{tr}(v_{xx}\Sigma(t,x)\Sigma^\top(t,x)) + v_x^\top b(t,x) + h(t,x,\Sigma^\top(t,x)v_x) = 0, & (t,x) \in [0,T) \times \mathcal{G}, \\ v(T,x) = g(x), & x \in \mathcal{G}, \\ v(t,x) = \psi(t,x), & (t,x) \in [0,T) \times \partial\mathcal{G} \end{cases} \tag{104}$$

in which $b(\cdot)$ and $h(\cdot)$ may take any of the forms given by equations (25) - (26), or (67) - (68), depending on whether the running cost in (101) is of the $\mathcal{L}^2$ or $\mathcal{L}^1$ type, respectively. In the context of differential games, the associated $b(\cdot)$ and $h(\cdot)$ are given by (91) and (92), (93) or (94). The corresponding FBSDEs that yield a probabilistic solution to this problem are [138]

$$\begin{cases} \mathrm{d}X_s = b(s,X_s)\mathrm{d}s + \Sigma(s,X_s)\mathrm{d}W_s, & s \in [t,\mathcal{T}], \\ X_t = x. \end{cases} \tag{105}$$

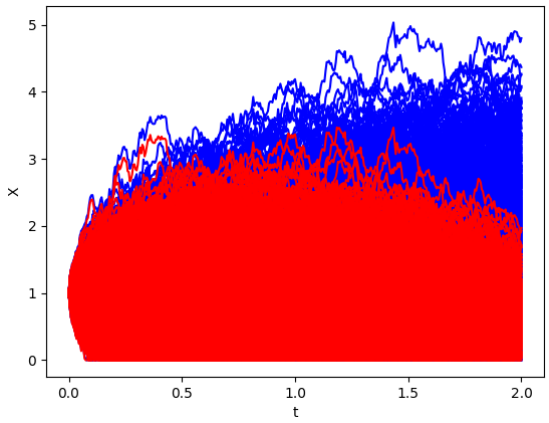
and

$$\begin{cases} \mathrm{d}Y_s = -h(s,X_s,Z_s)\mathrm{d}s + Z_s^\top \mathrm{d}W_s, & s \in [t,\mathcal{T}], \\ Y_\mathcal{T} = \Psi(X_\mathcal{T}). \end{cases} \tag{106}$$

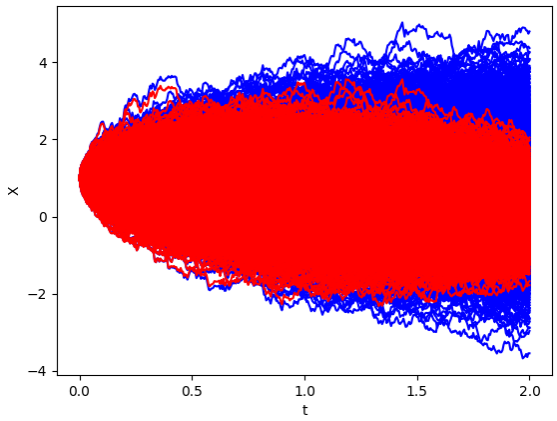
The same procedure can be applied to a differential game setting. We conclude this section by noting that the resulting FBSDE problem can be solved iteratively using the importance sampling algorithm of Section 5.2.

8.2 Simulations

To illustrate the theory, we simulated the algorithm for the scalar system $\mathrm{d}x = (-0.2x + 0.5u)\mathrm{d}t + 0.5\mathrm{d}w$. With an initial condition $x(0) = 1$, we assume that the target is $x = 0$, and therefore terminate the process early, once the origin is crossed, without any penalty ($\psi(t, x) = 0$). After a maximum duration of $T = 2$ has passed, we terminate the process and penalize deviation from the origin at that time instant using $g(x_T) = 5x_T^2$. Furthermore, we set $q(t, x) = 0$, and $R = 1$. For comparison, we also simulate the same system without the presence of a terminal surface. For the purposes of simulation, two thousand trajectories were generated on a time grid of $\Delta t = 0.005$, while the set of basis functions for Y was selected to be $[1\ x\ x^2]^\top$. The results are depicted in Fig. 15. The cost mean and variance for the first exit problem is 1.7 and 10.7 respectively, while for the fixed final time problem the cost mean and variance are 3.5 and 14.7 respectively, indicating that there is a significant decrease in the cost if we relax the requirement of a fixed final time for the task to be accomplished.



(a) Trajectories of the uncontrolled ($u = 0$) system (blue), and the optimally controlled system (red). The process may terminate early once the goal ($x = 0$) has been achieved. Cost mean and variance are 1.7 and 10.7 respectively.



(b) Trajectories of the uncontrolled ($u = 0$) system (blue), and the optimally controlled system (red). The process terminates only when $T = 2$. Cost mean and variance are 3.5 and 14.7 respectively.

Figure 15: System trajectories: (a) with early termination at the target $x = 0$, and (b) with fixed time of termination T.

IX

APPLICATION: THE SOFT LANDING PROBLEM

In this chapter, we will apply the proposed algorithm to the *soft landing problem* (SLP). Therein, the goal is to find the optimal control, i.e., the optimal thrust profile, for a spacecraft attempting to make a soft landing on a planet, using a minimum amount of fuel. The problem was originally addressed by considering only one spatial dimension (namely the altitude with respect to the planet), in which case its deterministic formulation offers a closed-form solution (initially obtained by Miele [88, 89] during the 1960's, see also [38, 87]). In more recent years, there has been renewed interest in the topic, which appears under the name *Powered-Descent Guidance* (PDG), mainly due to the success of NASA's Mars Science Laboratory program. Several results appear in the literature, treating a more complex problem involving all three spatial dimensions, more accurate modeling of the dynamics to account for planetary rotation, and several state and control constraints [1, 31, 116]. The challenges faced in the implementation of planetary PDG controllers are the twofold: (a) the environmental uncertainty and stochastic disturbances present, and (b) the limited capabilities for onboard computation.

In this dissertation, we address both of these issues by an application of the proposed algorithm on the $\mathcal{L}^1$-optimal SLP. We shall demonstrate that the algorithm offers superior performance in the presence of stochastic disturbances, compared to both an open-loop, as well as a closed-loop implementation of the deterministic solution, offering a much lower mean and variance on the touchdown speed. Depending on given safety specifications, we can further reduce this mean and variance, thus

gaining a more robust, safer controller, at the expense of slightly increased fuel expenditure. Furthermore, the nature of the algorithm allows for a complete solution of the problem a priori and off-line, thus minimizing the required onboard computing capabilities of the spacecraft.

9.1 Problem Description

In this section, we formally define the SLP. We first introduce the deterministic setting and its closed-form solution, which we will later use for validation and comparison purposes. We then present a stochastic version of the problem, on which we will apply the proposed algorithm. Finally, in the simulation section, we compare the numerical results obtained by the proposed algorithm to those of the closed-form solution (both in open-loop and closed-loop implementation).

9.1.1 Deterministic Setting

Consider the problem of a spacecraft attempting to make a soft landing on a planet, using a minimum amount of fuel. The dynamical equations are given by

$$\begin{aligned}
\dot{h}(t) &= v(t), \\
\dot{v}(t) &= -g + \frac{u(t)}{m(t)}, \qquad u(t) \in [u_{\min}, u_{\max}], \\
\dot{m}(t) &= -\alpha u(t), \\
t \in &[0, t_f], \quad h(0) = h_0, \quad v(0) = v_0, \quad m(0) = m_0,
\end{aligned}$$

wherein $h : [0, t_f] \to \mathbb{R}_+$, $v : [0, t_f] \to \mathbb{R}$, and $m : [0, t_f] \to \mathbb{R}_+$ denote the altitude, vertical speed, and mass of the spacecraft at time t, respectively, g is the gravitational acceleration, assumed to be constant, α is a positive constant that describes the mass flow rate, and $u(t) : [0, t_f] \to [u_{\min}, u_{\max}]$, is the control input (thrust), with $u_{\min}, u_{\max} \in \mathbb{R}_+$. As admissible controls, we consider all piecewise continuous control

functions taking values in the aforementioned interval. The initial conditions are (h_0, v_0, m_0), whereas the terminal conditions are $h(t_f) = v(t_f) = 0$. Here, t_f denotes the time instant of landing, whose particular value is otherwise left unspecified. For the mass, we assume that a reasonable value has been assigned to m_0 so that landing with remaining mass at or above the dry mass (mass of the spaceship without fuel) is feasible. We wish to obtain the optimal control $u^*(t)$ that satisfies the above conditions, while minimizing the amount of fuel spent:

$$J_{\text{det}}(u(\cdot); h_0, v_0, m_0) = \int_0^{t_f} |u(t)| \mathrm{d}t. \tag{107}$$

