On the Time Discretization of the Feynman-Kac Forward-Backward Stochastic Differential Equations for Value Function Approximation

Kelsey P. Hawkins, Ali Pakniyat, Panagiotis Tsiotras

***Abstract*— Novel numerical estimators are proposed for the forward-backward stochastic differential equations (FBSDE) appearing in the Feynman-Kac representation of the value function. In contrast to the current numerical approaches based on discretization of the continuous-time FBSDE results, we propose a converse approach, by first obtaining a discrete-time approximation of the on-policy value function, and then developing a discrete-time result which resembles the continuous-time counterpart. This approach yields improved numerical estimators in the function approximation phase, and demonstrates enhanced error analysis for those value function estimators. Numerical results and error analysis are demonstrated on a scalar nonlinear stochastic optimal control problem, and they show improvements in the performance of the proposed estimators in comparison with the state-of-the-art methodologies.**

I. INTRODUCTION

Recent investigations have shown that stochastic control problems with nonlinear dynamics and non-quadratic costs can be solved with an iterative application of Feynman-Kac representation theory and its associated forward-backward stochastic differential equations (FBSDEs) [1], [2], [3], [4]. Applying Girsanov's theorem (see, e.g., [5, Chapter 5, Theorem 10.1]) to the typical FBSDE formulation results in the association of a broad class of FBSDEs to the same stochastic optimal control problem. This theoretical result gives rise to an *iterative-FBSDE* (iFBSDE) method, which alternates between solving for the value function over an arbitrary distribution of trajectories, and refining the distribution of trajectories to more closely match an optimally-controlled distribution. Although this theoretical result is well understood, numerical methods which approximate trajectory distributions with Monte-Carlo sample distributions show instability across iterations and offer poor accuracy guarantees on even simple problems such as the linear quadratic regulator (LQR). Improving the stability and accuracy of iFBSDE methods is important for making FBSDE-based approaches viable alternatives to other stochastic control numerical methods.

The solution of a single pair of FBSDEs has been addressed previously in the literature and still remains an active field of study, especially in the mathematical finance community [6], [7], [8], [9], [10], [11]. These methods generally assume that the forward distribution is available and not changing – so that enough of the state space can be covered by dense sampling of the given forward SDE in the order of 10^5 paths or more [7], [11] is satisfactory to produce a desirable FBSDE solution. The focus of such approaches is to show asymptotic convergence of the FBSDE solution as more samples are added or as the approximation is refined, either through Picard-iteration schemes [7] or multi-level schemes [8]. However, since iFBSDE methods iteratively update the forward SDE, dense sampling of any particular forward SDE or highly accurate solutions lead to very slow convergence in iterative application of those methods. Instead, iFBSDE methods seek to produce fast approximations with smaller numbers of samples and good extrapolative properties, since the purpose of solving the backward SDE is to arrive at an improved forward SDE policy in the next iteration.

This paper focuses on improving the accuracy of the value function approximation step, associated with the solution of the backward SDE, under the assumption that the number of trajectory samples is relatively small compared to traditional FBSDE numerical methods. Specifically, we investigate the discrete-time approximation of the backward SDE in the context of solving for the value function in the backward pass. Although some algorithms use analytic solutions of the backward SDEs over short intervals [12], for many nonlinear problems analytic solutions are not available. In this case, one can use Euler-Maruyama approximations for both the continuous-time forward and backward SDEs [1] to solve for the continuous-time value function. In this paper we propose to first form the Euler-Maruyama approximation of the dynamics, costs, and value function, and then target the discrete-time value function for approximation. Instead of approximating relationships arriving from Feynman-Kac FBSDE theory, we derive discrete-time relationships which resemble their continuous-time counterparts. By doing so, we arrive at a set of alternative estimators for the value function of higher numerical accuracy.

The primary contributions of this paper are as follows:

- We propose a pair of alternative estimators for the value function for the backward pass of a Girsanov-shifted Feynman-Kac FBSDE numerical method.
- We characterize the theoretical bias and variance of these estimators and showing its theoretic superiority to previously proposed estimators.
- We numerically confirm the theoretical results on a scalar nonlinear stochastic optimal control problem.

The structure of the paper is as follows. In Section II

K. Hawkins is with Toyota Research Institute, Ann Arbor, Michigan kelsey.hawkins@tri.global

A. Pakniyat is with Faculty of Mechanical Engineering, University of Alabama, Tuscaloosa, Alabama apakniyat@ua.edu

P. Tsiotras is with Faculty of the Institute for Robotics and Intelligent Machines, Georgia Institute of Technology, Atlanta, Georgia tsiotras@gatech.edu

Support for this work has been provided by NSF award IIS-2008686.

we introduce the stochastic optimal control problem we are interested in solving, as well as the continuous-time approach to solving for an on-policy value function using drifted FBSDEs. At the end of this section we describe a discrete-time method of approximating the backward SDE which we propose to improve upon. In Section III we discuss a novel approach to this theory, beginning by replacing the continuous-time problem with a discrete-time approximation. We then use discrete-time relationships to arrive at estimators which resemble the estimators arrived at through continuous-time theory. Finally, in Section IV we present results from a numerical experiment which confirms the error analysis and that they perform better than previously proposed estimators.

Due to space limitations, we present the theorems without proofs. Proofs can be found in the extended version of the paper in [13].

