

A Multilevel Approach for Stochastic Nonlinear Optimal Control

BY AJAY JASRA¹, JEREMY HENG², YAXIAN XU³ & ADRIAN N. BISHOP⁴

^{1,3}Department of Statistics & Applied Probability, National University of Singapore, Singapore, 117546, SG. E-Mail:

staja@nus.edu.sg, a0078115@u.nus.edu

²ESSEC Business School, 5 Nepal Park, Singapore 139408, SG. E-Mail: *heng@essec.edu*

⁴University of Technology Sydney, AUS. E-Mail: *adrian.bishop@uts.edu.au*

Abstract

We consider a class of finite time horizon nonlinear stochastic optimal control problem, where the control acts additively on the dynamics and the control cost is quadratic. This framework is flexible and has found applications in many domains. Although the optimal control admits a path integral representation for this class of control problems, efficient computation of the associated path integrals remains a challenging Monte Carlo task. The focus of this article is to propose a new Monte Carlo approach that significantly improves upon existing methodology. Our proposed methodology first tackles the issue of exponential growth in variance with the time horizon by casting optimal control estimation as a smoothing problem for a state space model associated with the control problem, and applying smoothing algorithms based on particle Markov chain Monte Carlo. To further reduce computational cost, we then develop a multilevel Monte Carlo method which allows us to obtain an estimator of the optimal control with $\mathcal{O}(\epsilon^2)$ mean squared error with a computational cost of $\mathcal{O}(\epsilon^{-2} \log(\epsilon)^2)$. In contrast, a computational cost of $\mathcal{O}(\epsilon^{-3})$ is required for existing methodology to achieve the same mean squared error. Our approach is illustrated on two numerical examples, which validate our theory.

Key words: Optimal Control; Multilevel Monte Carlo; Markov chain Monte Carlo, Sequential Monte Carlo.

1 Introduction

We consider a class of finite time horizon nonlinear stochastic optimal control problem, where the control acts additively on the dynamics and the control cost is quadratic [14, 15]. This framework is flexible and has applications in domains such as robotics [23], epidemiology [26, 28], reinforcement learning [23], and nonlinear particle smoothing [17, 22]. For this class of control problems, the nonlinear Hamilton-Jacobi-Bellman equation can be reduced to a linear equation by applying a suitable logarithmic transformation. Although this allows the optimal control to admit a closed form expression via the Feynman-Kac formula, efficient computation of the associated path integrals remains a challenging Monte Carlo task. Simple approaches based on simulating the uncontrolled dynamics and performing normalized importance sampling [4] often suffer from exponential growth in variance with the time horizon. Hence to accurately estimate the optimal control, one might require an exponentially commensurate number of Monte Carlo simulations. As optimal importance sampling, i.e. zero variance estimation of the optimal control, is achieved when one simulates from the optimally controlled dynamics [25, Theorem 2], this prompts an iterative procedure [24, 25, 21] to estimate optimal control. Although these iterative importance control methods can often give substantial variance reduction, they typically require parameterizing the form of the control. The focus of this article is to propose a new Monte Carlo approach to the above path integral control problem. Our proposed methodology first tackles the issue of exponential growth in variance with the time horizon by casting optimal control estimation as a smoothing problem for a state space model associated with the control problem, and applying state-of-the-art smoothing algorithms based on particle Markov chain Monte Carlo (MCMC) [1].

To further reduce computational cost, we then consider the multilevel Monte Carlo (MLMC) method [8, 9, 10] which is particularly well-suited to the problem at hand as path integrals are expectations w.r.t. a continuum model, defined by the probability law of the uncontrolled stochastic differential equation (SDE). For numerical implementation to be tractable, one must typically resort to discretizing the continuum model (for instance using Euler discretizations of SDEs) and considering expectations w.r.t. the discretized model. As the time discretization becomes more precise, the approximation becomes more accurate, but simultaneously, the associated cost of computation increases. The MLMC method considers a seemingly trivial telescoping sum representation of the expectation w.r.t. the most precise discretization, where the summands are differences of expectations of increasingly coarsely discretized approximations. The key idea is to approximate the differences by sampling dependent *couplings* of the two probability measures in the difference, independently for each difference. In some contexts, if the couplings are appropriately constructed (see e.g. [8]) then relative to Monte Carlo estimation for the most precise expectation, the computational cost can be significantly reduced. In the case of optimal control estimation, one cannot adopt the MLMC method in [8] directly, as exact sampling from smoothing distributions and their couplings is intractable. To circumvent this difficulty, we adapt the ideas in [13] (see also [7]) to our context to develop a MLMC method that is based on MCMC sampling methods. We then establish that, for some given $\epsilon > 0$, for our proposed estimator of

the optimal control to have $\mathcal{O}(\epsilon^2)$ mean squared error (MSE), the computational cost required is $\mathcal{O}(\epsilon^{-2} \log(\epsilon)^2)$. In contrast, a computational cost of $\mathcal{O}(\epsilon^{-3})$ is required for existing methodology to achieve the same MSE. In the case where one assumes a parametric form of the control, we note that previous work in [21] have considered applying the standard MLMC method in [8] to reduce the computational cost within an iterative importance control scheme to update the control parameters.

This article is structured as follows. In Section 2, we begin by detailing the stochastic optimal control problem of interest and its path integral formulation. We then describe our proposed methodology to compute the optimal control in Section 3, and state some theoretical results on its complexity in Section 4. In Section 5, we validate our theory on two examples, including a nonlinear stochastic compartmental model for an epidemic with cost-controlled vaccination. The appendix features the assumptions and proofs for our complexity theorem in Section 4.

2 Nonlinear stochastic optimal control

We consider a nonlinear controlled process $\{X_t, 0 \leq t \leq T\}$ in \mathbb{R}^n , defined as the solution of the following SDE

$$dX_t = f(X_t)dt + e(X_t)u(t, X_t)dt + g(X_t)dW_t \quad (1)$$

with initial condition $X_0 = x_0 \in \mathbb{R}^n$. The above is to be understood in the Ito sense [2] and $\{W_t, 0 \leq t \leq T\}$ is a standard Brownian motion in \mathbb{R}^d . We assume that $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $e : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m}$ and $g : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times d}$ are twice differentiable and there exists a constant $c_1 > 0$ such that

$$|f(x) - f(y)| + |g(x) - g(y)| + |e(x) - e(y)| \leq c_1|x - y|,$$

for all $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$. Without any loss of generality, we suppose that e and g (which may be non-square) have full rank. Note that the latter implies existence and uniqueness of left-inverses, i.e. functions $e^{-1} : \mathbb{R}^n \rightarrow \mathbb{R}^{m \times n}$ and $g^{-1} : \mathbb{R}^n \rightarrow \mathbb{R}^{d \times n}$ such that $(e^{-1}e)(x) = I_m$ and $(g^{-1}g)(x) = I_d$ for all $x \in \mathbb{R}^n$. Lastly, we assume that $g(x)g(x)^\top$ is uniformly positive definite over $x \in \mathbb{R}^n$. For a fixed time interval $[t_0, t_1] \subseteq [0, T]$, the set of admissible controls $\mathcal{U}_{[t_0, t_1]}$ we shall consider are Borel measurable functions $u : [t_0, t_1] \times \mathbb{R}^n \rightarrow \mathbb{R}^m$ satisfying

$$\mathbb{E}_u^{t_0, x} \left[\int_{t_0}^{t_1} |u(t, X_t)|^q dt \right] < \infty$$

for all $x \in \mathbb{R}^n$ and $q \geq 1$, where $\mathbb{E}_u^{t, x}$ denotes conditional expectations w.r.t. the law of (1) on the event $X_t = x \in \mathbb{R}^n$. These conditions are sufficient for the existence of a unique, continuous, (strong) solution to (1); see for e.g. [2, 27].

For each $(t, x) \in [0, T] \times \mathbb{R}^n$ and $u \in \mathcal{U}_{[t, T]}$, we associate the process defined by (1) with the following cost functional

$$w(t, x, u) = \mathbb{E}_u^{t, x} \left[\phi(X_T) + \int_t^T \{ \ell(X_s) + u(s, X_s)^\top R u(s, X_s) \} ds \right], \quad (2)$$

where $\phi, \ell : \mathbb{R}^n \rightarrow [0, \infty)$ are continuous terminal and running cost functions, respectively, and $R \in \mathbb{R}^{m \times m}$ is a positive-definite symmetric. We then define the value function as

$$v(t, x) = \inf_{u \in \mathcal{U}_{[t, T]}} w(t, x, u). \quad (3)$$

If there exists a unique minimizer of (3) across the entire time horizon $t \in [0, T]$, we will refer to it as the optimal control and denote it by $u^* : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^m$.

For any suitably smooth function $\varphi : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}$, we will denote its partial derivative w.r.t. the time by $\partial_t \varphi : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}$, its gradient w.r.t. the spatial variable by $\nabla \varphi : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ and its Hessian by $\nabla^2 \varphi : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$. For any $A \in \mathbb{R}^{n \times n}$, we write its trace as $\text{tr}(A)$. Under appropriate conditions, the value function (3) can be associated with the following Hamilton-Jacobi-Bellman (HJB) equation

$$-\partial_t v(t, x) = \inf_{u \in \mathcal{U}_{[0, T]}} \left\{ \ell(x) + u(t, x)^\top R u(t, x) + u(t, x)^\top e(x)^\top \nabla v(t, x) + f(x)^\top \nabla v(t, x) + \frac{1}{2} \text{tr} [g(x)g(x)^\top \nabla^2 v(t, x)] \right\} \quad (4)$$

defined for $(t, x) \in [0, T] \times \mathbb{R}^n$, with a boundary condition $v(T, \cdot) = \phi(\cdot)$ at the terminal time T . This association is made precise in the following.

Assumption 2.1. Suppose that the value function $v : [0, T] \times \mathbb{R}^n \rightarrow [0, \infty)$ defined in (3) is once continuously differentiable in the time variable, twice continuously differentiable the spatial variable, and is a (classical) solution to the HJB equation (4).

Sufficient conditions for this assumption to hold, in addition to the modeling hypotheses introduced thus far, can be found for e.g. in [6]. These conditions typically take the form of further regularity or boundedness assumptions on the system (1) and cost functions in (2) and/or their derivatives and are rather standard¹.