It can be shown [38,87,88] that the solution to this $\mathcal{L}^1$-optimal control problem of free final time yields a unique optimal *bang-bang* controller, and that the problem is *normal* (meaning singular control does not appear within the optimal control sequence), and that there is at most one switch time. The optimal control sequence is

$$u^*(t) = \begin{cases} u_{\min}, & t \in [0, t_s), \\ u_{\max}, & t \in [t_s, t_f], \end{cases} \tag{108}$$

in which t_s denotes the switching time. It can be shown (see Appendix A) that the switching and final time satisfy the following system of equations:

$$\begin{aligned} h_0 + v_0 t_s + \frac{t_f}{\alpha} - \frac{1}{\alpha}(t_s - \frac{m_0}{\alpha u_{\min}})\ln(1 - \frac{\alpha u_{\min}}{m_0} t_s) - \frac{1}{2} g t_s^2 \\ + \frac{m_0 - \alpha u_{\min} t_s}{\alpha^2 u_{\max}} \ln\left(1 - \frac{\alpha u_{\max}}{m_0 - \alpha u_{\min} t_s}(t_f - t_s)\right) + \frac{1}{2} g (t_f - t_s)^2 = 0, \end{aligned} \tag{109}$$

$$\alpha(u_{\max} - u_{\min}) t_s = \alpha u_{\max} t_f + m_0 \Big(\exp(\alpha(v_0 - g t_f) - 1) \Big). \tag{110}$$

The above system can be numerically solved for t_s and t_f.

9.1.2 Stochastic Setting

We consider a stochastic version of the above problem by introducing the dynamics

$$\begin{aligned}
\mathrm{d}h(t) &= v(t)\mathrm{d}t, \\
\mathrm{d}v(t) &= (-g + \frac{u(t)}{m(t)})\mathrm{d}t + \sigma\frac{bu_{\max}}{m(t)}\mathrm{d}W_t, \qquad u(t) \in [u_{\min}, u_{\max}], \\
\mathrm{d}m(t) &= -\alpha u(t)\mathrm{d}t - \sigma\alpha b u_{\max}\mathrm{d}W_t, \\
t &\in [0, \mathcal{T}], \quad h(0) = h_0, \quad v(0) = v_0, \quad m(0) = m_0,
\end{aligned}$$

Since the time of termination, t_f is not specified a priori, we will consider a *first exit* problem, in which the process terminates when the hyperplane $h = 0$ is crossed, or an upper bound T on the time duration has passed. Thus the state space is $\mathcal{G} = \{h, v, m : h \in \mathbb{R}_+, v \in \mathbb{R}, m \in R_+\}$, with $\partial\mathcal{G} = \{h, v, m \in \mathcal{G} : h = 0\}$. Enforcing the terminal equality constraints $h(t_f) = v(t_f) = 0$ in a stochastic setting is not meaningful, since the probability of hitting those states *exactly* is zero. We shall introduce them as *soft* constraints in the cost, which we define as the following:

$$J(u(\cdot); h_0, v_0, m_0, T) = \mathbb{E}\big[\Psi(\mathcal{T}, h(\mathcal{T}), v(\mathcal{T})) + \int_0^{\mathcal{T}} q|u(t)|\mathrm{d}t\big], \tag{111}$$

with q being a positive constant, $\mathcal{T}$ the minimum between the time of first exit, τ_{exit}, and the upper bound T, and

$$\Psi(t, h, v) \triangleq \begin{cases} c_1 h^2(t) + c_2 v^2(t), & (t, h, v) \in \{T\} \times \mathcal{G}, \\ c_3 v^2(t), & (t, h, v) \in [0, T) \times \partial\mathcal{G}, \end{cases} \tag{112}$$

where in c_1, c_2, and c_3 are positive constants. The motivation behind this choice of terminal cost $\Psi(\cdot)$ is that trajectories that terminate earlier than $t = T$ because of touchdown ($h = 0$) are penalized a high touchdown speed, whereas trajectories

that terminate at $t = T$ (i.e., without a touchdown) are penalized for both residual altitude and speed.

9.2 Simulation Results

For the purposes of simulation, we assumed the following constants: (taken from [136], which investigates safe landing on Mars) $g = 3.71$, $b = 0.02$, $\alpha = 4.83E-4$, $u_{\min} =$ 4.97E3, $u_{\max} =$ 1.33E4, $\sigma = 3$, and initial conditions $(h_0, v_0, m_0) = (80, -10, 1905)$. Comparison of performance is done via two metrics, namely the touchdown speed, and the fuel mass used; in both cases, both mean and variance are calculated. Another indicator is the percentage of trajectories that lead to a touchdown.

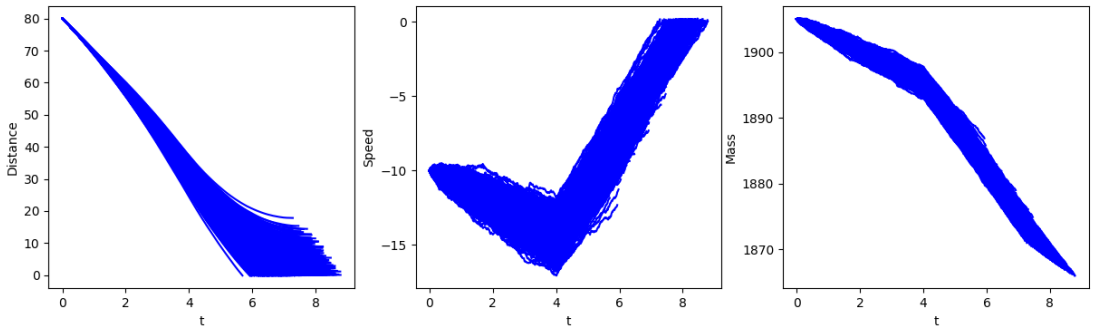
9.2.1 Deterministic Control- Open-Loop Implementation

In this case, we calculate the switching time and apply the deterministic control law (108) in an open-loop fashion. The results are depicted in Figure 16. Out of the 1000 trajectories simulated, only 50.3% lead to touchdown. The remaining trajectories lead to a hovering above the ground, which also explains the spike in fuel expenditure, seen in Figure 16(b). Of the 50.3% of the trajectories for which a touchdown occurs, most of them are considered a crash, due to the high speed impact. Indeed, the mean touchdown speed is -5.24 m/s, with a variance of 5.40.

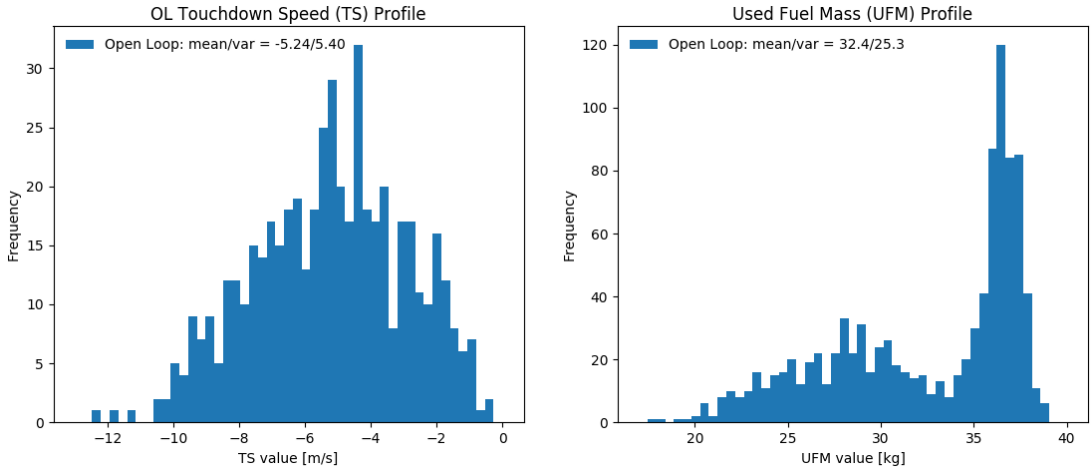
9.2.2 Deterministic Control- Closed-Loop Implementation

We now simulate the control law (108) in a closed-loop fashion, i.e., at each time instant we recalculate the switching time. Switching back and forth between controls (due to the influence of the noise) is allowed.

The results are depicted in Figure 17. All of the 1000 trajectories simulated now lead to a touchdown. However, most of them are still considered a crash, due to the high speed impact. Indeed, the mean touchdown speed this time is -3.19 m/s, with a variance of 1.96.



(a) Trajectories resulting from the open-loop implementation of control law (108). Only 50.3% of trajectories lead to touchdown, the rest ends up hovering above the ground (see left figure)).