II. CONTINUOUS-TIME FEYNMAN-KAC FBSDES

A. Stochastic Optimal Control Problem

We start with a complete, filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathsf{Q})$, on which W_s^{Q} is an n-dimensional standard Brownian (Wiener) process with respect to the probability measure Q and adapted to the filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$. Consider a stochastic nonlinear system governed by the Itô differential equation

$$\mathrm{d}X_s = f(s, X_s, u_s)\,\mathrm{d}s + \sigma(s, X_s)\,\mathrm{d}W_s^{\mathsf{Q}}, \quad X_0 = x_0, \quad (1)$$

where X_s is a $\mathcal{F}_s$-progressively measurable state process on the interval $s \in [0,T]$ taking values in $\mathbb{R}^n$, $u_{[0,T]}$ is a progressively measurable input process on the same interval, taking values in the compact set $U \subseteq \mathbb{R}^m$, and $f : [0,T]\times\mathbb{R}^n\times U \to \mathbb{R}^n$, $\sigma : [0,T]\times\mathbb{R}^n \to \mathbb{R}^{n\times n}$ are the Markovian drift and diffusion functions, respectively. The cost associated with a given control signal $u_{[t,T]}$ is

$$S_t(u_{[t,T]}) := \int_t^T \ell(s, X_s, u_s)\,\mathrm{d}s + g(X_T), \quad (2)$$

where $\ell : [0,T]\times\mathbb{R}^n\times U \to \mathbb{R}_+$ is the running cost, and $g : \mathbb{R}^n \to \mathbb{R}_+$ is the terminal cost. Let membership of a function in $C_b^{l,k}$ denote that the function and its partial derivatives in t of order $\leq l$ and in x of order $\leq k$ are continuous and bounded on its domain, defining C_b^k similarly. We assume the functions $f, a := \sigma\sigma^\top, \ell \in C_b^{1,2}$, $g \in C_b^3$, and that σ^{-1} exists and is uniformly bounded on its domain. The stochastic optimal control (SOC) problem is to determine the optimal value function

$$V^*(t,x) = \inf_{u_{[t,T]}} \left\{\mathbf{E}_{\mathsf{Q}}[\, S_t(u_{[t,T]}) | X_t = x\,]\right\}. \quad \text{(SOC)}$$

B. On-Policy Value Function

In numerical methods, optimal value functions and optimal policies can only be estimated, so in practice we work with generic Markov policies $\mu : [0,T]\times\mathbb{R}^n \to U$ and the value functions associated with them V^μ, and use iterative methods to approximate V^* and π^* from V^μ and μ. The on-policy value function V^μ is defined as

$$V^\mu(t,x) = \mathbf{E}_{\mathsf{Q}}[\, S_t^\mu | X_t = x\,], \qquad S_t^\mu := \int_t^T \ell_s^\mu\,\mathrm{d}s + g(X_T), \quad (3)$$

with the process X_s satisfying the forward SDE (FSDE)

$$\mathrm{d}X_s = f_s^\mu\,\mathrm{d}s + \sigma_s\,\mathrm{d}W_s^{\mathsf{Q}}, \qquad X_0 = x_0, \quad (4)$$

where we abbreviate $f^\mu := (t,x) \mapsto f(t,x,\mu(t,x))$, $f_s^\mu := (s) \mapsto f^\mu(s, X_s)$, and similarly for ℓ, σ. We call μ an admissible Markov policy if it is Borel-measurable and V^μ has a classic, unique, $C_b^{1,2}$ solution for its associated Hamilton-Jacobi PDE

$$\partial_t V^\mu + \frac{1}{2}\,\mathrm{tr}[\sigma\sigma^\top \partial_{xx} V^\mu] + (\partial_x V^\mu)^\top f^\mu + \ell^\mu\big|_{t,x} = 0, \qquad V^\mu(T,x) = g(x), \quad (5)$$

for $(t,x) \in [0,T)\times\mathbb{R}^n$, where ∂_t and ∂_x are the partial derivative operators in t and x, and ∂_{xx} is the Hessian in x. Since the boundedness of σ^{-1} makes the PDE non-degenerate parabolic, a sufficient condition for the existence of a classical solution is that f^μ and ℓ^μ are in $C_b^{1,2}$ [14, p. 156; Chapter 3, Theorem 4.2, Theorem 4.4]. The same reference guarantees $V^* \equiv V^{\pi^*}$.

C. On-Policy Backward SDE

The positivity of $\sigma\sigma^\top$ yields that (5) is a parabolic PDE and, hence, by the Feynman-Kac Theorem (see, e.g. [15]) it is linked to to the solution (X_s, Y_s, Z_s) of the pair of FBSDEs composed of the FSDE (4) and the backward SDE (BSDE)

$$\mathrm{d}Y_s = -\ell_s^\mu\,\mathrm{d}s + Z_s^\top\,\mathrm{d}W_s^{\mathsf{Q}}, \qquad Y_T = g(X_T), \quad (6)$$

where Y_s and Z_s are, respectively, one and n-dimensional adapted processes. Due largely to [16, Chapter 7, Theorem 4.5, (4.29)], we arrive at the following theorem.