Under Assumption 2.1, the value-to-go satisfies the HJB equation

$$-\partial_t v(t, x) = \ell(x) + u^*(t, x)^\top R u^*(t, x) + u^*(t, x)^\top e(x)^\top \nabla v(t, x) + f(x)^\top \nabla v(t, x) + \frac{1}{2} \text{tr} [g(x)g(x)^\top \nabla^2 v(t, x)] \quad (5)$$

for $(t, x) \in [0, T] \times \mathbb{R}^n$, with a boundary condition $v(T, \cdot) = \phi(\cdot)$. From (4), we find that the corresponding optimal control is given by

$$u^*(t, x) = -R^{-1}e(x)^\top \nabla v(t, x) \quad (6)$$

for $(t, x) \in [0, T] \times \mathbb{R}^n$. Substituting the form of the optimal control back into the HJB equation (5) gives

$$-\partial_t v(t, x) = \ell(x) - \frac{1}{2} \nabla v(t, x)^\top e(x) R^{-1} e(x)^\top \nabla v(t, x) + f(x)^\top \nabla v(t, x) + \frac{1}{2} \text{tr} [g(x)g(x)^\top \nabla^2 v(t, x)]$$

which is a nonlinear partial differential equation defined on $[0, T] \times \mathbb{R}^n$. However, the latter can be simplified by considering a logarithmic transformation of the value function

$$\psi(t, x) = \exp \left[\frac{-v(t, x)}{\gamma} \right] \quad (7)$$

for $(t, x) \in [0, T] \times \mathbb{R}^n$ and some $\gamma > 0$ satisfying the following assumption.

Assumption 2.2. Suppose that there exists $\gamma \in \mathbb{R}$ such that $\gamma e(x) R^{-1} e(x)^\top = g(x)g(x)^\top$ for all $x \in \mathbb{R}^d$.

This assumption² is standard in the path integral formulation of optimal control [14], but it also appears more generally in the stochastic optimal control literature [6]. Assumption 2.2 allows us to write

$$-\partial_t \psi(t, x) = -\frac{1}{\gamma} \ell(x) \psi(t, x) + f(x)^\top \nabla \psi(t, x) + \frac{1}{2} \text{tr} [g(x)g(x)^\top \nabla^2 \psi(t, x)]$$

which is a linear partial differential equation on $[0, 1] \times \mathbb{R}^n$, with boundary condition $\psi(T, \cdot) = \exp[-\phi(\cdot)/\gamma]$. By the Feynman-Kac formula, the solution is given by

$$\psi(t, x) = \mathbb{E}^{t,x} \left[\exp \left\{ -\frac{1}{\gamma} \phi(Z_T) - \frac{1}{\gamma} \int_t^T \ell(Z_s) ds \right\} \right],$$

where $\mathbb{E}^{t,x}$ denotes conditional expectations w.r.t. the law of the uncontrolled process $\{Z_t\}$ defined by

$$dZ_s = f(Z_s) ds + g(Z_s) dW_s \quad (8)$$

with initial condition $Z_t = x \in \mathbb{R}^n$. The following result uses the relationships in (6) and (7) to deduce an expression of the optimal control.

Proposition 2.1. Suppose Assumptions 2.1 and 2.2, and the modeling hypotheses hold. We have for $(t, x) \in [0, T] \times \mathbb{R}^n$

$$\begin{aligned} u^*(t, x) &= -R^{-1}e(x)^\top \nabla v(t, x) = \gamma R^{-1}e(x)^\top \nabla \log \psi(t, x) \\ &= \lim_{r \rightarrow 0} \frac{1}{r} \frac{\mathbb{E}^{t,x} \left[\exp \left\{ -\frac{1}{\gamma} \left(\phi(Z_T) + \int_t^T \ell(Z_s) ds \right) \right\} \int_0^r e^{-1}(Z_s) g(Z_s) dW_s \right]}{\mathbb{E}^{t,x} \left[\exp \left\{ -\frac{1}{\gamma} \left(\phi(Z_T) + \int_t^T \ell(Z_s) ds \right) \right\} \right]} \end{aligned} \quad (9)$$

where expectations are path integrals defined by the uncontrolled SDE (8) with initial condition $Z_t = x$.

Proof. This result appears in [25] with $e = g$ and it is straightforward to generalize. \square

The controller form in Proposition 2.1 (and variations of such) is often referred to as the path integral formulation of optimal control [14].

¹It is noteworthy that while a classical solution to the HJB equation arising in deterministic optimal control is not typical, it is well-known [18, 6, 19, 27] that the stochastic optimal control problem is quite generally ‘more regular’.

²The interpretation of this relationship is that along directions where the noise variance is small, the control is deemed more expensive while, conversely, in those directions in which the noise has larger variance the control is cheap [14]. Indeed, this may be desirable in practice since it forces control energy to be spent mostly in those directions in which the noise level may be problematic [23].

3 Computation of optimal control

To simplify notation, throughout this section, we will set the terminal time as $T = 1$ and consider estimating the optimal control at time $t = 0$.

3.1 Standard approach

From Proposition 2.1, the objective is to compute for $r > 0$ small

$$u^*(0, x_0) = \frac{1}{r} \frac{\mathbb{E}^{0, x_0} \left[\exp \left\{ -\frac{1}{\gamma} \left(\phi(Z_1) + \int_0^1 \ell(Z_s) ds \right) \right\} \int_0^r e^{-1}(Z_s) g(Z_s) dW_s \right]}{\mathbb{E}^{0, x_0} \left[\exp \left\{ -\frac{1}{\gamma} \left(\phi(Z_1) + \int_0^1 \ell(Z_s) ds \right) \right\} \right]}. \quad (10)$$

We note that (10) neglects the additional bias incurred by truncating $r > 0$ and refer the reader to [4] for a discussion on the impact of this parameter. To numerically approximate (10), the standard approach to path integral control would rely on a sufficiently precise time discretization of the model. For a sufficiently large $l \in \mathbb{N}$, we consider the Euler-Maruyama discretization of (8) with step size $h = 2^{-l}$

$$Z_{kh} = Z_{(k-1)h} + f(Z_{(k-1)h})h + g(Z_{(k-1)h})W_k, \quad k \in \{1, \dots, 2^l\}, \quad (11)$$

with initial condition $Z_0 = x_0$, where $W_k \sim \mathcal{N}_d(0, hI_d)$ denote independent Brownian increments that are distributed according to a Gaussian distribution mean zero and covariance hI_d . By taking $r = 2^{-(M-1)}$ for $1 < M \leq l$, this prompts the following time discretization of (10)

$$u^l(0, x_0) = \frac{1}{r} \frac{\mathbb{E}_l^{0, x_0} \left[\exp \left\{ -\frac{1}{\gamma} \left(\phi(Z_1) + h \sum_{k=0}^{2^l-1} \ell(Z_{kh}) \right) \right\} \sum_{k=0}^{2^{l-M+1}-1} e^{-1}(Z_{kh}) g(Z_{kh}) W_{k+1} \right]}{\mathbb{E}_l^{0, x_0} \left[\exp \left\{ -\frac{1}{\gamma} \left(\phi(Z_1) + h \sum_{k=0}^{2^l-1} \ell(Z_{kh}) \right) \right\} \right]}, \quad (12)$$

where \mathbb{E}_l^{0, x_0} denotes conditional expectations w.r.t. the law of (11) with initial condition $Z_0 = x_0$.

To approximate the expectations in (12), we can simulate N trajectories $Z_{h:1}^i = (Z_h^i, \dots, Z_1^i)$, $i \in \{1, \dots, N\}$ and consider the standard Monte Carlo approximation

$$u^{l,N}(0, x_0) = \frac{1}{r} \frac{N^{-1} \sum_{i=1}^N \exp \left\{ -\frac{1}{\gamma} \left(\phi(Z_1^i) + h \sum_{k=0}^{2^l-1} \ell(Z_{kh}^i) \right) \right\} \sum_{k=0}^{2^{l-M+1}-1} e^{-1}(Z_{kh}^i) g(Z_{kh}^i) W_{k+1}^i}{N^{-1} \sum_{i=1}^N \exp \left\{ -\frac{1}{\gamma} \left(\phi(Z_1^i) + h \sum_{k=0}^{2^l-1} \ell(Z_{kh}^i) \right) \right\}}. \quad (13)$$

Noting that we can write $u^{l,N}(0, x_0) = r^{-1} \sum_{i=1}^N \omega_l^i \varphi_l(Z_{0:r}^i)$ with normalized weights

$$\omega_l^i = \frac{\exp \left\{ -\frac{1}{\gamma} \left(\phi(Z_1^i) + h \sum_{k=1}^{2^l-1} \ell(Z_{kh}^i) \right) \right\}}{\sum_{j=1}^N \exp \left\{ -\frac{1}{\gamma} \left(\phi(Z_1^j) + h \sum_{k=1}^{2^l-1} \ell(Z_{kh}^j) \right) \right\}}$$

and test function

$$\varphi_l(z_{0:r}) = \sum_{k=0}^{2^{l-M+1}-1} e^{-1}(z_{kh}) g(z_{kh}) g^{-1}(z_{kh}) \{z_{(k+1)h} - z_{kh} - f(z_{kh})h\},$$

it follows that (13) can be seen as the normalized importance sampling estimator of $u^l(0, x_0) = r^{-1} \mathbb{E}_{\pi^l} [\varphi_l(Z_{0:r})]$, where \mathbb{E}_{π^l} denotes expectation w.r.t. the distribution

$$\pi^l(dz_{h:1}) = G^l(z_{h:1}) p^l(dz_{h:1}) / C^l \quad (14)$$

defined on \mathbb{R}^{n2^l} , equipped with the Borel σ -algebra $\mathcal{B}(\mathbb{R}^{n2^l})$. The notation $p^l(dz_{h:1})$ denotes the law of (11) with initial condition x_0 ,

$$G^l(z_{h:1}) = \prod_{k=1}^{2^l} G_k^l(z_{kh}), \quad G_k^l(z_{kh}) = \begin{cases} \exp \left\{ -\frac{h}{\gamma} \ell(z_{kh}) \right\} & \text{for } k \in \{1, \dots, 2^l - 1\}, \\ \exp \left\{ -\frac{1}{\gamma} \phi(z_1) \right\} & \text{for } k = 2^l, \end{cases} \quad (15)$$

Algorithm 1 Sequential Monte Carlo for model (14)

Input: number of particles $N_p \in \mathbb{N}$.