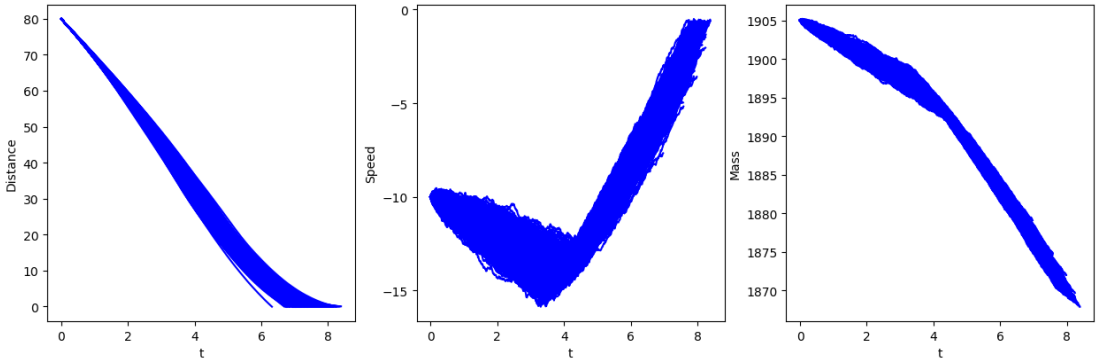


(b) Touchdown speed profile (left) and fuel consumption profile (right). Out of the 50.3% of the trajectories that lead to touchdown, most of them are considered a crash. The high fuel expense in the right is explained by the hovering above the ground.

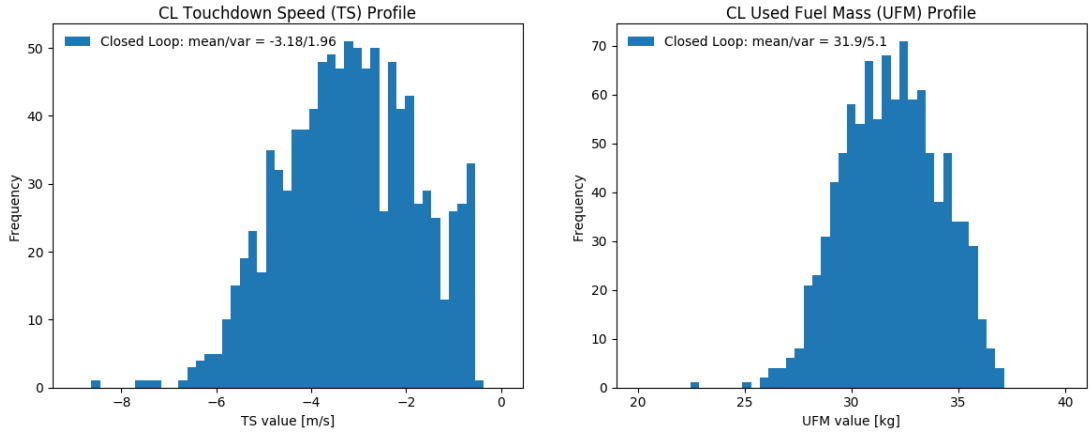
Figure 16: SLP: solution of the open-loop implementation of control law (108).

9.2.3 Proposed Algorithm

For $T = 8.5$, $q = 1$, we used three thousand trajectory samples on a time grid of $\Delta t = 0.005$, and a trajectory blending ratio of 0.98. The results are depicted in Figure 18. After the final iteration of the proposed algorithm, we evaluate the performance of the control law by simulating 1000 trajectories for time intervals long enough to achieve touchdown, see Figure 18(b). For $t > T$, we use the same control law as for $t = T$. In contrast to the deterministic setting, the cost given by (111) can be used to shape trajectories based on whether we place more importance on minimizing the touchdown speed even for worst-case disturbances (at the expense



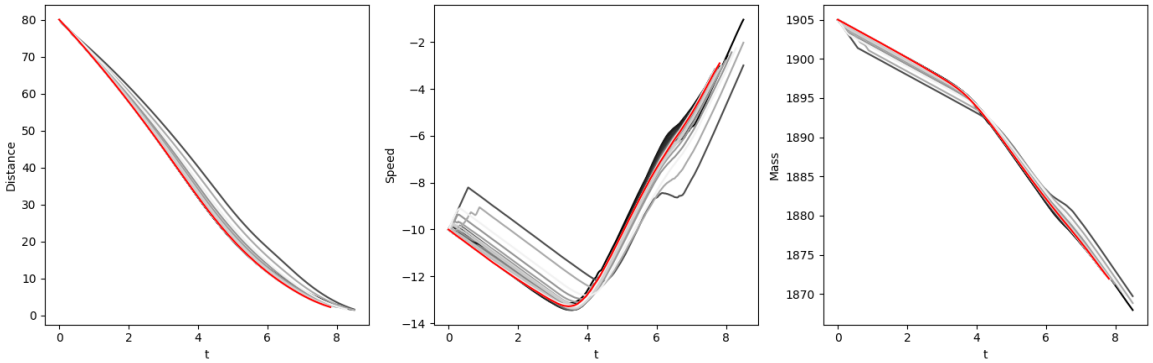
(a) Trajectories resulting from the closed-loop implementation of control law (108). 100% of trajectories lead to touchdown.



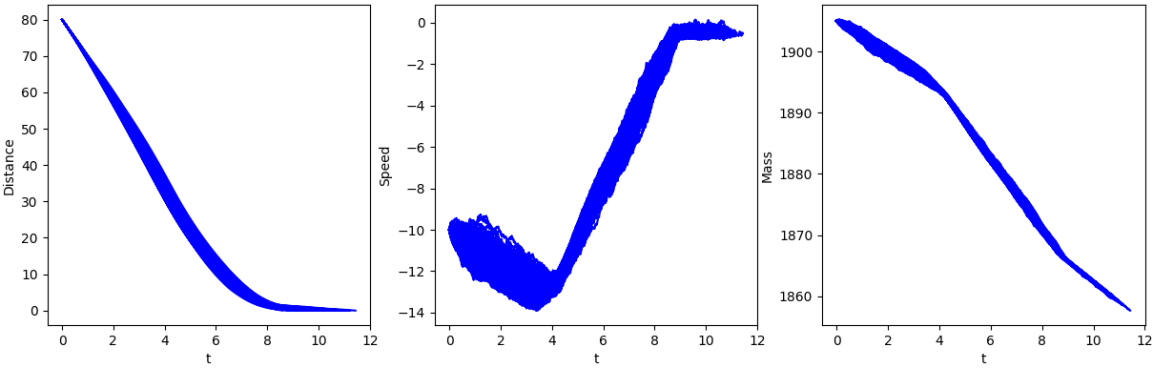
(b) Touchdown speed profile (left) and fuel consumption profile (right). Most of the trajectories still lead to a crash.

Figure 17: SLP: solution of the closed-loop implementation of control law (108).

of increased fuel usage), or whether fuel expenditure is critical and should be thus done in a parsimonious manner. Two such cases are depicted in Figure 19. In Case I, fuel is relatively expensive, thus for some noise profiles the spacecraft has a high touchdown speed (mean -0.62 m/s, variance 0.061). In contrast, Case II corresponds to relatively cheap fuel, and thus the algorithm increases the effort to contain the spread of trajectories, thus avoiding a crashing impact even for bad noise profiles (mean touchdown speed -0.55m/s, variance 0.006). This increases the fuel expenditure (used fuel mass of Case II: mean 43.2kg variance 1.5, as opposed to 39.7/1.1 for Case I). Assuming that any touchdown speed higher than 5ft/sec (1.52m/sec) is



(a) The mean of the controlled system trajectories of each iteration (grayscale) and after the final iteration (red).



(b) Optimally controlled trajectories, simulated until touchdown. 100% of the trajectories lead to touchdown.

Figure 18: SLP: solution of the proposed algorithm.

considered a crash[1], we may summarize the comparison results in Table 2. The results are also shown in Figure 20. The superiority of the proposed algorithm in providing a control solution leading to a smooth landing, which is furthermore robust to stochastic disturbances, is evident. In addition, all computations can be performed off-line, leading to a simple implementation, which does not require high on-board computational capability for the spacecraft.