Theorem II.1 (Feynman-Kac Representation)**.** *For the solution (X_s, Y_s, Z_s) to the FBSDE characterized by* (4) *and* (6)*, it holds that*

$$Y_s = V^\mu(s, X_s), \qquad s \in [0,T], \qquad Z_s = \sigma_s^\top \partial_x V^\mu(s, X_s), \quad \textit{a.e. } s \in [0,T], \quad (7)$$

Q*-almost surely (a.s.).*

D. Drifted FBSDE

We now show how an alternative pair of *drifted* FBSDEs with a different trajectory distribution can be used to estimate the same value function V^μ. Starting with a new, complete, filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathsf{P})$, let K_s be a given $\mathcal{F}_s$-progressively measurable and bounded process on the interval generating the forward SDE,

$$\mathrm{d}X_s = K_s\,\mathrm{d}s + \sigma_s\,\mathrm{d}W_s^{\mathsf{P}}, \qquad X_0 = x_0, \quad (8)$$

where W_s^{P} is Brownian in P. Consider further the corresponding drifted backward SDE,

$$\mathrm{d}Y_s = -(\ell_s^\mu + Z_s^\top D_s)\,\mathrm{d}s + Z_s^\top \mathrm{d}W_s^{\mathsf{P}}, \quad Y_T = g(X_T), \quad (9)$$

where

$$D_s := \sigma_s^{-1}(f_s^\mu - K_s). \tag{10}$$

Under very mild conditions on K_s, Girsanov's theorem [5, Chapter 5, Theorem 10.1] can be used to show that the solution (X_s, Y_s, Z_s) to this FBSDE also solves the on-policy FBSDE characterized by (4) and (6). The following drifted representation theorem can then be proved.

Theorem II.2. *For the solution (X_s, Y_s, Z_s) to the drifted FBSDE characterized by* (8) *and* (9)*, it holds that*

$$\begin{aligned} Y_s &= V^\mu(s, X_s), && s \in [0,T], \\ Z_s &= \sigma_s^\top \partial_x V^\mu(s, X_s), && \textit{a.e. } s \in [0,T], \end{aligned} \tag{11}$$

P-*a.s.*.

The relationship over short intervals follows.

Corollary II.1. *Let (X_s, Y_s, Z_s) be the solution to the drifted FBSDE characterized by* (8) *and* (9)*, then,*

$$Y_t = \mathbf{E}_\mathsf{P}[\widehat{Y}_{t,\tau}^{\text{noisy}} | X_t] = \mathbf{E}_\mathsf{P}[\widehat{Y}_{t,\tau}^{\text{noiseless}} | X_t] = V^\mu(t, X_t), \tag{12}$$

P-*a.s. for $0 \le t \le \tau \le T$ for*

$$\widehat{Y}_{t,\tau} := Y_\tau - \Delta\widehat{Y}_{t,\tau}, \tag{13}$$

where $\Delta\widehat{Y}_{t,\tau}$ is either

$$\Delta\widehat{Y}_{t,\tau}^{\text{noisy}} := -\int_t^\tau (\ell_s^\mu + Z_s^\top D_s)\mathrm{d}s + \int_t^\tau Z_s^\top \mathrm{d}W_s^\mathsf{P}, \tag{14}$$

$$\textit{or} \quad \Delta\widehat{Y}_{t,\tau}^{\text{noiseless}} := -\int_t^\tau (\ell_s^\mu + Z_s^\top D_s)\mathrm{d}s. \tag{15}$$

This result suggests that we can choose a drift process K_s at-will, picking any state trajectory distribution over which to solve for the backward process Y_s, which corresponds to the value function. Computation of the backward process estimators $\widehat{Y}_{t,\tau}$ requires the addition of a correction term $Z_s^\top D_s$ to compensate the change of drift in the forward SDE due to K_s. The primary constraint is that the diffusion matrix σ is consistent across different representations of the forward SDE.

E. Euler-Maruyama FBSDE Approximation

Many approaches to solving the FBSDEs, e.g. [1], propose approximating both the forward and backward steps with Euler-Maruyama-like SDE approximations. For the drifted FSDE the approximation is

$$X_\tau - X_t = K_t \Delta t + \sigma_t (\Delta t)^{1/2} \Delta W_t, \tag{16}$$

where $\Delta t := \tau - t$. For the drifted BSDE step variable we use in LSMC, we have

$$\widehat{Y}_{t,\tau} = V(\tau, X_\tau) - \Delta\widehat{Y}_{t,\tau}, \tag{17}$$

where $\Delta\widehat{Y}_{t,\tau}$ is taken in [1] to be

$$\Delta\widehat{Y}_{t,\tau}^{\text{noiseless}} = -(\ell_t^\mu + Z_\tau^\top D_t)\Delta t. \tag{18}$$

The variable Z_τ is evaluated at the end of the interval so that it can utilize the latest approximation of the value function gradient. Although the $Z_\tau^\top \Delta W_t$ term does not, in general, have a conditional expectation of zero, and thus introduces bias into the estimator, excluding it reduces variance and additional error introduced via the approximation of Z_τ. The primary contribution of this paper, discussed in Section III, is proposing new estimators for $\widehat{Y}_{t,\tau}$.