1. At time 0 and particle $i \in \{1, \dots, N_p\}$:
 - (a) set $Z_0^i = x_0$;
 - (b) set ancestor index $A_0^i = i$.
2. For time step $k \in \{1, \dots, 2^l - 1\}$ and particle $i \in \{1, \dots, N_p\}$:
 - (a) sample Brownian increment $W_k^i \sim \mathcal{N}_d(0, hI_d)$;
 - (b) set $Z_{kh}^i = Z_{(k-1)h}^{A_{k-1}^i} + f(Z_{(k-1)h}^{A_{k-1}^i})h + g(Z_{(k-1)h}^{A_{k-1}^i})W_k^i$;
 - (c) sample ancestor index $A_k^i \sim \mathcal{R}\left(G_k^l(Z_{kh}^1), \dots, G_k^l(Z_{kh}^{N_p})\right)$.
3. For time step 2^l :
 - (a) sample Brownian increment $W_{2^l}^i \sim \mathcal{N}_d(0, hI_d)$ for particle $i \in \{1, \dots, N_p\}$;
 - (b) set $Z_1^i = Z_{(2^l-1)h}^{A_{2^l-1}^i} + f(Z_{(2^l-1)h}^{A_{2^l-1}^i})h + g(Z_{(2^l-1)h}^{A_{2^l-1}^i})W_{2^l}^i$;
 - (c) sample an ancestor index $B_{2^l} \sim \mathcal{R}\left(G_{2^l}^l(Z_1^1), \dots, G_{2^l}^l(Z_1^{N_p})\right)$.
4. Trace ancestry by setting $B_k = A_k^{B_{k+1}}$ for $k \in \{1, \dots, 2^l - 1\}$.

Output: trajectory $Z_{h:1} = (Z_h^{B_1}, \dots, Z_1^{B_{2^l}})$ and normalizing constant estimator $C^{l, N_p} = \prod_{k=1}^{2^l} N_p^{-1} \sum_{i=1}^{N_p} G_k^l(Z_{kh}^i)$.

and $C^l < \infty$ is the normalizing constant of (14).

Although the estimator (13) is straightforward to implement and amenable to parallel computation, its variance will often grow exponentially with the time horizon T (taken as 1 in our notation), and particularly so when $l \in \mathbb{N}$ is not sufficiently large [11]. The following section presents an alternative Monte Carlo approach that circumvents this difficulty.

3.2 Smoothing approach

Our proposed methodology is based on the observation that (14) can be seen as the smoothing distribution of a state space model with (11) as the latent Markov process and (15) as the observation densities. This connection between optimal control and smoothing has been previously noted in [5, 16]. This perspective explains why the variance of (13) is often large, and allows us to propose better optimal control estimators by exploiting state-of-the-art smoothing algorithms based on sequential Monte Carlo (SMC) methods.

A basic SMC method known as the bootstrap particle filter is detailed in Algorithm 1, where $\mathcal{R}(\omega_1, \dots, \omega_N)$ refers to a resampling operation based on a vector of nonnegative unnormalized weights $\omega_i, i \in \{1, \dots, N\}$. For example, this is the categorical distribution on $\{1, \dots, N\}$ with probabilities $\omega_i / \sum_{j=1}^N \omega_j, i \in \{1, \dots, N\}$, when multinomial resampling is employed; other lower variance and adaptive resampling schemes can also be considered. The algorithm requires specifying the number of particles $N_p \in \mathbb{N}$ as input, which determines the accuracy and cost of the approximation, and outputs an approximate sample from (14) and an unbiased estimator of its normalizing constant.

We now consider a particular smoothing algorithm known as the particle independent Metropolis-Hastings (PIMH) method, which uses Algorithm 1 as a building block to design a MCMC method to sample from (14); see [1] for additional details and other alternatives. An algorithmic description of PIMH is given in Algorithm 2. From the Markov chain $Z_{h:1}^i, i \in \{1, \dots, N_l\}$ generated by PIMH, one obtains a consistent estimator of the optimal control for the time discretized model (12)

$$u^{l, N_l}(0, x_0) = r^{-1} N_l^{-1} \sum_{i=1}^{N_l} \varphi_l(Z_{0:r}^i) \quad (16)$$

Algorithm 2 Particle independent Metropolis-Hastings for model (14)

Input: number of particles $N_p \in \mathbb{N}$ and iterations $N_l \in \mathbb{N}$.

1. Initialization:

(a) run Algorithm 1 to obtain trajectory $Z_{h:1}^0$ and normalizing constant estimator C_0^{l,N_p} .

2. For iteration $i \in \{1, \dots, N_l\}$:

(a) run Algorithm 1 to obtain trajectory $Z_{h:1}^*$ and normalizing constant estimator C_*^{l,N_p} ;

(b) with probability $\min\left\{1, C_*^{l,N_p}/C_{i-1}^{l,N_p}\right\}$ set $Z_{h:1}^i = Z_{h:1}^*$ and $C_i^{l,N_p} = C_*^{l,N_p}$;

(c) otherwise set $Z_{h:1}^i = Z_{h:1}^{i-1}$ and $C_i^{l,N_p} = C_{i-1}^{l,N_p}$.

Output: trajectories $Z_{h:1}^i, i \in \{1, \dots, N_l\}$.

as the number of iterations $N_l \rightarrow \infty$, for any number of particles $N_p \in \mathbb{N}$. The latter also impacts the quality of the approximation: since the normalizing constant estimator given by Algorithm 1 is consistent as $N_p \rightarrow \infty$, the acceptance probability in Step 2(b) would be close to one if N_p is large.

3.3 Multilevel approach

To further improve the computational efficiency of (16), we now leverage upon the MLMC method [8, 9, 10] by considering a hierarchy of time discretizations with time steps $h_l = 2^{-l}, l \in \{M, \dots, L\}$ for $M < L$. The multilevel approach is based on the following telescopic sum

$$\mathbb{E}_{\pi^L} [\varphi_L(Z_{0:r})] = \mathbb{E}_{\pi^M} [\varphi_M(Z_{0:r})] + \sum_{l=M+1}^L \left\{ \mathbb{E}_{\pi^l} [\varphi_l(Z_{0:r})] - \mathbb{E}_{\pi^{l-1}} [\varphi_{l-1}(Z_{0:r})] \right\}, \quad (17)$$

where π^l refers to the distribution defined in (14) with time step $h = h_l, l \in \{M, \dots, L\}$. The first term in the sum $\mathbb{E}_{\pi^M} [\varphi_M(Z_{0:r})]$ can be approximated using the methodology described in Section 3.2, i.e. we employ Algorithm 2 with $N_M \in \mathbb{N}$ iterations to obtain a Markov chain $Z_{h_M:1}^i, i \in \{1, \dots, N_M\}$, and return the estimator

$$u^{M,N_M}(0, x_0) = r^{-1} N_M^{-1} \sum_{i=1}^{N_M} \varphi_M(Z_{0:r}^i). \quad (18)$$

In standard MLMC [8], the summands

$$\mathbb{E}_{\pi^l} [\varphi_l(Z_{0:r})] - \mathbb{E}_{\pi^{l-1}} [\varphi_{l-1}(Z_{0:r})] \quad (19)$$

are then estimated independently, for $l \in \{M+1, \dots, L\}$, by sampling from appropriately constructed couplings of π_l and π_{l-1} that induce sufficient correlation to reduce the computational cost of just approximating $\mathbb{E}_{\pi^L} [\varphi_L(Z_{0:r})]$ in isolation. However, in our context, exact sampling from these smoothing distributions and their couplings is not feasible. In the following, we will adapt the ideas in [13] to our setup to approximate (19).

We now construct a smoothing distribution

$$\pi^{l,l-1}(dz_{h_l:1}(l), dz_{h_{l-1}:1}(l-1)) = \check{G}^{l,l-1}(z_{h_l:1}(l), z_{h_{l-1}:1}(l-1)) p^{l,l-1}(dz_{h_l:1}(l), dz_{h_{l-1}:1}(l-1)) / C^{l,l-1} \quad (20)$$

defined on the product space $\mathbb{R}^{n2^l} \times \mathbb{R}^{n2^{l-1}}$, equipped with the product σ -algebra $\mathcal{B}(\mathbb{R}^{n2^l}) \times \mathcal{B}(\mathbb{R}^{n2^{l-1}})$, that would allow us to couple our approximation of the smoothing distributions at time discretization levels l and $l-1$. This corresponds to a new state space model: $p^{l,l-1}$ denotes the law of a latent process $(Z_{h_l:1}(l), Z_{h_{l-1}:1}(l-1))$ that evolves according to

$$\begin{aligned} Z_{kh_l}(l) &= Z_{(k-1)h_l}(l) + f(Z_{(k-1)h_l}(l))h_l + g(Z_{(k-1)h_l}(l))W_k(l), \quad k \in \{1, \dots, 2^l\}, \\ Z_{kh_{l-1}}(l-1) &= Z_{(k-1)h_{l-1}}(l-1) + f(Z_{(k-1)h_{l-1}}(l-1))h_{l-1} + g(Z_{(k-1)h_{l-1}}(l-1))W_k(l-1), \quad k \in \{1, \dots, 2^{l-1}\}, \end{aligned} \quad (21)$$

with initial condition $Z_0(l) = Z_0(l-1) = x_0$,

$$\check{G}^{l,l-1}(z_{h_l:1}(l), z_{h_{l-1}:1}(l-1)) = \prod_{k \in K_l^1} \check{G}_k^l(z_{kh_l}(l)) \prod_{k \in K_l^2} \check{G}_k^l(z_{kh_l}(l), z_{kh_{l-1}/2}(l-1))$$

where the observation densities $(\check{G}_1^l, \dots, \check{G}_{2^l}^l)$ given by

$$\begin{aligned} \check{G}_k^l(z_{kh_l}(l)) &= G_k^l(z_{kh_l}(l)) + 1, \quad k \in K_l^1 = \{1, 3, \dots, 2^l - 1\}, \\ \check{G}_k^l(z_{kh_l}(l), z_{kh_{l-1}/2}(l-1)) &= \max \left\{ G_k^l(z_{kh_l}(l)), G_{k/2}^{l-1}(z_{kh_{l-1}/2}(l-1)) \right\}, \quad k \in K_l^2 = \{2, 4, \dots, 2^l\}, \end{aligned} \quad (22)$$

and $C^{l,l-1} < \infty$ is the normalization constant. In (21), the Brownian increments $(W_1(l), \dots, W_{2^l}(l))$ and $(W_1(l-1), \dots, W_{2^{l-1}}(l-1))$ at levels l and $l-1$, respectively, are coupled by independently sampling $W_k(l) \sim \mathcal{N}_d(0, h_l I_d)$ for $k \in \{1, \dots, 2^l\}$ and setting $W_k(l-1) = W_{2(k-1)+1}(l) + W_{2k}(l) \sim \mathcal{N}_d(0, h_{l-1} I_d)$ for $k \in \{1, \dots, 2^{l-1}\}$. To approximate (19), we will rely on the identity