[1]See NASA specifications, e.g., `https://www.nasa.gov/mission_pages/station/structure/elements/soyuz/landing.html`

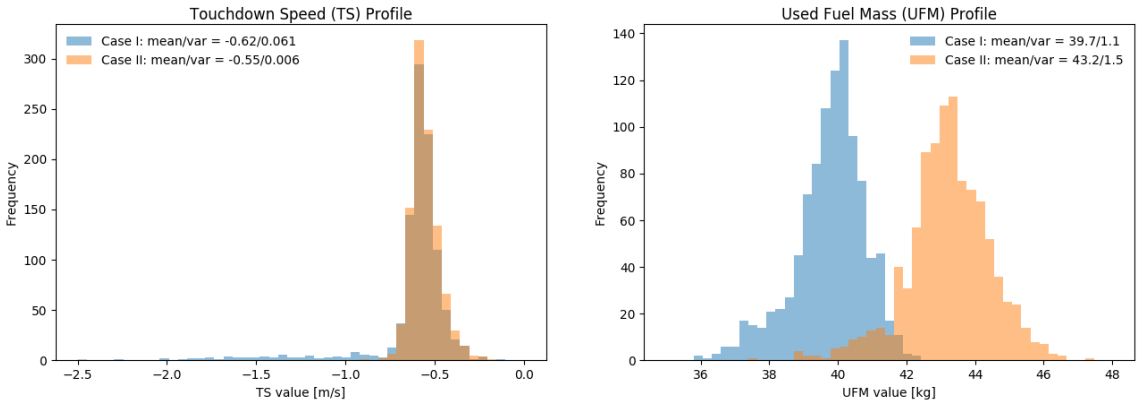
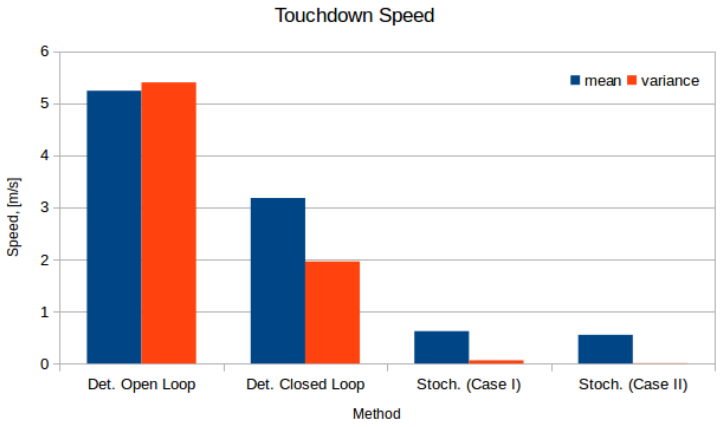


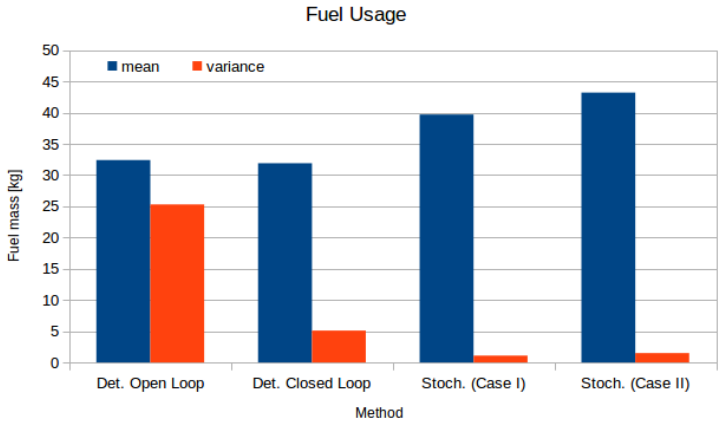
Figure 19: Comparison between the touchdown speed and fuel consumption profiles for cases I and II. In case I, fuel is relatively expensive, thus it is used frugally, leading to low fuel consumption (right figure). This however also leads to a few realizations corresponding to high touchdown speed (spacecraft crashes, left figure). Case II, which is characterized by relatively cheap fuel, greatly reduces the variance of the touchdown speed, thereby avoiding realizations that lead to crashes, at the expense of increased fuel consumption.

Method	Touchdown Speed [m/s] (mean/variance)	Fuel Usage [kg] (mean/variance)	Touchdown Percentage	Crash Percentage
Deterministic, Open Loop	-5.24/5.40	32.4/25.3	50.3%	95.0%
Deterministic, Closed Loop	-3.18/1.96	31.9/5.1	100%	86.8%
Stochastic, Case I	-0.62/0.061	39.7/1.1	100%	2.3%
Stochastic, Case II	-0.55/0.006	43.2/1.5	100%	0*%

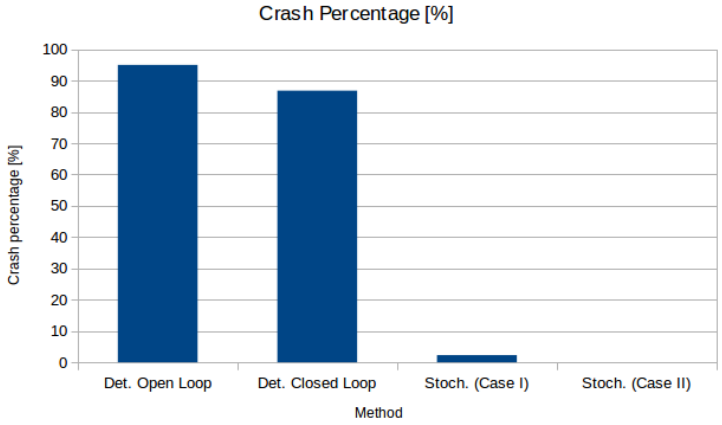
Table 2: Comparison of all methods in terms of touchdown speed, fuel mass used, percentage of trajectories that lead to touchdown, and percentage of trajectories leading to crash. A crash is classified as a trajectory with a touchdown speed greater than 5ft/s (1.52 m/s). For Case II, no crashes occur; the Chebyshev-Cantelli Inequality gives an upper bound of 0.6 % on the probability of a crash occurring in this case.



(a) Touchdown speed comparison (mean/variance)



(b) Fuel usage comparison (mean/variance).



(c) Crash percentage

Figure 20: Performance comparison of the three methods. For (c), a crash is classified as a trajectory with a touchdown speed greater than 5ft/s (1.52 m/s). For Case II of the proposed method, no crashes occur; the Chebyshev-Cantelli Inequality gives an upper bound of 0.6 % on the probability of a crash occurring in this case.

X

CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

In this dissertation, we presented a novel framework to address various stochastic optimal control problems and differential games. In light of a nonlinear Feynman-Kac lemma, we transform the Hamilton-Jacobi-Bellman boundary value problem to a problem involving a system of forward and backward stochastic differential equations. This framework relaxes some of the restrictive conditions present within the existing literature on sampling-based methods for stochastic optimal control. We then develop an efficient numerical scheme to solve the resulting systems of FBSDEs. In particular, the proposed numerical scheme requires only one regression operation per time step (as opposed to $p + 1$ – where p is the dimensionality of the noise– as it is the case in the most established scheme in the literature), and is furthermore enhanced with importance sampling. The latter is derived by means of Girsanov's theorem on the change of measure and allows us to obtain solutions in an iterative manner, whenever the problem is characterized by complex, nonlinear dynamics. We have applied the framework on various problems of stochastic optimal control, including $\mathcal{L}^2$, $\mathcal{L}^1$, and risk-sensitive control in both fixed final time and first-exit settings, as well as differential games. The usefulness of the framework is also illustrated with an application on the soft landing problem.

As future directions of research, we may propose the following:

- **Convergence / Error analysis of the proposed scheme**. A proof of convergence of the proposed scheme was constructed, but it was not complete before the dissertation submission deadline, and thus its publication will be postponed

until a future date. Furthermore, an interesting extension would be the investigation of the various error sources, along with the calculation of their associated upper bounds.

- **Constraint investigation**. It may be feasible to introduce constraints, which will make the framework more useful in addressing practical engineering problems. The herein presented framework can handle "soft-constraints" introduced in the running cost. Alternatively, a first-exit type of formulation can be considered, in which parts of the termination surface are associated with a high terminal cost. Introduction of state constraints in stochastic optimal control is an open research topic, with only a few recent results in the literature [114,115].

- **Mixed $\mathcal{L}^2$-$\mathcal{L}^1$ penalty**. The framework can address control effort penalties that are a combination of $\mathcal{L}^2$ and $\mathcal{L}^1$. The combined penalty allows for a closed-form expression of the optimal control [97], and under the decomposability condition assumed in this dissertation, the HJB PDE is associated to a system of FBSDEs. Some caution is needed in the algorithmic implementation however, since the optimal control is expressed in terms of saturation functions.

- **Higher-dimensional problems: data dimensionality reduction**. Although sampling-based methods do not suffer from the curse of dimensionality, there is still increased difficulty in dealing with high dimensional problems. In the proposed framework, the main bottleneck is expected to be the selection of a good set of basis functions for regression, whenever high-dimensional data are expected to lie on a lower-dimensional manifold. This suggests an interesting research direction towards the application of data dimensionality reduction, various projection methods including the tensor-train decomposition method etc., on the proposed framework.

- **Alternative methods for regression**. In this dissertation, we used linear

regression for the approximation of conditional expectations. This is the fastest and simplest approach, but it is nevertheless plausible that the numerical scheme can benefit from more sophisticated regression methods.

- **Local-to-global: imitation learning**. The algorithm yields a control law which is accurate for an area of the state space that was visited by the sampling trajectories, i.e., it is a local solution. However, during the several iterations of the algorithm, different areas are visited during sampling, with most of the solution information being discarded every time. One can therefore attempt to aggregate this information using a universal function approximator, e.g., a neural network. In this case, the neural network may learn to replace the entire algorithm, similar to what is done in the field of imitation learning.