F. Least Squares Monte Carlo

Least squares Monte Carlo (LSMC) is a scheme for obtaining the parameters of a model of the value function V^μ, originally credited to [12]. Let $\phi(x;\alpha)$ be a function representation with parameters $\alpha \in \mathcal{A}$ and $\{(x_t^k, \widehat{y}_t^k)\}_{k=1}^M$ be a set of Monte Carlo samples approximating the joint distribution $(X_t, \widehat{Y}_{t,\tau})$. The value function approximation is obtained by enforcing the constraint (12), minimizing

$$\alpha_t^* = \arg\min_{\alpha \in \mathcal{A}} \sum_{k=1}^M \frac{1}{M} (\widehat{y}_t^k - \phi(x_t^k; \alpha_i))^2, \tag{19}$$

and letting $\phi(x;\alpha_t^*) =: \widetilde{V}^\mu(t,x) \approx V^\mu(t,x)$.

III. Forward-Backward Difference Equations

In this section we begin by forming a discrete-time approximation of the dynamics and value function, then rediscover relationships which resemble those arrived at previously. In doing so, we make two contributions: first, we arrive at better estimators compared to direct discretization of continuous time relations because we are able to exploit characteristics of the discrete-time formulation the continuous-time problem obscures, and, secondly, we provide a discrete-time intuition for what the continuous-time theory is actually doing.

A. Discrete Time SOC Approximation

The interval $[0,T]$ is partitioned into N subintervals of length Δt with the partition $\{t_0 = 0, t_1 = \Delta t, ..., t_{N-1} = T - \Delta t, t_N = T\}$. We abbreviate variables $X_{t_i} =: X_i$ for brevity. Let $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \{\widetilde{\mathcal{F}}_i\}_{i\in\{0,\dots,N\}}, \widetilde{\mathsf{Q}})$ be the discrete-time filtered probability space where W_i^Q is a discrete time Brownian process in $\widetilde{\mathsf{Q}}$ such that $W_i^\mathsf{Q} \sim \mathcal{N}(0, I_n)$ is normally distributed. The on-policy forward difference equation is

$$X_{i+1} - X_i = F_i^\mu + \Sigma_i W_i^\mathsf{Q}, \qquad X_0 = x_0, \tag{20}$$

where, using the Euler-Maruyama approximation method, $F_i^\mu := f(t_i, x, \mu_i(x))\Delta t$, $\Sigma_i(x) := \sigma(t_i, x)\sqrt{\Delta t}$, and the on-policy value function is

$$V_i^\mu(x) := \mathbf{E}_{\widetilde{\mathsf{Q}}}[\sum_{j=i}^{N-1} L_j^\mu + g(X_N) | X_i = x\,], \tag{21}$$

$L_i^\mu := \ell(t_i, x, \mu_i(x))\Delta t$. The value function satisfies the on-policy Bellman equation

$$V_i^\mu(X_i) = L_i^\mu + \mathbf{E}_{\widetilde{\mathsf{Q}}_{i+1}[\mu_i]}[V_{i+1}^\mu(X_{i+1}) | X_i = x]. \tag{22}$$

B. Discrete-Time BSDE Approximation

For the discrete-time value function $\{V_i^\mu\}_i$ and forward process $\{X_i\}_i$ we define the process $\{Y_i := V_i^\mu(X_i)\}_i$. Further, we define the term ΔY_i as one that satisfies the backward difference,

$$Y_i = Y_{i+1} - \Delta Y_i. \tag{23}$$

In our numerical methods, we use estimators $\overline{Y}_{i+1} \approx Y_{i+1}$ and $\Delta\widehat{Y}_i \approx \Delta Y_i$ to obtain a combined estimator

$$\widehat{Y}_i := \overline{Y}_{i+1} - \Delta\widehat{Y}_i. \tag{24}$$

with the interpretation $\widehat{Y}_i \approx V_i^\mu(X_i)$. Both $\overline{Y}_{i+1}$ and $\Delta\widehat{Y}_i$ can be chosen according to different approximation schemes; these choices are investigated in the later parts of this section.

C. On-Policy Taylor-Expanded BSDE

We now propose an alternative expression for $\Delta\widehat{Y}_i$, an analogue to the on-policy terms defined in (14) and (15) when $D_s \equiv 0$. This method of approximation uses second order Taylor-expansions and the on-policy Bellman equation to produce a similar but more accurate backward step,

$$\Delta\widehat{Y}_i^{\textit{taylor}} := -L_i^\mu + \overline{Z}_{i+1}^\top W_i^{\mathsf{Q}} + \frac{1}{2}\operatorname{tr}(\overline{M}_{i+1}(W_i^{\mathsf{Q}} W_i^{\mathsf{Q}\top} - I)), \tag{25}$$

where

$$\overline{Z}_{i+1} := \Sigma_i^\top \partial_x \widetilde{V}_{i+1}^\mu(\overline{X}_{i+1}^{\mathsf{Q}}), \tag{26}$$

$$\overline{M}_{i+1} := \Sigma_i^\top \partial_{xx} \widetilde{V}_{i+1}^\mu(\overline{X}_{i+1}^{\mathsf{Q}})\Sigma_i, \tag{27}$$

and $\overline{X}_{i+1}^{\mathsf{Q}} := X_i + F_i^\mu$. If we compare (25) to the previous method, it is different in two ways. First, the gradient of the value function is evaluated at $\overline{X}_{i+1}^{\mathsf{Q}}$ instead of X_{i+1}. Second, the trace term does not appear in the previous method, though its appearance in this equation is not uncommon. Note that this term has zero mean since $\mathbf{E}_{\mathsf{Q}}[W_i^{\mathsf{Q}} W_i^{\mathsf{Q}\top}|X_i] = I$. In the continuous-time counterpart, these second order effects are infinitesimally small, but in a discrete-time approximation they can no longer be ignored.