$$\begin{aligned} \mathbb{E}_{\pi^l} [\varphi_l(Z_{0:r})] - \mathbb{E}_{\pi^{l-1}} [\varphi_{l-1}(Z_{0:r})] &= \\ \frac{\mathbb{E}_{\pi^{l,l-1}} \left[\varphi_l(Z_{0:r}(l)) \check{H}^{l,1}(Z_{h_l:1}(l), Z_{h_{l-1}:1}(l-1)) \right]}{\mathbb{E}_{\pi^{l,l-1}} \left[\check{H}^{l,1}(Z_{h_l:1}(l), Z_{h_{l-1}:1}(l-1)) \right]} &- \frac{\mathbb{E}_{\pi^{l,l-1}} \left[\varphi_{l-1}(Z_{0:r}(l-1)) \check{H}^{l,2}(Z_{h_l:1}(l), Z_{h_{l-1}:1}(l-1)) \right]}{\mathbb{E}_{\pi^{l,l-1}} \left[\check{H}^{l,2}(Z_{h_l:1}(l), Z_{h_{l-1}:1}(l-1)) \right]} \end{aligned} \quad (23)$$

where $\mathbb{E}_{\pi^{l,l-1}}$ denotes expectation w.r.t. $\pi^{l,l-1}$ and

$$\check{H}^{l,1}(z_{h_l:1}(l), z_{h_{l-1}:1}(l-1)) = \frac{G^l(z_{h_l:1}(l))}{\check{G}^{l,l-1}(z_{h_l:1}(l), z_{h_{l-1}:1}(l-1))}, \quad (24)$$

$$\check{H}^{l,2}(z_{h_l:1}(l), z_{h_{l-1}:1}(l-1)) = \frac{G^{l-1}(z_{h_{l-1}:1}(l-1))}{\check{G}^{l,l-1}(z_{h_l:1}(l), z_{h_{l-1}:1}(l-1))}. \quad (25)$$

The choice of observation densities in (22) is such that the Radon-Nikodym derivatives (24) are upper-bounded by a finite constant that does not depend on the level l . This ensures that the change of measure approach in (23) would not introduce too much variance relative to exact sampling from a dependent coupling of π_l and π_{l-1} .

Since (20) is just another smoothing distribution on an extended state space, we can also employ the methodology of Section 3.2 to construct a MCMC method to sample from it. The resulting algorithm based on the PIMH method is detailed in Algorithm 4. Like before, this uses a SMC method on model (20), described in Algorithm 3, as a building block. Given the Markov chain $Z_{h_l:1}^i(l), Z_{h_{l-1}:1}^i(l-1), i \in \{1, \dots, N_p\}$ generated by Algorithm 4, we approximate (23) using the estimator

$$\begin{aligned} r \left\{ u^{l, N_l}(0, x_0) - u^{l-1, N_l}(0, x_0) \right\} &= \\ \frac{N_l^{-1} \sum_{i=1}^{N_l} \varphi_l(Z_{0:r}^i(l)) \check{H}^{l,1}(Z_{h_l:1}^i(l), Z_{h_{l-1}:1}^i(l-1))}{N_l^{-1} \sum_{i=1}^{N_l} \check{H}^{l,1}(Z_{h_l:1}^i(l), Z_{h_{l-1}:1}^i(l-1))} &- \frac{N_l^{-1} \sum_{i=1}^{N_l} \varphi_{l-1}(Z_{0:r}^i(l-1)) \check{H}^{l,2}(Z_{h_l:1}^i(l), Z_{h_{l-1}:1}^i(l-1))}{N_l^{-1} \sum_{i=1}^{N_l} \check{H}^{l,2}(Z_{h_l:1}^i(l), Z_{h_{l-1}:1}^i(l-1))}. \end{aligned}$$

Using the above approach independently for levels $l \in \{M+1, \dots, L\}$ and (18) for level $l = M$ gives the following multilevel estimator of the optimal control (10)

$$u^{M:L, N_{M:L}}(0, x_0) = u^{M, N_M}(0, x_0) + \sum_{l=M+1}^L \left\{ u^{l, N_l}(0, x_0) - u^{l-1, N_l}(0, x_0) \right\}. \quad (26)$$

It follows under mild assumptions that (26) is a consistent estimator. In the next section, we will establish, under appropriate assumptions, that the multilevel estimator has a reduced computational cost compared to the single level estimator (16).

Algorithm 3 Sequential Monte Carlo for model (20) at level $l \in \{M + 1, \dots, L\}$

Input: number of particles $N_p \in \mathbb{N}$.

1. At time 0 and particle $i \in \{1, \dots, N_p\}$:
 - (a) set $Z_0^i(l) = x_0$ and $Z_0^i(l-1) = x_0$;
 - (b) set ancestor index $A_0^i = i$.
2. For time step $k \in \{1, \dots, 2^l - 1\}$ and particle $i \in \{1, \dots, N_p\}$:
 - (a) sample Brownian increment $W_k^i(l) \sim \mathcal{N}_d(0, h_l I_d)$ for level l ;
 - (b) set $Z_{kh_l}^i(l) = Z_{(k-1)h_l}^{A_{k-1}^i}(l) + f(Z_{(k-1)h_l}^{A_{k-1}^i}(l))h_l + g(Z_{(k-1)h_l}^{A_{k-1}^i}(l))W_k^i(l)$;
 - (c) if $k \in \{2, 4, \dots, 2^l - 2\}$:
 - i. set $n = k/2$;
 - ii. set Brownian increment $W_n^i(l-1) = W_{2(n-1)+1}^i(l) + W_{2n}^i(l)$ for level $l-1$;
 - iii. set $Z_{nh_{l-1}}^i(l-1) = Z_{(n-1)h_{l-1}}^{A_{n-1}^i}(l-1) + f(Z_{(n-1)h_{l-1}}^{A_{n-1}^i}(l-1))h_{l-1} + g(Z_{(n-1)h_{l-1}}^{A_{n-1}^i}(l-1))W_n^i(l-1)$;
 - iv. sample ancestor index $A_k^i \sim \mathcal{R}\left(\check{G}_k^l(Z_{kh_l}^1(l), Z_{nh_{l-1}}^1(l-1)), \dots, \check{G}_k^l(Z_{kh_l}^{N_p}(l), Z_{nh_{l-1}}^{N_p}(l-1))\right)$.
3. For time step 2^l :
 - (a) sample Brownian increment $W_{2^l}^i(l) \sim \mathcal{N}_d(0, h_l I_d)$ for level l and particle $i \in \{1, \dots, N_p\}$;
 - (b) set $Z_1^i(l) = Z_{(2^l-1)h_l}^{A_{2^l-1}^i}(l) + f(Z_{(2^l-1)h_l}^{A_{2^l-1}^i}(l))h_l + g(Z_{(2^l-1)h_l}^{A_{2^l-1}^i}(l))W_{2^l}^i(l)$;
 - (c) set Brownian increment $W_{2^{l-1}}^i(l-1) = W_{2^{l-1}}^i(l) + W_{2^l}^i(l)$ for level $l-1$ and particle $i \in \{1, \dots, N_p\}$;
 - (d) set $Z_1^i(l-1) = Z_{(2^{l-1}-1)h_{l-1}}^{A_{2^{l-1}-1}^i}(l-1) + f(Z_{(2^{l-1}-1)h_{l-1}}^{A_{2^{l-1}-1}^i}(l-1))h_{l-1} + g(Z_{(2^{l-1}-1)h_{l-1}}^{A_{2^{l-1}-1}^i}(l-1))W_{2^{l-1}}^i(l-1)$;
 - (e) sample an ancestor index $B_{2^l} \sim \mathcal{R}\left(\check{G}_{2^l}^l(Z_1^1(l), Z_1^1(l-1)), \dots, \check{G}_{2^l}^l(Z_1^{N_p}(l), Z_1^{N_p}(l-1))\right)$.
4. Trace ancestry by setting $B_{2^k} = A_{2^k}^{B_{2^{k+1}}}$ for $k \in \{1, \dots, 2^{l-1} - 1\}$ and $B_{2^{k-1}} = B_{2^k}$ for $k \in \{1, \dots, 2^{l-1}\}$.

Output: trajectories $Z_{h_l:1}(l) = (Z_{h_l}^{B_1}, \dots, Z_1^{B_{2^l}})$ and $Z_{h_{l-1}:1}(l-1) = (Z_{h_{l-1}}^{B_2}, \dots, Z_1^{B_{2^l}})$, and normalizing constant estimator $C^{l,l-1,N_p} = \prod_{k \in K_l^1} N_p^{-1} \sum_{i=1}^{N_p} \check{G}_k^l(Z_{kh_l}^i(l)) \prod_{k \in K_l^2} N_p^{-1} \sum_{i=1}^{N_p} \check{G}_k^l(Z_{kh_{l-1}/2}^i(l-1))$.

Algorithm 4 Particle independent Metropolis-Hastings for model (20)

Input: number of particles $N_p \in \mathbb{N}$ and iterations $N_l \in \mathbb{N}$.

1. Initialization:
 - (a) run Algorithm 3 to obtain trajectories $Z_{h_l:1}^0(l)$ and $Z_{h_{l-1}:1}^0(l-1)$, and normalizing constant estimator $C_0^{l,l-1,N_p}$.
2. For iteration $i \in \{1, \dots, N_l\}$:
 - (a) run Algorithm 3 to obtain trajectories $Z_{h_l:1}^*(l)$ and $Z_{h_{l-1}:1}^*(l-1)$, and normalizing constant estimator $C_*^{l,l-1,N_p}$;
 - (b) with probability $\min\left\{1, C_*^{l,l-1,N_p}/C_{i-1}^{l,l-1,N_p}\right\}$ set $Z_{h_l:1}^i(l) = Z_{h_l:1}^*(l)$, $Z_{h_{l-1}:1}^i(l-1) = Z_{h_{l-1}:1}^*(l-1)$ and $C_i^{l,l-1,N_p} = C_*^{l,l-1,N_p}$;
 - (c) otherwise set $Z_{h_l:1}^i(l) = Z_{h_l:1}^{i-1}(l)$, $Z_{h_{l-1}:1}^i(l-1) = Z_{h_{l-1}:1}^{i-1}(l-1)$ and $C_i^{l,l-1,N_p} = C_{i-1}^{l,l-1,N_p}$.

Output: trajectories $Z_{h_l:1}^i(l), Z_{h_{l-1}:1}^i(l-1), i \in \{1, \dots, N_l\}$.

4 Theoretical results

In our context, to sample $\pi^{l,l-1}$ for $l \in \{M+1, \dots, L\}$ (resp. π^M) we will generate a Markov chain on a potentially enlarged space $\mathcal{W}_l \supseteq \mathbb{R}^{n(2^l+2^{l-1})}$ with the σ -algebra \mathcal{W}_l (resp. $(\mathcal{W}_M, \mathcal{W}'_M)$). The purpose of this construction is to facilitate the application of advanced Markov chain samplers such as in [1]. We denote the associated Markov kernel as $K_l : \mathcal{W}_l \times \mathcal{W}_l \rightarrow [0, 1]$, $l \in \{M, \dots, L\}$.