APPENDIX A

SUPPLEMENTARY MATERIAL TO THE SOFT LANDING PROBLEM

In this appendix, we derive the system of equations that can be solved to obtain the switching time t_s, and final time t_f, in the Soft Landing Problem. It can be shown [38, 87, 88] that the solution to the SLP yields a unique optimal *bang-bang* controller, and that the problem is *normal* (meaning singular control does not appear within the optimal control sequence), and that there is at most one switch time. The optimal sequence is $\{u_{\min}, u_{\max}\}$, with a switching time t_s. The total time duration is t_f. Recall that the dynamics are given by:

$$\dot{h} = v, \tag{113}$$

$$\dot{v} = -g + \frac{u}{m(t)}, \qquad u(t) \in [u_{\min}, u_{\max}], \tag{114}$$

$$\dot{m} = -\alpha u, \tag{115}$$

$$h(0) = h_0, \quad v(0) = v_0, \quad m(0) = m_0, \tag{116}$$

For the first segment $t \in [0, t_s)$, where $u = u_{\min}$, we have by virtue of (115)

$$m(t) = m_0 - \alpha u_{\min} t. \tag{117}$$

Substituting the above in (114) and integrating yields

$$v(t) = v_0 - \frac{1}{\alpha} \ln\left(1 - \frac{\alpha u_{\min}}{m_0} t\right) - gt. \tag{118}$$

Again, we substitute the above expression in (113) and perform integration to obtain

$$h(t) = h_0 + v_0 t + \frac{t}{\alpha} - \frac{1}{\alpha}(t - \frac{m_0}{\alpha u_{\min}})\ln\left(1 - \frac{\alpha u_{\min}}{m_0}t\right) - \frac{1}{2}gt^2. \tag{119}$$

We evaluate the previous three expressions at $t = t_s^-$:

$$h(t_s) = h_0 + v_0 t_s + \frac{t_s}{\alpha} - \frac{1}{\alpha}(t_s - \frac{m_0}{\alpha u_{\min}})\ln\left(1 - \frac{\alpha u_{\min}}{m_0}t_s\right) - \frac{1}{2}gt_s^2, \tag{120}$$

$$v(t_s) = v_0 - \frac{1}{\alpha}\ln\left(1 - \frac{\alpha u_{\min}}{m_0}t_s\right) - gt_s, \tag{121}$$

$$m(t_s) = m_0 - \alpha u_{\min} t_s. \tag{122}$$

We now move to the interval $[t_s, t_f]$, wherein $u = u_{\max}$. Following the same procedure, the state equations are given by

$$h(t) = h(t_s) + v(t_s)(t - t_s) + \frac{t - t_s}{\alpha} - \frac{1}{\alpha}(t - t_s - \frac{m(t_s)}{\alpha u_{\max}})\ln\left(1 - \frac{\alpha u_{\max}}{m(t_s)}(t - t_s)\right) - \frac{1}{2}g(t - t_s)^2, \tag{123}$$

$$v(t) = v(t_s) - \frac{1}{\alpha}\ln\left(1 - \frac{\alpha u_{\max}}{m(t_s)}(t - t_s)\right) - g(t - t_s), \tag{124}$$

$$m(t) = m(t_s) - \alpha u_{\max}(t - t_s). \tag{125}$$

We apply the boundary condition $v(t_f) = 0$ in (124) to obtain

$$v(t_s) = \frac{1}{\alpha}\ln\left(1 - \frac{\alpha u_{\max}}{m(t_s)}(t_f - t_s)\right) + g(t_f - t_s). \tag{126}$$

The above, along with the boundary condition $h(t_f) = 0$, are applied in (123) to obtain

$$h(t_s) = -\frac{t_f - t_s}{\alpha} - \frac{m(t_s)}{\alpha^2 u_{\max}}\ln\left(1 - \frac{\alpha u_{\max}}{m(t_s)(t_f - t_s)}\right) - \frac{1}{2}g(t_f - t_s)^2. \tag{127}$$

We now enforce continuity of states and equate the above to the expressions given by (120) and (121), and simplify to obtain the final expressions

$$h_0 + v_0 t_s + \frac{t_f}{\alpha} - \frac{1}{\alpha}(t_s - \frac{m_0}{\alpha u_{\min}})\ln(1 - \frac{\alpha u_{\min}}{m_0} t_s) - \frac{1}{2} g t_s^2 + \frac{m_0 - \alpha u_{\min} t_s}{\alpha^2 u_{\max}} \ln\left(1 - \frac{\alpha u_{\max}}{m_0 - \alpha u_{\min} t_s}(t_f - t_s)\right) + \frac{1}{2} g (t_f - t_s)^2 = 0, \quad (128)$$

$$\alpha(u_{\max} - u_{\min}) t_s = \alpha u_{\max} t_f + m_0 \left(\exp(\alpha(v_0 - g t_f) - 1) \right). \quad (129)$$

APPENDIX B

AUTHOR PUBLICATIONS

The complete list of publications of the author is as follows:

B.1 Journal Publications

i) **I. Exarchos**, E. Theodorou, and P. Tsiotras, *Stochastic Differential Games: A Sampling Approach via FBSDEs*, submitted to the Journal of Dynamic Games and Applications, under review.

ii) **I. Exarchos**, E. Theodorou, and P. Tsiotras, L^1-*Optimal Control via Forward and Backward Stochastic Differential Equations*, submitted to the Journal of Optimization Theory and Applications, under review.

iii) **I. Exarchos** and E. Theodorou, *Stochastic Optimal Control via Forward and Backward Stochastic Differential Equations and Importance Sampling*, submitted to Automatica, accepted for publication, to appear.

iv) **I. Exarchos**, P. Tsiotras, and M. Pachter, *On the Suicidal Pedestrian Differential Game*, Journal of Dynamic Games and Applications, Springer US, 2014, (article url)

B.2 Conference Publications

i) **I. Exarchos**, E. Theodorou, and P. Tsiotras, *Game-Theoretic and Risk-Sensitive Stochastic Optimal Control via Forward and Backward Stochastic Differential Equations*, 55th IEEE Conference on Decision and Control (CDC), Las Vegas NV, December 12-14, 2016.

ii) **I. Exarchos** and E. Theodorou, *Learning Optimal Control via Forward and Backward Stochastic Differential Equations*, The American Control Conference (ACC), Boston MA, July 6-8, 2016.

iii) **I. Exarchos**, P. Tsiotras, and M. Pachter, *UAV Collision Avoidance based on the Solution of the Suicidal Pedestrian Differential Game*, AIAA Guidance, Navigation, and Control Conference (SciTech), San Diego CA, January 4-8, 2016.

iv) **I. Exarchos** and P. Tsiotras, *An Asymmetric Version of the Two Car Game*, 53rd IEEE Conference on Decision and Control (CDC), Los Angeles CA, December 15-17, 2014.

REFERENCES

[1] Açikmeşe, B. and Ploen, S. R., "Convex programming approach to powered descent guidance for mars landing," *Journal of Guidance, Control, and Dynamics*, vol. 30, no. 5, pp. 1353–1366, 2007.

[2] Aguilar, C. O. and Krener, A. J., "Numerical solutions to the Bellman equation of optimal control," *Journal of Optimization Theory and Applications*, vol. 160, no. 2, pp. 527–552, 2014.

[3] Arapostathis, A., Borkar, V., and Ghosh, M., *Ergodic Control of Diffusion Processes*, vol. 143 of *Encyclopedia of Mathematics and Its Applications*. Cambridge University Press, 2012.

[4] Athans, M. and Falb, P., *Optimal Control- An Introduction to the Theory and Its Applications*. Dover Publications, Inc., 2007.

[5] Başar, T. and Bernhard, P., H^{∞}- *Optimal Control and Related Minimax Design Problems*. Birkhäuser Boston, 2nd ed., 2008.

[6] Bally, V. and Pagès, G., "A quantization algorithm for solving multi-dimensional discrete-time optimal stopping problems," *Bernoulli*, vol. 9, no. 6, pp. 1003–1049, 2003.

[7] Beard, R., Saridis, G., and Wen, J., "Galerkin approximation of the Generalized Hamilton-Jacobi-Bellman equation," *Automatica*, vol. 33, no. 12, pp. 2159–2177, 1997.

[8] Bender, C. and Denk, R., "A forward scheme for backward SDEs," *Stochastic Processes and their Applications*, vol. 117, pp. 1793–1812, December 2007.

[9] Bender, C. and Steiner, J., "Least squares Monte Carlo for backward SDEs," *Numerical Methods in Finance (Springer Proceedings in Mathematics)*, pp. 257–289, 2012.

[10] Bensoussan, A. and Van Schuppen, J. H., "Optimal control of partially observable stochastic systems with an exponential-of-integral performance index," *SIAM J. Control Optim.*, vol. 23, pp. 599–613, July 1985.

[11] Berkovitz, L., "A variational approach to differential games," *RAND Corporation Report*, 1961.