The following theorem suggests that this choice of approximation of ΔY_i has relatively small residual error.

Theorem III.1. *The choice* (25) *is an unbiased estimator of the actual value function difference* $\Delta Y_i := V_{i+1}^\mu(X_{i+1}) - V_i^\mu(X_i)$*, i.e.,*

$$\mathbf{E}_{\widetilde{\mathsf{Q}}}[\Delta\widehat{Y}_i|X_i] = \mathbf{E}_{\widetilde{\mathsf{Q}}}[\Delta Y_i|X_i]. \tag{28}$$

Further, its residual error is

$$\Delta Y_i - \Delta\widehat{Y}_i = \delta_{i+1}^{\Delta\widehat{Y}} - \mathbf{E}_{\widetilde{\mathsf{Q}}}[\delta_{i+1}^{\Delta\widehat{Y}}|X_i], \tag{29}$$

$$\delta_{i+1}^{\Delta\widehat{Y}} := \delta_{i+1}^{\widetilde{V}} + \delta_{i+1}^{\text{h.o.t.}}, \tag{30}$$

where $\delta_{i+1}^{\text{h.o.t.}}$ *is the 3rd and higher order terms in the Taylor expansion of* $\widetilde{V}_{i+1}^\mu(X_{i+1})$ *centered at* $\overline{X}_{i+1}^{\mathsf{Q}} := X_i + F_i^\mu$*, and* $\delta_{i+1}^{\widetilde{V}} := V_{i+1}^\mu(X_{i+1}) - \widetilde{V}_{i+1}^\mu(X_{i+1})$ *is the error in the* $(i+1)^{st}$ *step value function representation.*

Under a very basic function approximation scheme, we can dismiss this term entirely.

Proposition III.1. *If the value function approximation* $\widetilde{V}_{i+1}^\mu$ *is quadratic then* $\delta_{i+1}^{\text{h.o.t.}} \equiv 0$.

Note that this does not require the true value function to be quadratic, only its approximation. Although using a less expressive representation in this way improves the error coming from the term $\delta_{i+1}^{\text{h.o.t.}}$, there may be a trade-off in terms of increasing the magnitude of error in $\delta_{i+1}^{\widetilde{V}}$, since the function V_{i+1}^μ might be less appropriately modeled.

The most remarkable aspect of Proposition III.1 is that it suggests that for linear-quadratic-regulator (LQR) problems these estimators are exact up to function approximation error, due to the fact that for LQR problems V_i^μ itself is quadratic and its best function approximation is in the class of quadratic functions. This provides a fundamental guarantee for these estimators.

D. Estimators of $\overline{Y}_{i+1}$

We propose two potential estimators for $\overline{Y}_{i+1} \approx V_{i+1}^\mu(X_{i+1})$ in (24). First, we can use the value function approximation associated with the previous backward step to reestimate the $\overline{Y}_{i+1}$ values,

$$\overline{Y}_{i+1}^{\textit{reestimate}} = \widetilde{V}_{i+1}^\mu(X_{i+1}). \tag{31}$$

We can also choose an estimate expression $\overline{Y}_{i+1}^{\text{noiseless}}$ which ends up cancelling out the terms with W_i^{Q} in them, leaving only

$$\widehat{Y}_i^{\text{noiseless}} = L_i^\mu + \widetilde{V}_{i+1}^\mu(\overline{X}_{i+1}^{\mathsf{Q}}) + \frac{1}{2}\operatorname{tr}(\overline{M}_{i+1}), \tag{32}$$

in place of (24). The following theorem establishes the error analysis of these estimators.

Theorem III.2. *The bias of the reestimate estimator is*

$$\mathbf{E}_{\widetilde{\mathsf{Q}}}[Y_i - \widehat{Y}_i^{\text{reestimate}}|X_i] = \mathbf{E}_{\widetilde{\mathsf{Q}}}[\delta_{i+1}^{\widetilde{V}}|X_i], \tag{33}$$

and the variance of the estimator is

$$\operatorname{Var}_{\widetilde{\mathsf{Q}}}[\widehat{Y}_i^{\text{reestimate}}|X_i] = \operatorname{Var}_{\widetilde{\mathsf{Q}}}[\delta_{i+1}^{\text{h.o.t.}}|X_i]. \tag{34}$$

The bias of the noiseless estimator is

$$\mathbf{E}_{\widetilde{\mathsf{Q}}}[Y_i - \widehat{Y}_i^{\text{noiseless}}|X_i] = \mathbf{E}_{\widetilde{\mathsf{Q}}}[\delta_{i+1}^{\widetilde{V}} + \delta_{i+1}^{\text{h.o.t.}}|X_i], \tag{35}$$

and the variance of the estimator is

$$\operatorname{Var}_{\widetilde{\mathsf{Q}}}[\widehat{Y}_i^{\text{noiseless}}|X_i] = 0. \tag{36}$$

This theorem shows that the *reestimate* condition has less bias than the *noiseless* condition, but is a higher variance estimator.