We will consider studying (component-wise)

$$\mathbb{E} \left[\left(\left\{ u^{M:L, N_{M:L}}(0, x_0) - u^L(0, x_0) \right\} + u^L(0, x_0) - u^*(0, x_0) \right)^2 \right] \quad (27)$$

where \mathbb{E} denotes an expectation w.r.t. the Markov chains that have been simulated to estimate $u^L(0, x_0)$. Our objective is to verify that the MSE (27) can be made, for $\epsilon > 0$ given, of $\mathcal{O}(\epsilon^2)$, with a cost that is smaller than using a single MCMC algorithm to approximate $u^L(0, x_0)$.

The assumptions for the following result are in Appendix A.1 and the proof is in Appendix A.2.

Theorem 4.1. *Assume (A1-2). Then there exists a $C < +\infty$ such that for any $L > M > 1$, $N_{M:L} \geq 1$:*

$$\mathbb{E} \left[\left(\left\{ u^{M:L, N_{M:L}}(0, x_0) - u^L(0, x_0) \right\} + u^L(0, x_0) - u^*(0, x_0) \right)^2 \right] \leq C \left(\sum_{l=M}^L \frac{h_l}{N_l} + \sum_{l=M+1}^L \sum_{q=M+1}^L \mathbb{I}_A(l, q) \frac{h_l^{1/2} h_q^{1/2}}{N_l N_q} + h_L \right) \quad (28)$$

where $A = \{(l, q) \in \{M+1, \dots, L\}^2 : l \neq q\}$, \mathbb{I} is the indicator function.

To understand the significance of this result, suppose that one iteration of each Markov chain costs $\mathcal{O}(h_l^{-1})$. The latter is the cost of sampling Euler approximations (note Remark A.1 in the appendix implies that the $\mathcal{O}(h_l^{-1})$ cost per iteration is possible). Let $\epsilon > 0$ be given and set $L = \mathcal{O}(|\log(\epsilon)|)$, so that the square bias is $\mathcal{O}(\epsilon^2)$ (see (36) in the appendix, but this is the h_L term on the R.H.S. of (28)). Now choosing $N_l = \mathcal{O}(\epsilon^{-2} h_l L)$ as in [8] yields

$$\begin{aligned} \sum_{l=M}^L \frac{h_l}{N_l} &= \mathcal{O}(\epsilon^2) \\ \sum_{l=M+1}^L \sum_{q=M+1}^L \mathbb{I}_A(l, q) \frac{h_l^{1/2} h_q^{1/2}}{N_l N_q} &= \mathcal{O}(\epsilon^2). \end{aligned}$$

So, we have achieved a MSE of at most $\mathcal{O}(\epsilon^2)$. The cost to achieve this MSE is

$$\sum_{l=M}^L h_l^{-1} N_l = \mathcal{O}(\epsilon^{-2} \log(\epsilon)^2).$$

If one simply used a Markov chain simulation for π^L or using the original approach in [4] the cost to achieve this same MSE would be $\mathcal{O}(\epsilon^{-3})$; a significant increase.

5 Numerical results

5.1 Linear quadratic Gaussian control

We consider a linear quadratic Gaussian (LQG) control problem where the underlying continuous-time linear controlled process is given by

$$dX_t = AX_t dt + Bu(t, X_t) dt + dW_t \quad (29)$$

with $t \in [0, T]$, $X_t \in \mathbb{R}$, $u(t, X_t) \in \mathbb{R}$ and $W_t \in \mathbb{R}$ denoting a standard Brownian motion in \mathbb{R} . The estimation of the optimal control that minimizes the following quadratic cost function is considered

$$w(t, x, u) = \mathbb{E}_u^{t, x} \left[FX_T^2 + \int_t^T \{QX_s^2 + Ru(s, X_s)^2\} ds \right]. \quad (30)$$

In our experiments, we set $A = -1$, $B = 1$, $F = 1$, $Q = 1$, $R = 0.1$, $M = 4$ and initialize the process at $X_0 = -0.1$. The estimation of the optimal control at time $t = 0$ using the different methods considered is compared:

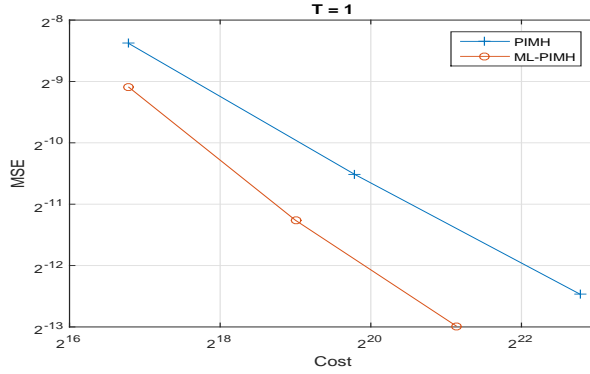


Figure 1: MSE v.s. cost for LQG model with time horizon $T = 1$.

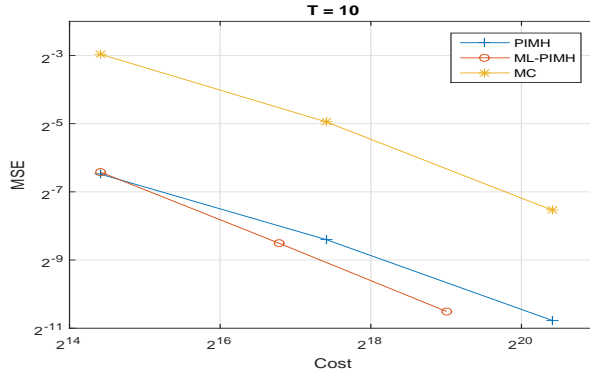


Figure 2: MSE v.s. cost for LQG model with time horizon $T = 10$.

the standard Monte Carlo (MC) approach described in Section 3.1 (see also [3, 15]), single level PIMH introduced Section 3.2, and multilevel Monte Carlo (MLMC) approach based on PIMH as detailed in Section 3.3. For the MLMC method, we set the number of MCMC samples according to the multilevel analysis, i.e. $N_l = \mathcal{O}(\epsilon^{-2}h_l L)$ and $L = \mathcal{O}(|\log(\epsilon)|)$, with the number of particles in SMC to be fixed at $N_p = 500$.

For the single level PIMH method, the number of MCMC samples is $N = \mathcal{O}(\epsilon^{-2})$. The single level PIMH algorithm and the multilevel PIMH algorithm are compared for a time horizon of length $T = 1$, with the MSE v.s. cost plot presented in Figure 1. For the same level of MSE, the cost reduction in the multilevel approach is clear. All three approaches are then compared for a longer time horizon of length $T = 10$. The number of samples is set so that the computational cost of the standard MC approach and that of the PIMH method are the same. From Figure 2, we see a marked reduction in computational cost when employing single level PIMH as compared to the standard MC approach, and a further reduction in cost with the multilevel PIMH approach.

5.2 Nonlinear compartmental model

5.2.1 Model specification

We consider the optimal control of a stochastic compartmental model for an epidemic with cost-controlled vaccination. The state variables are $X_t = (S_t, I_t, V_t, R_t) \in \mathbb{R}^4$ corresponding to susceptible (S), infected (I), vaccinated (V), and removed (R) individuals in a population with the constraint

$$S_t + I_t + V_t + R_t = 1 \tag{31}$$

for all $t \geq 0$. The controlled model considered herein respects this constraint for all $t > 0$, whenever it is enforced to hold at time 0; see [26]. We consider a modification of [26] suitable for our purposes,

$$\begin{aligned}
dS_t &= (\beta - \beta S_t - \kappa I_t S_t + \theta V_t - S_t u(t, X_t)) dt - \sigma S_t dW_t \\
dI_t &= (\kappa S_t I_t + \varepsilon \kappa V_t I_t - \lambda I_t - \beta I_t + \varrho S_t u(t, X_t)) dt + \sigma (S_t - \varepsilon S_t - \sigma_\varrho \varrho S_t) dW_t \\
dV_t &= (-\varepsilon \kappa I_t V_t - \beta V_t - \theta V_t + (1 - \varrho) S_t u(t, X_t)) dt + \sigma (\varepsilon S_t + \sigma_\varrho \varrho S_t) dW_t \\
dR_t &= (\lambda I_t - \beta R_t) dt
\end{aligned} \tag{32}$$

where $W_t \in \mathbb{R}$ denotes a standard Brownian motion in \mathbb{R} . In this model, the birth and death rate are given by $\beta \in (0, 1)$, and the infection rate is controlled by $\kappa > 0$, known as the contact rate. The parameter $\lambda \geq 0$ controls the curing rate, $\theta \geq 0$ controls the loss of vaccine effectiveness, and $\varepsilon \in (0, 1)$ controls the efficacy of the vaccination protocol, i.e. letting $\varepsilon = 0$ would imply the vaccine is perfectly effective, while $\varepsilon = 1$ implies the vaccination has no effect. The parameter $0 < \varrho \approx 0$ is necessary for our model to be well-defined and taken small enough so that it has no effect qualitatively.

The control input $u \in \mathbb{R}$ specifies the fraction of the susceptible class being vaccinated at any moment. Note that although we would like $u \in [0, 1]$, values outside this constraint cause no mathematical difficulty and do not pose a problem for satisfaction of the constraint (31), i.e. $d(S_t + I_t + V_t + R_t) = 0$. Given some fixed terminal time $T > 0$, the cost function we aim to minimize is given by

$$w(t, x, u) = \mathbb{E}_u^{t,x} \left[I_T^2 + \int_t^T \{q I_s + r u(s, X_s)^2\} ds \right], \tag{33}$$

for $t \in [0, T]$, where $q, r > 0$ are weighting parameters. Note that the running cost is linear in the state.