[12] Bouchard, B., Ekeland, I., and Touzi, N., "On the Malliavin approach to Monte Carlo approximation of conditional expectations," *Finance and Stochastics*, vol. 8, pp. 45–71, 2004.

[13] Bouchard, B. and Elie, R., "Discrete-time approximation of decoupled forward-backward SDE with jumps," *Stochastic Processes and their Applications*, vol. 118, pp. 53–75, 2008.

[14] Bouchard, B., Elie, R., and Touzi, N., "Discrete-time approximation of BSDEs and probabilistic schemes for fully nonlinear PDEs," *Radon Series Comp. Appl. Math.*, vol. 8, pp. 91–124, 2009.

[15] Bouchard, B. and Touzi, N., "Discrete time approximation and Monte Carlo simulation of BSDEs," *Stochastic Processes and their Applications*, vol. 111, pp. 175–206, June 2004.

[16] Briand, P., Delyon, B., and Memin, J., "Donsker-type theorem for BSDEs," *Electronic Communications in Probability*, vol. 6, pp. 1–14, January 2001.

[17] Buckdahn, R. and Li, J., "Stochastic differential games and viscosity solutions of Hamilton–Jacobi–Bellman–Isaacs equations," *SIAM J. Control Optim.*, vol. 47, no. 1, pp. 444–475, 2008.

[18] Chan, C. C., "The state of the art of electric, hybrid, and fuel cell vehicles," *Proceedings of the IEEE*, vol. 95, no. 4, pp. 704–718, 2007.

[19] Chassagneux, J. F., Chotai, H., and Muûls, M., *A Forward-Backward SDEs Approach to Pricing in Carbon Markets*. Springer Briefs in Mathematics of Planet Earth, Springer International Publishing AG, October 2017.

[20] Chassagneux, J. F. and Richou, A., "Numerical simulation of quadratic BSDEs," *The Annals of Applied Probability*, vol. 26, no. 1, pp. 262–304, 2016.

[21] Cheridito, P. and Nam, K., "Multidimensional quadratic and subquadratic BSDEs with special structure," *Stochastics An International Journal of Probability and Stochastic Processes*, vol. 87, no. 5, pp. 871–884, 2015.

[22] Cheridito, P., Soner, H. M., Touzi, N., and Victoir, N., "Second-order backward stochastic differential equations and fully nonlinear parabolic PDEs," *Communications on Pure and Applied Mathematics*, vol. 60, no. 7, pp. 1081–1110, 2007.

[23] Crisan, D. and Manolarakis, K., "Sovling backward stochastic differential equations using the cubature method: Application to nonlinear pricing," *SIAM Journal of Financial Mathematics*, vol. 3, no. 1, pp. 534–571, 2012.

[24] Da Lio, F. and Ley, O., "Uniqueness results for second-order Bellman-Isaacs equations under quadratic growth assumptions and applications," *SIAM J. Control Optim.*, vol. 45, no. 1, pp. 74–106, 2006.

[25] Dai Pra, P., Meneghini, L., and Runggaldier, W. J., "Connections between stochastic control and dynamic games," *Mathematics of Control, Signals, and Systems (MCSS)*, vol. 9, no. 4, pp. 303–326, 1996.

[26] Delarue, F. and Menozzi, S., "A forward-backward stochastic algorithm for quasi-linear PDEs," *The Annals of Applied Probability*, vol. 16, no. 1, pp. 140–184, 2006.

[27] Delbaen, F., Hu, Y., and Richou, A., "On the uniqueness of solutions to quadratic BSDEs with convex generators and unbounded terminal conditions," *Annales de l'Institut Henri Poincarè, Probabilitès et Statistiques*, vol. 47, no. 2, pp. 559–574, 2011.

[28] Ding, D. and Liu, Y., "A regression-based numerical method for forward-backward stochastic differential equations," *Available at SSRN: http://ssrn.com/abstract=2513836*, October 2014.

[29] Dixon, M., Edelbaum, T., Potter, J., and Vandervelde, W., "Fuel optimal reorientation of axisymmetric spacecraft," *Journal of Spacecraft and Rockets*, vol. 7, no. 11, pp. 1345–1351, 1970.

[30] Douglas, J., Ma, J., and Protter, P., "Numerical methods for forward-backward stochastic differential equations," *Ann. Appl. Probab.*, vol. 6, pp. 940–968, 1996.

[31] Dueri, D., Açikmeşe, B., Scharf, D., and Harris, M., "Customized real-time interior-point methods for onboard powered-descent guidance," *Journal of Guidance, Control, and Dynamics*, vol. 40, no. 2, pp. 197–212, 2017.

[32] Duncan, T. and Pasik-Duncan, B., "Some stochastic differential games with state dependent noise," *54th IEEE Conference on Decision and Control, Osaka, Japan*, December 15-18, 2015.

[33] Dunham, B., "Automatic on/off switching gives 10-percent gas saving," *Popular Science*, vol. 205, no. 4, p. 170, 1974.

[34] Dvijotham, K. and Todorov, E., "Linearly solvable optimal control," *Reinforcement Learning and Approximate Dynamic Programming for Feedback Control*, pp. 119–141, 2012.

[35] El Karoui, N., Peng, S., and Quenez, M. C., "Backward stochastic differential equations in finance," *Mathematical Finance*, vol. 7, January 1997.

[36] Fahim, A., Touzi, N., and Warin, X., "A probabilistic numerical method for fully nonlinear parabolic PDEs," *The Annals of Applied Probability*, vol. 21, no. 4, pp. 1322–1364, 2011.

[37] Fleming, W., "Exit probabilities and optimal stochastic control," *Applied Math. Optim.*, vol. 9, pp. 329–346, 1971.

[38] Fleming, W. and Rishel, R., *Deterministic and Stochastic Optimal Control*. Springer-Verlag New York Inc., 1975.

[39] Fleming, W. and Soner, H., *Controlled Markov Processes and Viscosity Solutions*. Stochastic Modelling and Applied Probability, Springer, 2nd ed., 2006.

[40] Fleming, W. and Souganidis, P., "On the existence of value functions of two player zero-sum stochastic differential games," *Indiana University Mathematics Journal*, 1989.

[41] Fleming, W. H. and McEneaney, W. M., "Risk-sensitive control on an infinite time horizon," *SIAM J. Control Optim.*, vol. 33, pp. 1881–1915, Nov. 1995.

[42] Friedman, A., *Stochastic Differential Equations and Applications*. Academic Press, 1975.

[43] Gardiner, C., *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*. Springer, 2004.

[44] Gobet, E. and Labart, C., "Error expansion for the discretization of backward stochastic differential equations," *Stochastic Processes and their Applications*, vol. 117, pp. 803–829, 2007.

[45] Gobet, E. and Labart, C., "Solving BSDE with adaptive control variate," *SIAM J. Numer. Anal.*, vol. 48, no. 1, pp. 257–277, 2010.

[46] Gobet, E., Lemor, J. P., and Warin, X., "A regression-based Monte Carlo method to solve backward stochastic differential equations," *The Annals of Applied Probability*, vol. 15, no. 3, 2005.

[47] Gobet, E. and Turkedjiev, P., "Approximation of discrete BSDE using least-squares regression," *hal-00642685v1*, 2011.

[48] Gorodetsky, A., Karaman, S., and Marzouk, Y., "Efficient high-dimensional stochastic optimal motion control using tensor-train decomposition," in *Robotics: Science and Systems (RSS)*, 2015.

[49] Guo, W., Zhang, J., and Zhuo, J., "A monotone scheme for high dimensional fully nonlinear pdes," *Ann. Appl. Probab.*, vol. 25, pp. 1540–1580, 2015.

[50] Györfi, L., Kohler, M., Krzyzak, A., and Walk, H., *A Distribution-Free Theory of Nonparametric Regression*. Springer Series in Statistics, Springer-Verlag New York, Inc., 2002.

[51] Hamadene, S. and Lepeltier, J. P., "Zero-sum stochastic differential games and backward equations," *Systems & Control Letters*, vol. 24, pp. 259–263, 1995.

[52] Ho, Y., Bryson, A., and Baron, S., "Differential games and optimal pursuit-evasion strategies," *IEEE Transactions on Automatic Control*, vol. 10, pp. 385–389, 1965.

[53] Horowitz, M. B. and Burdick, J. W., "Semidefinite relaxations for stochastic optimal control policies," in *American Control Conference, Portland, OR*, pp. 3006–3012, June 4-6, 2014.

[54] Horowitz, M. B., Damle, A., and Burdick, J. W., "Linear Hamilton Jacobi Belman equations in high dimensions," in *53rd IEEE Conference on Decision and Control, Los Angeles, California, USA*, December 15-17 2014.

[55] Isaacs, R., *Differential Games: A Mathematical Theory with Applications to Warfare and Pursuit, Control and Optimization*. New York: Willey, 1965.