E. Drifted Taylor-Expanded BSDE

We now offer a discrete-time approximation of the drifted off-policy FBSDEs. Let $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \{\widetilde{\mathcal{F}}_i\}_{i\in\{0,\dots,N\}}, \widetilde{\mathsf{P}})$ be an alternative discrete-time filtered probability space where W_i^{P} is the associated Brownian process. Define also on this space the process $\{K_i\}_{i=0}^{N-1}$ which may be selected using the function $K_i(\omega) = h_i(X_i(\omega), \omega)$. We assume K_i is bounded, $\widetilde{\mathcal{F}}_i$-measurable, and independent of W_i^{P}. Further, we assume K_i is Markovian with respect to X_i.

Choose as the FSDE

$$X_{i+1} - X_i = K_i + \Sigma_i W_i^{\mathsf{P}}, \qquad X_0 = x_0. \tag{37}$$

We choose for the backward step variable, instead of (25),

$$\begin{aligned}\Delta\widehat{Y}_i^{\text{drift}} := &-L_i^\mu + \overline{Z}_{i+1}^\top W_i^{\mathsf{P}} - \overline{Z}_{i+1}^\top D_i \\ &+ \frac{1}{2}\operatorname{tr}(\overline{M}_{i+1}(W_i^{\mathsf{P}} W_i^{\mathsf{P}\top} - I - D_i D_i^\top)),\end{aligned} \tag{38}$$

where

$$D_i := \Sigma_i^{-1}(F_i^\mu - K_i), \tag{39}$$

$$\overline{Z}_{i+1} := \Sigma_i^\top \partial_x \widetilde{V}_{i+1}^\mu(\overline{X}_{i+1}^{\mathsf{P}}), \tag{40}$$

$$\overline{M}_{i+1} := \Sigma_i^\top \partial_{xx} \widetilde{V}_{i+1}^\mu(\overline{X}_{i+1}^{\mathsf{P}})\Sigma_i, \tag{41}$$

$$\overline{X}_{i+1}^{\mathsf{P}} := X_i + K_i. \tag{42}$$

The only difference between these discrete-time off-policy FBSDEs and their on-policy equivalents is the drift term K_i in the FSDE and the two terms with D_i in the BSDE. Indeed, when $K_i = F_i^\mu$ then $D_i = 0$ and the pair returns to the on-policy form, making the proposed off-policy method a generalization of the on-policy method.

The reestimate estimator remains $\overline{Y}_{i+1}^{\textit{reestimate}} = \widetilde{V}_{i+1}^\mu(X_{i+1})$ but the *noiseless* condition now resolves to

$$\begin{aligned}\widehat{Y}_i^{\text{noiseless}} = &L_i^\mu + \widetilde{V}_{i+1}^\mu(\overline{X}_{i+1}^{\mathsf{P}}) + \overline{Z}_{i+1}^\top D_i \\ &+ \frac{1}{2}\operatorname{tr}(\overline{M}_{i+1}(I + D_i D_i^\top)).\end{aligned} \tag{43}$$

Note that the residual error $\Delta Y_i - \Delta\widehat{Y}_i$ is a random variable whose distribution depends on the measure. Since we are now constructing the measure $\widetilde{\mathsf{P}}$ instead of $\widetilde{\mathsf{Q}}$, this estimator is no longer unbiased.

Theorem III.3. *For the solution of the FBSDEs in the drifted measure $\widetilde{P}$, the bias of each of the estimators given in Section III-D is*

$$\mathbf{E}_{\widetilde{\mathsf{P}}}[Y_i - \widehat{Y}_i | X_i, K_i] = \mathbf{E}_{\widetilde{\mathsf{Q}}}[Y_i - \widehat{Y}_i | X_i, K_i] + \varepsilon_{i+1}^{\mathsf{P}|\mathsf{Q}}, \tag{44}$$

where the first term is the bias given in Theorem III.2 and

$$\varepsilon_{i+1}^{\mathsf{P}|\mathsf{Q}} := \mathbf{E}_{\widetilde{\mathsf{P}}}[\delta_{i+1}^{\Delta\widehat{Y}} | X_i, K_i] - \mathbf{E}_{\widetilde{\mathsf{Q}}}[\delta_{i+1}^{\Delta\widehat{Y}} | X_i, K_i], \tag{45}$$

and the variance of the estimator is equivalent

$$\operatorname{Var}_{\widetilde{\mathsf{P}}}[\widehat{Y}_i | X_i, K_i] = \operatorname{Var}_{\widetilde{\mathsf{Q}}}[\widehat{Y}_i | X_i, K_i]. \tag{46}$$

We can characterize this added bias exclusively in the measure $\widetilde{\mathsf{P}}$ using the next result.

Proposition III.2. *The bias term appearing in Theorem III.3 is bounded as*

$$\begin{aligned}&|\mathbf{E}_{\widetilde{\mathsf{Q}}}[\delta_{i+1}^{\Delta\widehat{Y}} | X_i, K_i]| \\ &\quad \le \exp(\frac{1}{2}\|D_i\|^2)\ \mathbf{E}_{\widetilde{\mathsf{P}}}[(\delta_{i+1}^{\Delta\widehat{Y}})^2 | X_i, K_i]^{1/2}.\end{aligned} \tag{47}$$

Although the error bound suggests that the bias grows rapidly with the magnitude $\|D_i\|$, when the magnitude $\|D_i\|$ is small, the first term in the product does not grow much faster than linearly. Further, it is still the case that if the value function approximation is quadratic then the higher order terms $\delta_{i+1}^{\text{h.o.t.}}$ drop out.