The basic reproduction rate is $\mathfrak{R}_0 = \kappa/(\beta + \lambda)$ and, with $\theta = \varrho = 0$, it is shown in [28] that for $\mathfrak{R}_0 < 1$ the stochastic system (32) is almost surely exponentially stable with any constant u to the equilibrium

$$(S^*, I^*, V^*, R^*) = \left(\frac{\beta}{\beta + u}, 0, 1 - \frac{\beta}{\beta + u}, 0 \right).$$

Hence, we take parameters such that $\mathfrak{R}_0 > 1$ going forward. In particular, we take $\beta = 0.016$, $\kappa = 0.55$, $\lambda = 0.45$ and $\varepsilon = 0.4$. We also take $\theta = 0.1$, $\varrho = 0.01$ and $\sigma = 0.4$. The initial condition is $(S_0, I_0, V_0, R_0) = (0.75, 0.15, 0.05, 0.05)$. To check our assumptions stated in Section 2, we note that

$$\begin{aligned}
&\gamma r^{-1} \begin{pmatrix} -S_t \\ \varrho S_t \\ (1 - \varrho) S_t \\ 0 \end{pmatrix} \begin{pmatrix} -S_t, & \varrho S_t, & (1 - \varrho) S_t, & 0 \end{pmatrix} \\
&= \sigma^2 \begin{pmatrix} -S_t \\ (1 - \varepsilon - \sigma_\varrho \varrho) S_t \\ (\varepsilon + \sigma_\varrho \varrho) S_t \\ 0 \end{pmatrix} \begin{pmatrix} -S_t, & (1 - \varepsilon - \sigma_\varrho \varrho) S_t, & (\varepsilon + \sigma_\varrho \varrho) S_t, & 0 \end{pmatrix}
\end{aligned}$$

for some $\gamma \in \mathbb{R}$. Indeed, we require $\varepsilon + (\sigma_\varrho + 1)\varrho = 1$ in order to find γ uniquely. Given our prior parameters this implies $\sigma_\varrho = 59$ and then $\gamma = 0.16r$. We only need the left inverse of $(-S_t, \varrho S_t, (1 - \varrho) S_t)^\top$. Note in the control computation we can act as if the system is three-dimensional and ignore the dynamics of R_t . The left inverse in this case is

$$\begin{pmatrix} -S_t \\ \varrho S_t \\ (1 - \varrho) S_t \end{pmatrix}^{-1} = \left[\begin{pmatrix} -S_t, & \varrho S_t, & (1 - \varrho) S_t \end{pmatrix} \begin{pmatrix} -S_t \\ \varrho S_t \\ (1 - \varrho) S_t \end{pmatrix} \right]^{-1} \begin{pmatrix} -S_t, & \varrho S_t, & (1 - \varrho) S_t \end{pmatrix}$$

which in general exists since the inverse on the right hand side in general exists.

5.2.2 Numerical results

In our experiments, we set $q = 1$, $r = 0.05$, $T = 3$ and $M = 3$. In the following, we will compare the single level PIMH algorithm to the multilevel PIMH algorithm for the task of optimal control estimation. As the standard MC approach did not perform well in the simple LQG model, it is not considered for this application. A trajectory of the

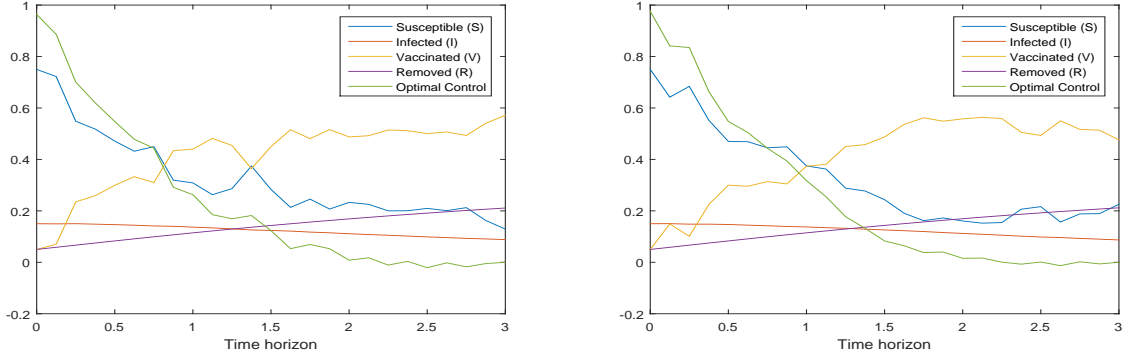


Figure 3: A trajectory of the controlled process and the corresponding approximation of the optimal control generated by the single level PIMH algorithm (left) and the multilevel PIMH algorithm (right).

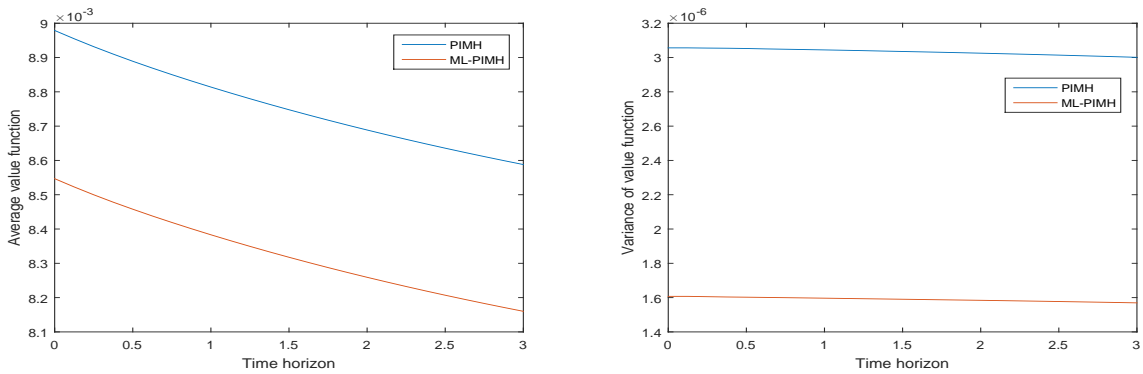


Figure 4: Sample average (left) and sample variance (right) of value function over time interval $[0, T]$.

controlled process generated by the two algorithms are shown in Figure 3. It is clear from Figure 3 that the infected and the susceptible compartments are decreasing over time, while the vaccinated and the removed compartments are increasing over time.

In Figure 4, we compare the two algorithms at a fixed computational cost by reporting the sample average (left panel) and sample variance (right panel) of the value function, computed using 20 independent repetitions of each algorithm. We observe that the multilevel approach achieves lower values on average in terms of the objective (33) with much smaller variance.

Lastly, for the task of estimating the optimal control at time $t = 0$, we present a MSE v.s. cost plot in Figure 5. For the MLMC method, we set the number of MCMC samples according to the multilevel analysis i.e. $N_l = \mathcal{O}(\epsilon^{-2}h_lL)$ and $L = \mathcal{O}(|\log(\epsilon)|)$, with the number of particles in SMC to be fixed at $N_p = 200$. For the single level PIMH method, the number of MCMC samples is taken as $N = \mathcal{O}(\epsilon^{-2})$. The true value is computed by running the latter algorithm at one plus the most precise level (i.e. $L + 1$) and the MSE is computed using 20 independent repetitions of each algorithm. The results illustrate that the multilevel approach offers significant reduction in computational cost.

Acknowledgements

AJ & YX were supported by an ACRF tier 2 grant: R-155-000-161-112. AJ is affiliated with the Risk Management Institute, the Center for Quantitative Finance and the OR & Analytics cluster at NUS. AJ was supported by a KAUST CRG4 grant ref: 2584.

A Technical results

Throughout our proofs C is a finite constant that does not depend upon l and whose value may change on each appearance.

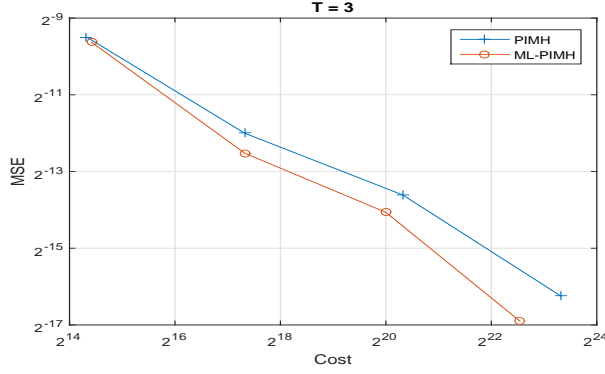


Figure 5: MSE v.s. cost for nonlinear compartmental model.

A.1 Assumptions

In the context of Theorem 4.1, to shorten our proofs, we will suppose that the Markov chain(s) are started in stationarity. This latter assumption can be removed with some work, but is unnecessary in order to convey the point of our approach. Note also that we are assuming that r is fixed throughout and there is an additional bias which is not addressed. In order to derive our theoretical results, in addition to the assumptions that have already been made, we will make the following assumptions. Below $\mathcal{P}(W_l)$ denotes the collection of probability measures on W_l .

- (A1) ϕ, ℓ, e, f, g are all bounded and measurable. In addition ϕ, ℓ are Lipschitz. Set $\alpha = e^{-1}g$, then we assume each element of α is bounded and Lipschitz.
- (A2) K_l is reversible. There exists a $\kappa \in (0, 1)$ such that for each $l \in \{M, \dots, L\}$ there exists a $\nu \in \mathcal{P}(W_l)$ such that for any $\eta : W_l \rightarrow \mathbb{R}$ bounded, measurable and Lipschitz, $x \in W_l$

$$\int_{W_l} \eta(x') K_l(x, dx') \geq \kappa \int_{W_l} \eta(x') \nu(dx').$$

Remark A.1. In (A2) we have assumed a mixing rate that will be independent of l . At first glance, it may seem that this is not possible in practice. However, using Lemma A.1 below, one can easily establish that (for example) the PIMH algorithm in [1] has such a property.

A.2 Proof of Theorem 4.1

The proof is constructed by using several results, which are first quoted and proved later on in the appendix.

Theorem A.1. Assume (A1-2). Then there exists a $C < \infty$ such that for any $l \in \{M+1, \dots, L\}$, $N_l \geq 1$:

$$\mathbb{E}[(u^{l, N_l}(0, x_0) - u^{l-1, N_l}(0, x_0) - \{u^l(0, x_0) - u^{l-1}(0, x_0)\})^2] \leq \frac{Ch_l}{N_l}.$$

Proof. This essentially the same as [13, Theorem 3.1] except that one needs Proposition A.2 and Lemma A.1 in Appendix A.3. \square

The proof of the following result is in Appendix A.4.

Proposition A.1. Assume (A1-2). Then there exists a $C < \infty$ such that for any $l \in \{M+1, \dots, L\}$, $N_l \geq 1$:

$$|\mathbb{E}[u^{l, N_l}(0, x_0) - u^{l-1, N_l}(0, x_0) - \{u^l(0, x_0) - u^{l-1}(0, x_0)\}]| \leq \frac{Ch_l^{1/2}}{N_l}.$$

Remark A.2. We note that, via [13, Proposition A.1.], a simple decomposition:

$$\frac{a}{b} - \frac{c}{d} = \frac{a-c}{b} + \frac{c[d-b]}{bd} \tag{34}$$

for any $(a, b, c, d) \in \mathbb{R}$, $b \neq 0$, $d \neq 0$ and Lemma A.1 one can show that

$$\mathbb{E}[(u^{M, N_M}(0, x_0) - u^M(0, x_0))^2] \leq \frac{C}{N_M} \quad (35)$$

for $C < \infty$ independent of N_M, l . In addition, in the proof of Proposition A.2 we have established that

$$|u^L(0, x_0) - u^*(0, x_0)| \leq Ch_L^{1/2} \quad (36)$$

where $C < \infty$ does not depend on l . (36) can be obtained using the bound for (39) ($l = L$), (34) and Lemma A.1.