[56] Jacobson, D. H., "Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games," *IEEE Transactions on Automatic Control*, vol. 18, pp. 124–131, 1973.

[57] Jeong, D. G. and Jeon, W. S., "Performance of adaptive sleep period control for wireless communications systems," *IEEE Transactions on Wireless Communications*, vol. 5, no. 11, pp. 3012–3016, 2006.

[58] Kappen, H. J., "Linear theory for control of nonlinear stochastic systems," *Physical Review Letters*, vol. 95, November 2005.

[59] Kappen, H. J., "Path integrals and symmetry breaking for optimal control theory," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 11, November 2005.

[60] Kappen, H. J., "An introduction to stochastic control theory, path integrals and reinforcement learning," *Cooperative Behavior in Neural Systems, American Institute of Physics Conference Series*, vol. 887, pp. 149–181, February 2007.

[61] Karatzas, I. and Shreve, S., *Brownian Motion and Stochastic Calculus*. Springer-Verlag New York Inc., 2nd ed., 1991.

[62] Kharroubi, I., Langrené, N., and Pham, H., "Discrete time approximation of fully nonlinear HJB equations via BSDEs with nonpositive jumps," *The Annals of Applied Probability*, vol. 25, no. 4, pp. 2301–2338, 2015.

[63] Kharroubi, I. and Pham, H., "Feynman–Kac representation for Hamilton–Jacobi–Bellman IPDE," *The Annals of Probability*, vol. 43, no. 4, pp. 1823–1865, 2015.

[64] Kharroubi, I., Langrené, N., and Pham, H., "A numerical algorithm for fully nonlinear hjb equations: an approach by control randomization," *Monte Carlo Methods and Applications*, vol. 20, no. 2, pp. 145–165, 2014.

[65] King, J. T., *Introduction to Numerical Computation*. McGraw-Hill, Inc., 1984.

[66] Kloeden, P. and Platen, E., *Numerical Solution of Stochastic Differential Equations*, vol. 23 of *Applications in Mathematics, Stochastic Modelling and Applied Probability*. Springer-Verlag Berlin Heidelberg, 3rd ed., 1999.

[67] Kobylanski, M., "Backward stochastic differential equations and partial differential equations with quadratic growth," *Annals of Probability*, pp. 558–602, 2000.

[68] Kong, L., Wong, G. K., and Tsang, D. H., "Performance study and system optimization on sleep mode operation in IEEE 802.16 e," *IEEE transactions on wireless communications*, vol. 8, no. 9, pp. 4518–4528, 2009.

[69] Kushner, H., "Numerical approximations for stochastic differential games," *SIAM J. Control Optim.*, vol. 41, pp. 457–486, 2002.

[70] Kushner, H. and Chamberlain, S., "On stochastic differential games: Sufficient conditions that a given strategy be a saddle point, and numerical procedures for the solution of the game," *Journal of Mathematical Analysis and Applications*, vol. 26, pp. 560–575, 1969.

[71] Lamperti, J., *Probability- A Survey of the Mathematical Theory*. Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., second ed., 1996.

[72] Lasserre, J. B., Henrion, D., Prieur, C., and Trelat, E., "Nonlinear optimal control via occupation measures and LMI-relaxations," *SIAM Journal of Control and Optimization*, vol. 47, no. 4, pp. 1643–1666, 2008.

[73] Lemor, J. P., Gobet, E., and Warin, X., "Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations," *Bernoulli*, vol. 12, no. 5, pp. 889–916, 2006.

[74] Lepeltier, J. P. and Martìn, J. S., "Existence for BSDE with superlinear–quadratic coefficient," *Stochastics: An International Journal of Probability and Stochastic Processes*, vol. 63, no. 3-4, pp. 227–240, 1998.

[75] Lepeltier, J. P. and San Martin, J., "Backward stochastic differential equations with continuous coefficient," *Statistics & Probability Letters*, vol. 32, no. 4, pp. 425–430, 1997.

[76] Li, Y., Yang, J., and Zhao, W., "Convergence error estimates of the Crank-Nicolson scheme for solving decoupled FBSDEs," *Sci China Math*, vol. 60, pp. 923–948, 2017.

[77] Liang, G., Lionnet, A., and Qian, Z., "On Girsanov's transform for backward stochastic differential equations," *arXiv:1011.3228v1 [math.PR]*, 2010.

[78] Longstaff, F. A. and Schwartz, R. S., "Valuing American options by simulation: A simple least-squares approach," *Review of Financial Studies*, vol. 14, pp. 113–147, 2001.

[79] Luo, P. and Tangpi, L., "Solvability of coupled FBSDEs with quadratic and superquadratic growth," *arXiv preprint arXiv:1505.01796*, 2015.

[80] Ma, J., J., S., and Zhao, Y., "On numerical approximations of forward-backward stochastic differential equations," *SIAM Journal on Numerical Analysis*, vol. 46, no. 5, pp. 2636–2661, 2008.

[81] Ma, J., Protter, P., San Martin, J., and Torres, S., "Numerical method for backward stochastic differential equations," *The Annals of Applied Probability*, vol. 12, no. 1, pp. 302–316, 2002.

[82] Ma, J., Protter, P., and Yong, J., "Solving forward-backward stochastic differential equations explicitly- a four step scheme," *Probability Theory and Related Fields*, vol. 98, pp. 339–359, September 1994.

[83] Ma, J. and Yong, J., *Forward-Backward Stochastic Differential Equations and Their Applications*. Springer-Verlag Berlin Heidelberg, 1999.

[84] Mao, X., *Stochastic Differential Equations and Applications*. Horwood Pub Ltd, 2007.

[85] Mao, X. and Szpruch, L., "Strong convergence and stability of implicit numerical methods for stochastic differential equations with non-globally Lipschitz continuous coefficients," *Journal of Computational and Applied Mathematics*, vol. 238, pp. 14–28, January 2013.

[86] McEneaney, W. M., "A curse-of-dimensionality-free numerical method for solution of certain HJB PDEs," *SIAM Journal of Control and Optimization*, vol. 46, no. 4, pp. 1239–1276, 2007.

[87] Meditch, J., "On the problem of optimal thrust programming for a lunar soft landing," *IEEE Transactions on Automatic Control*, vol. 9, pp. 477–484, October 1964.

[88] Miele, A., "The calculus of variations in applied aerodynamics and flight mechanics," *Optimization Techniques: With Applications to Aerospace Systems*, vol. 5, pp. 99–170, 1962.

[89] Miele, A., "Extremization of linear integrals by Green's theorem," *Optimization Techniques: With Applications to Aerospace Systems*, vol. 5, pp. 69–98, 1962.

[90] Milstein, G. N. and Tretyakov, M. V., "Numerical algorithm for forward-backward stochastic differential equations," *SIAM J. Sci. Comput.*, vol. 28, no. 2, pp. 561–582, 2006.

[91] Mitchell, I. M. and Tomlin, C. J., "Overapproximating reachable sets by Hamilton-Jacobi projections," *Journal of Scientific Computing*, vol. 19, no. 1-3, pp. 323–346, 2003.

[92] Morimoto, J. and Atkeson, C., "Minimax differential dynamic programming: An application to robust biped walking," *Advances in Neural Information*

Processing Systems (NIPS), Vancouver, British Columbia, Canada, December 9-14, 2002.

[93] Morimoto, J., Zeglin, G., and Atkeson, C., "Minimax differential dynamic programming: Application to a biped walking robot," *IEEE/RSJ International Conference on Intelligent Robots and Systems, Las Vegas, NV*, vol. 2, pp. 1927–1932, October 27-31, 2003.

[94] Moseler, T. and Bender, C., "Importance sampling for backward SDEs," *Stochastic Analysis and Applications*, vol. 28, no. 2, pp. 226–253, 2010.

[95] Murphy, K. P., *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.

[96] Nadaraya, E. A., "On estimating regression," *Theory of Probability and its Applications*, vol. 9, no. 1, pp. 141–142, 1964.

[97] Nagahara, M., Quevedo, D. E., and Nešić, D., "Maximum hands-off control and L^1 optimality," in *52nd IEEE Conference on Decision and Control, Florence, Italy*, pp. 3825–3830, December 10-13, 2013.

[98] Nagahara, M., Quevedo, D. E., and Nešić, D., "Maximum hands-off control: a paradigm of control effort minimization," *IEEE Transactions on Automatic Control*, vol. 61, no. 3, pp. 735–747, 2016.

[99] Nogales, A. G., Pèrez, P., and Monfort, P., "A Monte Carlo method to approximate conditional expectations based on a theorem of Besicovitch: Application to equivariant estimation of the parameters of the general half-normal distribution," *arXiv:1306.1182*, 2013.

[100] Øksendal, B., *Stochastic Differential Equations- An Introduction with Applications*. Springer-Verlag Berlin Heidelberg, 6th ed., 2007.