IV. NUMERICAL RESULTS

We evaluated these estimators on a 1-dimensional problem with dynamics and costs

$$\begin{aligned}f(t,x,u) &= 0.1(x-3)^2 + 0.2u, \qquad x_0 = 7 \\ \ell(t,x,u) &= 12|x-6| + 0.4u^2, \qquad g(x) = 25x^2,\end{aligned}$$

for a time interval of length $T = 10$, and noise $\sigma = 0.8$. The ground-truth optimal value function is computed by directly evaluating the optimal Bellman equation using a finely grided state space and $N = 200$ timesteps. For the purposes of evaluating and comparing different estimators, we assume the optimal policy is known ahead of time, but in typical applications of these estimators only an approximation would be available. For the initial sampling distribution we select K_i according to $K_i = -0.2X_i$, which is equivalent to F_i^π with the sampling policy of $\pi_i(x) = -0.5(x-3)^2 - x$. We use Hermite polynomials to represent the value function basis functions.

TABLE I: Expressions for the proposed Taylor estimators, as well as the competing Euler-Maruyama estimators.

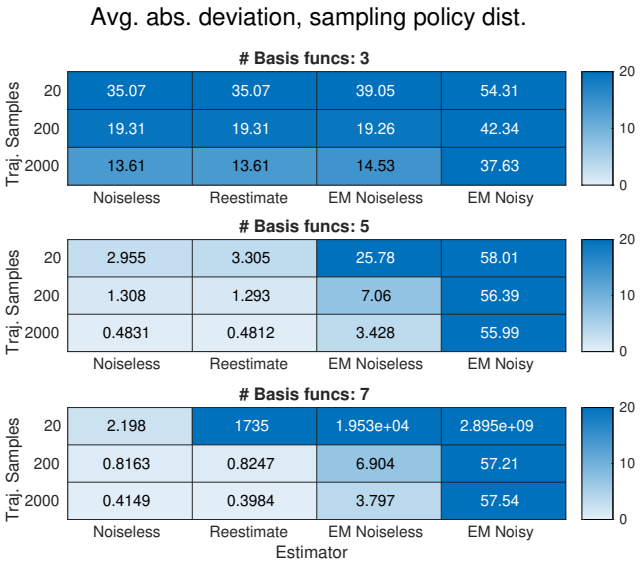
Estimator	$\widehat{Y}_i =$
Taylor Noiseless	$L_i^\mu + \widetilde{V}_{i+1}^\mu(\overline{X}_{i+1}^{\mathsf{P}}) + \overline{Z}_{i+1}^\top D_i + \frac{1}{2}\operatorname{tr}(\overline{M}_{i+1}(I + D_i D_i^\top))$
Taylor Reestimate	$\widetilde{V}_{i+1}^\mu(X_{i+1}) + L_i^\mu - \overline{Z}_{i+1}^\top W_i^{\mathsf{P}} + \overline{Z}_{i+1}^\top D_i + \frac{1}{2}\operatorname{tr}(\overline{M}_{i+1}(I + D_i D_i^\top - W_i^{\mathsf{P}} W_i^{\mathsf{P}\top}))$
Euler-Maruyama Noiseless [1]	$\widetilde{V}_{i+1}^\mu(X_{i+1}) + L_i^\mu + Z_{i+1}^\top D_i$
Euler-Maruyama Noisy	$\widetilde{V}_{i+1}^\mu(X_{i+1}) + L_i^\mu - Z_{i+1}^\top W_i^{\mathsf{P}} + Z_{i+1}^\top D_i$

For each estimator in Table I a single forward pass is performed with the initial sampling policy and a single backward pass is performed with the target policy $\mu = \pi^*$ where π^* is the optimal policy obtained from the ground truth. We repeated this single pair of forward-backward passes under different experimental conditions to investigate how each estimator performs, visualized in Fig. 1. The metric in Fig. 1(a) is the average absolute deviation taken over the sampling policy distribution, then averaged over time, computed as

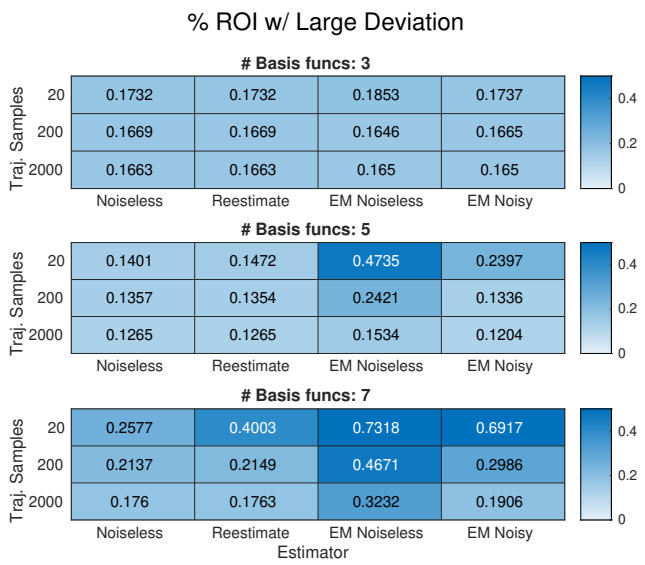
$$\sum_{i=1}^{N} \frac{1}{N} \mathbf{E}_{\widetilde{\mathsf{P}}}[|\widetilde{V}_i(X_i) - V_i^*(X_i)|], \tag{48}$$

where $\widetilde{V}$ is the value function estimate and V^* is the ground truth. The metric in Fig. 1(b) is the percentage of values in the estimate which deviate largely from the ground truth uniformly over the region of interest (ROI) $x \in [-10, 10]$, then averaged over time, computed as

$$\sum_{i=1}^{N} \frac{1}{N} \int_{-10}^{10} \mathbf{1}[|\widetilde{V}_i(x) - V_i^*(x)| > \varepsilon_i] \mathrm{d}x, \tag{49}$$