Proof of Theorem 4.1. Using the C_2 -inequality,

$$\begin{aligned} & \mathbb{E} \left[\left(\left\{ u^{M:L, N_{M:L}}(0, x_0) - u^L(0, x_0) \right\} + u^L(0, x_0) - u^*(0, x_0) \right)^2 \right] \leq \\ & 2 \left(\mathbb{E} \left[\left(u^{M:L, N_{M:L}}(0, x_0) - u^L(0, x_0) \right)^2 \right] + |u^L(0, x_0) - u^*(0, x_0)|^2 \right). \end{aligned} \quad (37)$$

First, apply the C_2 -inequality to the variance term (the bias is the L.H.S. of (36) and hence the variance is the left term on the R.H.S. of (37)), splitting $u^{M, N_M}(0, x_0) - u^M(0, x_0)$ and the other terms. Second, apply Theorem A.1, Proposition A.1 and (35) to the variance terms and (36) to the bias term. This allows one to complete the proof. \square

A.3 Proofs for Theorem A.1

Below for $i_1 \in \{1, \dots, m\}$ we write the i_1^{th} -element of m -vector $\varphi_l(z_{h_l: M_l h_l})$ as $\varphi_l(z_{h_l: M_l h_l})_{i_1}$. For $(i_1, i_2) \in \{1, \dots, m\} \times \{1, \dots, d\}$, we write the $i_1^{\text{th}}, i_2^{\text{th}}$ element of $\alpha(z)$ as $\alpha(z)_{i_1 i_2}$. For ease of notation set

$$T(i_1, l, q) :=$$

$$\mathbb{E}_{\pi^{l, l-1}} \left[\left(\varphi_l(Z_{h_l: M_l h_l}(l))_{i_1} \check{H}^{l, 1}(Z_{h_l: 1}(l), Z_{h_{l-1}: 1}(l-1)) - \varphi_{l-1}(Z_{h_{l-1}: M_{l-1} h_{l-1}}(l-1))_{i_1} \check{H}^{l, 2}(Z_{h_l: 1}(l), Z_{h_{l-1}: 1}(l-1)) \right)^q \right]^{3-q}.$$

Proposition A.2. *Assume (A1). Then for any $i_1 \in \{1, \dots, m\}$, $q \in \{1, 2\}$, there exist a $C < +\infty$ such that for any $l \in \{M+1, \dots, L\}$*

$$T(i_1, l, q) \leq Ch_l.$$

Proof. We give the proof for $q = 2$, the proof for the case $q = 1$ follows by Jensen's inequality.

Let $\mathcal{Z}^{l, l-1}$ denote the normalizing constant of $\pi^{l, l-1}$. By Lemma A.1 it easily follows that

$$\mathcal{Z}^{l, l-1} \geq C.$$

Thus it follows that

$$\begin{aligned} T(i_1, l, 2) & \leq C \int_{\mathbb{R}^{d(2l+2l-1)}} \left(\varphi_l(z_{h_l: M_l h_l}(l))_{i_1} \prod_{k=0}^{h_l^{-1}} G_k^l(z_{kh_l}(l)) - \right. \\ & \left. \varphi_{l-1}(z_{h_{l-1}: M_{l-1} h_{l-1}}(l-1))_{i_1} \prod_{k=0}^{h_{l-1}^{-1}} G_k^{l-1}(z_{kh_{l-1}}(l-1)) \right)^2 p^{l, l-1}(d(z_{h_l: 1}(l), z_{h_{l-1}: 1}(l-1))). \end{aligned} \quad (38)$$

Define

$$T_1(i_1, l) := \mathbb{E}_D \left[\left(\varphi_l(z_{h_l: M_l h_l}(l))_{i_1} \prod_{k=0}^{h_l^{-1}} G_k^l(z_{kh_l}(l)) - \left(\sum_{i_2=1}^d \int_0^r \alpha(Z_s)_{i_1 i_2} dW_s(i_2) \right) \exp \left\{ -\frac{1}{\gamma} \left(\phi(Z_1) + \int_0^1 \ell(Z_s) ds \right) \right\} \right)^2 \right]$$

where we are denoting expectations w.r.t. the law of the diffusion (8) as \mathbb{E}_D and $W_s(i_2)$ is the i_2^{th} -element of the Brownian motion in (8). Then it is clear that the integral on the R.H.S. of (38) is upper-bounded by

$$2(T_1(i_1, l) + T_1(i_1, l-1)).$$

Hence we focus upon $T_1(i_1, l)$ to conclude our result.

We have

$$\begin{aligned}
T_1(i_l, l) &\leq 2\mathbb{E}_D \left[\left(\varphi_l(z_{h_l:M_l h_l}(l))_{i_1} \prod_{k=0}^{h_l^{-1}} G_k^l(z_{kh_l}(l)) - \sum_{i_2=1}^d \int_0^r \alpha(Z_s)_{i_1 i_2} dW_s(i_2) \exp\left\{-\frac{1}{\gamma}(\phi(Z_1) + h_l \sum_{k=1}^{h_l^{-1}-1} \ell(Z_{kh_l}))\right\} \right)^2 \right] + \\
&2\mathbb{E}_D \left[\left(\left(\sum_{i_2=1}^d \int_0^r \alpha(Z_s)_{i_1 i_2} dW_s(i_2) \right) \exp\left\{-\frac{1}{\gamma}(\phi(Z_1) + h_l \sum_{k=1}^{h_l^{-1}-1} \ell(Z_{kh_l}))\right\} - \exp\left\{-\frac{1}{\gamma}(\phi(Z_1) + \int_0^1 \ell(Z_s) ds)\right\} \right)^2 \right].
\end{aligned} \tag{39}$$

We deal with the two expectations on the R.H.S. of (39) individually. For the first term on the R.H.S. we have that

$$\begin{aligned}
&\mathbb{E}_D \left[\left(\varphi_l(z_{h_l:M_l h_l}(l))_{i_1} \prod_{k=0}^{h_l^{-1}} G_k^l(z_{kh_l}(l)) - \sum_{i_2=1}^d \int_0^r \alpha(Z_s)_{i_1 i_2} dW_s(i_2) \exp\left\{-\frac{1}{\gamma}(\phi(Z_1) + h_l \sum_{k=1}^{h_l^{-1}-1} \ell(Z_{kh_l}))\right\} \right)^2 \right] \leq \\
&2\mathbb{E}_D \left[\left(\left(\varphi_l(z_{h_l:M_l h_l}(l))_{i_1} - \sum_{i_2=1}^d \int_0^r \alpha(Z_s)_{i_1 i_2} dW_s(i_2) \right) \left(\prod_{k=0}^{h_l^{-1}} G_k^l(z_{kh_l}(l)) \right) \right)^2 \right] + \\
&2\mathbb{E}_D \left[\left(\sum_{i_2=1}^d \int_0^r \alpha(Z_s)_{i_1 i_2} dW_s(i_2) \right)^2 \left(\prod_{k=0}^{h_l^{-1}} G_k^l(z_{kh_l}(l)) - \exp\left\{-\frac{1}{\gamma}(\phi(Z_1) + h_l \sum_{k=1}^{h_l^{-1}-1} \ell(Z_{kh_l}))\right\} \right)^2 \right].
\end{aligned}$$

Application of Lemmata A.2-A.3 yields the upper-bound of Ch_l . For the second term on the R.H.S. of (39), one can use Lemma A.4. Hence the proof is completed. \square

Lemma A.1. *Assume (A1). Then there exists a $0 < \underline{C} < \bar{C} < +\infty$ such that for any $l \in \{M, \dots, L\}$, $z_{h_l:1} \in \mathbb{R}^{d^{2^l}}$*

$$\underline{C} \leq \prod_{k=0}^{h_l^{-1}} G_k^l(z_{kh_l}) \leq \bar{C}$$

and for any $l \in \{M+1, \dots, L\}$, $(z_{h_l:1}(l), z_{h_{l-1}:1}(l)) \in \mathbb{R}^{d(2^{2^l} + 2^{2^{l-1}})}$

$$\underline{C} \leq \left\{ \prod_{k \in K_1^l} \check{G}_k^l(z_{kh_l}(l), z_{a_k(l)}(l-1)) \right\} \left\{ \prod_{k \in K_2^l} \check{G}_k^l(z_{kh_l}(l)) \right\} \leq \bar{C}.$$

Proof. Throughout $0 < \underline{C} < \bar{C} < +\infty$ are finite constants that do not depend upon l and whose value may change on each appearance. We note that for any $k \in \{0, \dots, h_l^{-1} - 1\}$, any $l \in \{M, \dots, L\}$ and any $z \in \mathbb{R}^d$

$$\exp\left\{-\frac{h_l}{\gamma} \sup_z |\ell(z)|\right\} \leq G_k^l(z) \leq \exp\left\{-\frac{h_l}{\gamma} \inf_z \ell(z)\right\}.$$

Clearly for any $z \in \mathbb{R}^d$ $\underline{C} \leq G_{h_l^{-1}}^l(z) \leq \bar{C}$, hence it follows that for any $l \in \{M, \dots, L\}$, $z_{h_l:1} \in \mathbb{R}^{d^{2^l}}$

$$\underline{C} \leq \prod_{k=0}^{h_l^{-1}} G_k^l(z_{kh_l}) \leq \bar{C}.$$

The second result is established using the relationship between \check{G}_k^l and G_k^l . \square

Recall we are denoting expectations w.r.t. the law of the diffusion (8) as \mathbb{E}_D and $W_s(i_2)$ is the i_2^{th} -element of the Brownian motion in (8).