[101] Pardoux, E. and Peng, S., "Backward stochastic differential equations and quasilinear parabolic partial differential equations," *Stochastic Partial Differential Equations and Their Applications*, vol. 176, pp. 200–217, September 2005.

[102] Pham, H., *Continuous- Time Stochastic Control and Optimization with Financial Applications*. Springer Berlin Heidelberg, 2005.

[103] Pham, H., "Feynman-Kac representation of fully nonlinear PDEs and applications," *arXiv:1409.0625*, 2014.

[104] Pontryagin, L. S., "On the theory of differential games," *Uspekhi Mat. Nauk*, vol. 21, pp. 219–274, 1966.

[105] Possamaï, D., "Second order backward stochastic differential equations with continuous coefficient," *arXiv:1201.1049v2*, 2012.

[106] Possamaï, D. and Zhou, C., "Second order backward stochastic differential equations with quadratic growth," *Stochastic Processes and their applications*, vol. 123, no. 10, pp. 3770–3799, 2013.

[107] Ramachandran, K. M. and Tsokos, C. P., *Stochastic Differential Games*. Atlantis Press, 2012.

[108] Rawlik, K., Toussaint, M., and Vijayakumar, S., "On stochastic optimal control and reinforcement learning by approximate inference," *Robotics: Science and Systems*, 2012.

[109] Rombokas, E., Malhotra, M., Theodorou, E., Matsuoka, Y., and Todorov, E., "Tensor-driven variable impedance control using reinforcement learning," *Robotics: Science and Systems, Sydney, Australia*, July 2012.

[110] Rombokas, E., Malhotra, M., Theodorou, E., Todorov, E., and Matsuoka, Y., "Reinforcement learning and synergistic control of the act

hand," *IEEE/ASME Transactions on Mechatronics*, vol. 18, no. 2, pp. 569–577, 2013.

[111] Ross, M. I., "How to find minimum-fuel controllers," *AIAA Guidance, Navigation, and Control Conference and Exhibit, Providence, RI, 16-19 August*, 2004.

[112] Ruijter, M. J. and Oosterlee, C. W., "A Fourier cosine method for an efficient computation of solutions to BSDEs," *SIAM J. Sci. Comput.*, vol. 37, no. 2, pp. A859–A889, 2015.

[113] Ruijter, M. J. and Oosterlee, C. W., "Numerical Fourier method and second-order Taylor scheme for backward SDEs in finance," *Applied Numerical Mathematics*, vol. 103, pp. 1–26, 2016.

[114] Rutquist, P., Breitholtz, C., and Wik, T., "Finite-time state-constrained optimal control for input-affine systems with actuator noise," *Proceedings of the 18th World Congress - The International Federation of Automatic Control, Milano (Italy), August 28 - September 2*, 2011.

[115] Rutquist, P., Wik, T., and Breitholtz, C., "Solving the Hamilton–Jacobi–Bellman equation for a stochastic system with state constraints," *53rd IEEE Conference on Decision and Control, Los Angeles, California, December 15-17*, 2014.

[116] Scharf, D., Açikmeşe, B., Dueri, D., Benito, J., and Cosoliva, J., "Implementation and experimental demonstration of onboard powered-descent guidance," *Journal of Guidance, Control, and Dynamics*, vol. 40, no. 2, pp. 213–229, 2017.

[117] Seywald, H., Kumar, R. R., Deshpande, S. S., and Heck, M. L., "Minimum fuel spacecraft reorientation," *Journal of guidance, control, and dynamics*, vol. 17, no. 1, pp. 21–29, 1994.

[118] Shreve, S., *Stochastic Calculus for Finance II: Continuous Time Models*. Springer Finance Textbooks, Springer, 2004.

[119] Soner, H. M., Touzi, N., and Zhang, J., "Wellposedness of second order backward SDEs," *Probability Theory and Related Fields*, vol. 153, no. 1-2, pp. 149–190, 2012.

[120] Song, Q., Yin, G., and Zhang, Z., "Numerical solutions for stochastic differential games with regime switching," *IEEE Transactions on Automatic Control*, vol. 53, pp. 509–521, March 2008.

[121] Stein, J., *Stochastic Optimal Control, International Finance, and Debt Crises*. Oxford University Press, 2006.

[122] Stengel, R. F., *Optimal Control and Estimation*. Dover Publications, Inc., 1994.

[123] Stulp, F., Theodorou, E., and Schaal, S., "Reinforcement learning with sequences of motion primitives for robust manipulation," *IEEE Transactions on Robotics*, vol. 28, no. 6, pp. 1360–1370, 2012.

[124] Sun, W., Theodorou, E. A., and Tsiotras, P., "Game-theoretic continuous time differential dynamic programming," *American Control Conference, Chicago, IL*, pp. 5593–5598, July 1–3, 2015.

[125] Tevzadze, R., "Solvability of backward stochastic differential equations with quadratic growth," *Stochastic processes and their Applications*, vol. 118, no. 3, pp. 503–515, 2008.

[126] Theodorou, E. A., "Nonlinear stochastic control and information theoretic dualities: Connections, interdependencies and thermodynamic interpretations," *Entropy*, vol. 17, no. 5, pp. 3352–3375, 2015.

[127] Theodorou, E. A., Buchli, J., and Schaal, S., "A generalized path integral control approach to reinforcement learning," *The Journal of Machine Learning Research*, vol. 11, pp. 3137–3181, January 2010.

[128] Theodorou, E. A., Tassa, Y., and Todorov, E., "Stochastic differential dynamic programming," *American Control Conference*, pp. 1125–1132, 2010.

[129] Theodorou, E. A. and Todorov, E., "Relative entropy and free energy dualities: Connections to path integral and KL control," *51st IEEE Conference onf Decision and Control*, pp. 1466–1473, 2012.

[130] Todorov, E., "Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system," *Neural Computation*, vol. 17, pp. 1084–1108, May 2005.

[131] Todorov, E., "Efficient computation of optimal actions," *Proceedings of the National Academy of Sciences*, vol. 106, no. 28, pp. 11478–11483, 2009.

[132] Todorov, E. and Jordan, M., "Optimal feedback control as a theory of motor coordination," *Nature Neuroscience*, 2002.

[133] Todorov, E. and Li, W., "A generalized iterative LQG method for locally optimal feedback control of constrained nonlinear stochastic systems," *American Control Conference*, pp. 300–306, 2005.

[134] Watson, G. S., "Smooth regression analysis," *Sankhyā: The Indian Journal of Statistics*, vol. 26, no. 4, pp. 359–372, 1964.

[135] Whittle, P., "Risk-sensitive linear/quadratic/gaussian control," *Advances in Applied Probability*, vol. 13, no. 4, pp. 764–777, 1981.

[136] Wong, E., Singh, G., and Masciarelli, J., "Guidance and control design for hazard avoidance and safe landing on Mars," *Journal of Spacecraft and Rockets*, vol. 43, pp. 378–384, March - April 2006.

[137] Xiu, D., *Numerical Methods for Stochastic Computations- A Spectral Method Approach*. Princeton University Press, 2010.

[138] Yong, J. and Zhou, X. Y., *Stochastic Controls: Hamiltonian Systems and HJB Equations*. Springer-Verlag New York Inc., 1999.

[139] Zhang, G., Gunzburger, M., and Zhao, W., "A sparse-grid method for multi-dimensional backward stochastic differential equations," *Journal of Computational Mathematics*, vol. 31, no. 3, 2013.

[140] Zhang, J., "A numerical scheme for BSDEs," *The Annals of Applied Probability*, vol. 14, no. 1, pp. 459–488, 2004.

[141] Zhang, J., *Backward Stochastic Differential Equations*. Probability Theory and Stochastic Modelling, Springer Science+Business Media LLC, 2017.

[142] Zhao, W., Li, Y., and Fu, Y., "Second-order schemes for solving decoupled forward backward stochastic differential equations," *Sci China Math*, vol. 57, pp. 665–686, 2014.

[143] Zhao, W., Zhang, W., and Ju, L., "A numerical method and its error estimates for the decoupled forward-backward stochastic differential equations," *Commun. Comput. Phys.*, vol. 15, pp. 618–646, 2014.

[144] Zhao, W., Zhou, T., and Kong, T., "High order numerical schemes for second-order fbsdes with applications to stochastic optimal control," *Commun. Comput. Phys.*, 2017.

[145] Zhao, W., Zhang, G., and Ju, L., "A stable multistep scheme for solving backward stochastic differential equations," *SIAM Journal on Numerical Analysis*, vol. 48, no. 4, pp. 1369–1394, 2010.

VITA

Ioannis Exarchos received his Diploma degree (valedictorian) in Mechanical Engineering and Aeronautics from the University of Patras, Greece, in 2010. He also received the M.S. degrees in both Aerospace Engineering, and Mathematics, from the Georgia Institute of Technology, in 2013 and 2015, respectively. He is currently a Ph.D. candidate in Aerospace Engineering at the Georgia Institute of Technology. He is an Onassis Foundation scholar.