(a) Average absolute deviation in the sampling policy distribution (48)



(b) Percentage of values in the ROI with a large deviation (49)

Fig. 1: Heatmaps of experiments comparing different estimators with varying numbers of basis functions and numbers of trajectory samples. Each matrix element is averaged over time and over 20 trials.

where $\varepsilon_i := 3\,\sigma_{[-10,10]}[(V_i^*(x))]$ is a multiple of the standard deviation of the ground truth over the ROI, and $\mathbf{1}$ is the indicator function. The metrics (48) and (49) measure local accuracy and extrapolation accuracy, respectively.

The results show that in all cases the proposed Taylor-based estimators perform as well as the Euler-Maruyama estimators and for the vast majority perform significantly better. Although the two proposed estimators show largely equivalent performance, we can see a confirmation of the theory in the bottom left heatmap. When the number of trajectory samples is small and the number of basis functions is high, high variance in the resetimate estimator yields very poor performance. However, when $M = 2000$ samples are collected, its accuracy is improved over the noiseless estimator. In practice, it is likely that the low variance of the noiseless estimator is preferable to its slightly higher bias.

V. CONCLUSION

In this paper we have proposed and justified novel numerical estimators for numerically solving Feynman-Kac FBSDEs for iFBSDE applications. While we have evaluated their effectiveness on a nonlinear one dimensional problem with a single pair of forward and backward passes, future work could evaluate higher dimensional problems which require multiple iterative pairs of passes. In addition, the discrete-time approach used here can improve the selection of optimizing policies.

REFERENCES

[1] I. Exarchos and E. A. Theodorou, "Stochastic optimal control via forward and backward stochastic differential equations and importance sampling," *Automatica*, vol. 87, pp. 159–165, 2018.

[2] I. Exarchos, E. A. Theodorou, and P. Tsiotras, "Stochastic L^1-optimal control via forward and backward sampling," *Systems and Control Letters*, vol. 118, pp. 101–108, 2018.

[3] ——, "Stochastic Differential Games: A Sampling Approach via FBSDEs," *Dynamic Games and Applications*, 2018.

[4] ——, "Game-theoretic and risk-sensitive stochastic optimal control via forward and backward stochastic differential equations," in *Conference on Decision and Control, Las Vegas, Nevada*, 2016, pp. 6154–6160.

[5] W. H. Fleming and R. W. Rishel, "Deterministic and stochastic optimal control," *Bulletin of the American Mathematical Society*, vol. 82, pp. 869–870, 1976.

[6] B. Bouchard, I. Ekeland, and N. Touzi, "On the malliavin approach to monte carlo approximation of conditional expectations," *Finance and Stochastics*, vol. 8, no. 1, pp. 45–71, 2004.

[7] C. Bender and R. Denk, "A forward scheme for backward SDEs," *Stochastic Processes and their Applications*, 2007.

[8] M. B. Giles, "Multilevel Monte Carlo path simulation," *Operations Research*, vol. 56, no. 3, pp. 607–617, 2008.

[9] C. Bender and J. Steiner, "Least-Squares Monte Carlo for Backward SDEs," *Springer Proceedings in Mathematics*, vol. 12, pp. 257–289, 2012.

[10] S. Alanko and M. Avellaneda, "Reducing variance in the numerical solution of bsdes," *Comptes Rendus Mathematique*, vol. 351, no. 3-4, pp. 135–138, 2013.

[11] E. Gobet, J. G. López Salas, P. Turkedjiev, and C. Vázquez, "Stratified regression monte-carlo scheme for semilinear pdes and bsdes with large scale parallelization on gpus," *SIAM Journal on Scientific Computing*, vol. 38, no. 6, pp. C652–C677, 2016.

[12] F. A. Longstaff and E. S. Schwartz, "Valuing American options by simulation: A simple least-squares approach," *Review of Financial Studies*, 2001.

[13] K. P. Hawkins, A. Pakniyat, and P. Tsiotras, "Value function estimators for Feynman-Kac Forward-Backward SDEs in stochastic optimal control," 2021. [Online]. Available: https://arxiv.org/abs/2103.14246

[14] W. H. Fleming and H. M. Soner, *Controlled Markov Processes and Viscosity Solutions*. Springer, 2006, vol. 25.

[15] S. Peng, "Probabilistic interpretation for systems of quasilinear parabolic partial differential equations," *Stochastics Stochastics Rep*, vol. 37, no. 1-2, pp. 61–74, 1991.

[16] J. Yong and X. Y. Zhou, *Stochastic Controls: Hamiltonian Systems and HJB Equations*. Springer, 1999, vol. 43.