Lemma A.2. *Assume (A1). Then for any $i_1 \in \{1, \dots, m\}$ there exist a $C < +\infty$ such that for any $l \in \{M, \dots, L\}$*

$$\mathbb{E}_D \left[\left(\left(\varphi_l(z_{h_l:M_l h_l}(l))_{i_1} - \sum_{i_2=1}^d \int_0^r \alpha(Z_s)_{i_1 i_2} dW_s(i_2) \right) \left(\prod_{k=0}^{h_l^{-1}} G_k^l(z_{kh_l}(l)) \right) \right)^2 \right] \leq Ch_l.$$

Proof. By Lemma A.1 we need only deal with

$$\mathbb{E}_D \left[\left(\varphi_l(z_{h_l; M_l h_l}(l))_{i_1} - \sum_{i_2=1}^d \int_0^r \alpha(Z_s)_{i_1 i_2} dW_s(i_2) \right)^2 \right].$$

This latter term, via the C_2 -inequality is upper-bounded by

$$\begin{aligned} & 2\mathbb{E}_D \left[\left(\varphi_l(z_{h_l; M_l h_l}(l))_{i_1} - \sum_{i_2=1}^d \sum_{k=1}^{M_l} \alpha(Z_{(k-1)h_l})_{i_1 i_2} W_k^l(i_2) \right)^2 \right] + \\ & 2\mathbb{E}_D \left[\left(\sum_{i_2=1}^d \sum_{k=1}^{M_l} \alpha(Z_{(k-1)h_l})_{i_1 i_2} W_k^l(i_2) - \sum_{i_2=1}^d \int_0^r \alpha(Z_s)_{i_1 i_2} dW_s(i_2) \right)^2 \right]. \end{aligned}$$

We treat these two terms independently, calling them T_1 and T_2 respectively.

Term: T_1 . By repeated use of the C_2 -inequality, T_1 is upper-bounded by

$$\begin{aligned} & C \sum_{i_2=1}^d \mathbb{E} \left[\left(\sum_{k=1}^{M_l} (\alpha(Z_{(k-1)h_l}(l))_{i_1 i_2} - \alpha(Z_{(k-1)h_l})_{i_1 i_2}) W_k^l(i_2) \right)^2 \right] = \\ & C \sum_{i_2=1}^d \sum_{k=1}^{M_l} \mathbb{E} \left[\left((\alpha(Z_{(k-1)h_l}(l))_{i_1 i_2} - \alpha(Z_{(k-1)h_l})_{i_1 i_2}) W_k^l(i_2) \right)^2 \right] \leq \\ & C \sum_{i_2=1}^d \sum_{k=1}^{M_l} \mathbb{E} \left[\left| Z_{(k-1)h_l}(l) - Z_{(k-1)h_l} \right|^2 W_k^l(i_2)^2 \right] \end{aligned}$$

where we have used the Lipschitz property of α to go to the last line. Splitting the expectation of the summand using Cauchy Schwarz we have

$$T_1 \leq C \sum_{i_2=1}^d \sum_{k=1}^{M_l} \mathbb{E} \left[\left| Z_{(k-1)h_l}(l) - Z_{(k-1)h_l} \right|^4 \right]^{1/2} \mathbb{E} [W_k^l(i_2)^4]^{1/2}.$$

Then using standard results from Euler-discretization of diffusions (see e.g. [20]) and Gaussian distributions:

$$\mathbb{E} \left[\left| Z_{(k-1)h_l}(l) - Z_{(k-1)h_l} \right|^4 \right]^{1/2} = \mathcal{O}(h_l) \quad \text{and} \quad \mathbb{E} [W_k^l(i_2)^4]^{1/2} = \mathcal{O}(h_l)$$

hence

$$T_1 \leq C \sum_{i_2=1}^d M_l h_l^2 \leq C h_l.$$

Term: T_2 . By repeated use of the C_2 -inequality, T_2 is upper-bounded by

$$\begin{aligned} & C \sum_{i_2=1}^d \mathbb{E}_D \left[\left(\sum_{k=1}^{M_l} \int_{(k-1)h_l}^{kh_l} \{ \alpha(Z_{(k-1)h_l})_{i_1 i_2} - \alpha(Z_s)_{i_1 i_2} \} dW_s(i_2) \right)^2 \right] = \\ & C \sum_{i_2=1}^d \sum_{k=1}^{M_l} \mathbb{E}_D \left[\left(\int_{(k-1)h_l}^{kh_l} \{ \alpha(Z_{(k-1)h_l})_{i_1 i_2} - \alpha(Z_s)_{i_1 i_2} \} dW_s(i_2) \right)^2 \right]. \end{aligned}$$

Using the Ito isometry formula, clearly

$$T_2 \leq C \sum_{i_2=1}^d \sum_{k=1}^{M_l} \int_{(k-1)h_l}^{kh_l} \mathbb{E}_D \left[\left(\alpha(Z_{(k-1)h_l})_{i_1 i_2} - \alpha(Z_s)_{i_1 i_2} \right)^2 \right] ds$$

Then using the Lipschitz property of α and then standard results for diffusion processes

$$T_2 \leq C h_l.$$

The proof is hence completed by the above arguments. \square

Lemma A.3. Assume (A1). Then for any $i_1 \in \{1, \dots, m\}$ there exist a $C < +\infty$ such that for any $l \in \{M, \dots, L\}$

$$\mathbb{E}_D \left[\left(\sum_{i_2=1}^d \int_0^r \alpha(Z_s)_{i_1 i_2} dW_s(i_2) \right)^2 \left(\prod_{k=0}^{h_l^{-1}} G_k^l(z_{kh_l}(l)) - \exp\left\{-\frac{1}{\gamma}(\phi(Z_1) + h_l \sum_{k=1}^{h_l^{-1}-1} \ell(Z_{kh_l}))\right\} \right)^2 \right] \leq Ch_l.$$

Proof. We begin by splitting the expectation via Cauchy Schwarz to yield the upper-bound

$$\mathbb{E}_D \left[\left(\sum_{i_2=1}^d \int_0^r \alpha(Z_s)_{i_1 i_2} dW_s(i_2) \right)^4 \right]^{1/2} \mathbb{E}_D \left[\left(\prod_{k=0}^{h_l^{-1}} G_k^l(z_{kh_l}(l)) - \exp\left\{-\frac{1}{\gamma}(\phi(Z_1) + h_l \sum_{k=1}^{h_l^{-1}-1} \ell(Z_{kh_l}))\right\} \right)^2 \right)^{1/2}. \quad (40)$$

By (a classic variation of) the Burkholder-Gundy-Davis inequality, it follows that

$$\mathbb{E}_D \left[\left(\sum_{i_2=1}^d \int_0^r \alpha(Z_s)_{i_1 i_2} dW_s(i_2) \right)^4 \right]^{1/2} \leq C$$

hence we consider the right-hand expectation in (40). Using the Lipschitz property of e^{-x} for bounded x , it follows that

$$\begin{aligned} & \mathbb{E}_D \left[\left(\prod_{k=0}^{h_l^{-1}} G_k^l(z_{kh_l}(l)) - \exp\left\{-\frac{1}{\gamma}(\phi(Z_1) + h_l \sum_{k=1}^{h_l^{-1}-1} \ell(Z_{kh_l}))\right\} \right)^2 \right)^{1/2} \leq \\ & C \mathbb{E}_D \left[\left(\phi(Z_1(l)) - \phi(Z_1) + h_l \sum_{k=1}^{h_l^{-1}-1} (\ell(Z_{kh_l}(l)) - \ell(Z_{kh_l})) \right)^4 \right)^{1/2}. \end{aligned}$$

Using the Lipschitz property of ϕ and ℓ it clearly follows that

$$\mathbb{E}_D \left[\left(\phi(Z_1(l)) - \phi(Z_1) + h_l \sum_{k=1}^{h_l^{-1}-1} (\ell(Z_{kh_l}(l)) - \ell(Z_{kh_l})) \right)^4 \right)^{1/2} \leq \mathbb{E}_D \left[\max_{k \in \{1, \dots, h_l^{-1}\}} |Z_{kh_l}(l) - Z_{kh_l}|^4 \right]^{1/2}.$$

Hence using standard results for Euler approximations of diffusion processes

$$\mathbb{E}_D \left[\left(\prod_{k=0}^{h_l^{-1}} G_k^l(z_{kh_l}(l)) - \exp\left\{-\frac{1}{\gamma}(\phi(Z_1) + h_l \sum_{k=1}^{h_l^{-1}-1} \ell(Z_{kh_l}))\right\} \right)^2 \right)^{1/2} \leq Ch_l$$

and one can conclude the proof by the above arguments. \square

Lemma A.4. Assume (A1). Then for any $i_1 \in \{1, \dots, m\}$ there exist a $C < +\infty$ such that for any $l \in \{M, \dots, L\}$

$$\begin{aligned} & \mathbb{E}_D \left[\left(\left(\sum_{i_2=1}^d \int_0^r \alpha(Z_s)_{i_1 i_2} dW_s(i_2) \right) \exp\left\{-\frac{1}{\gamma}(\phi(Z_1) + h_l \sum_{k=1}^{h_l^{-1}-1} \ell(Z_{kh_l}))\right\} \right. \right. \\ & \left. \left. - \exp\left\{-\frac{1}{\gamma}(\phi(Z_1) + \int_0^1 \ell(Z_s) ds)\right\} \right)^2 \right] \leq Ch_l. \end{aligned}$$

Proof. As in the proof of Lemma A.3 split the expectation via Cauchy Schwarz to yield the upper-bound

$$\mathbb{E}_D \left[\left(\sum_{i_2=1}^d \int_0^r \alpha(Z_s)_{i_1 i_2} dW_s(i_2) \right)^4 \right]^{1/2} \mathbb{E}_D \left[\left(\exp\left\{-\frac{1}{\gamma}(\phi(Z_1) + h_l \sum_{k=1}^{h_l^{-1}-1} \ell(Z_{kh_l}))\right\} - \exp\left\{-\frac{1}{\gamma}(\phi(Z_1) + \int_0^1 \ell(Z_s) ds)\right\} \right)^4 \right)^{1/2}.$$

Again (as in the proof of Lemma A.3) by the Burkholder-Gundy-Davis inequality the first expectation is $\mathcal{O}(1)$ so we concentrate upon the second expectation (call it T_1). Using the Lipschitz property of e^{-x} for bounded x , it follows that

$$\begin{aligned} T_1 & \leq C \mathbb{E}_D \left[\left(h_l \sum_{k=1}^{h_l^{-1}-1} \ell(Z_{kh_l}) - \int_0^1 \ell(Z_s) ds \right)^4 \right]^{1/2} \\ & = C \mathbb{E}_D \left[\left(\sum_{k=1}^{h_l^{-1}-1} \int_{(k-1)h_l}^{kh_l} \{\ell(Z_{kh_l}) - \ell(Z_s)\} ds + \int_{(1-h_l)}^1 \ell(Z_s) ds \right)^4 \right)^{1/2}. \end{aligned}$$

Then via the Minkowski inequality

$$T_1 \leq C \left[\sum_{k=1}^{h_l^{-1}-1} \mathbb{E}_D \left[\left(\int_{(k-1)h_l}^{kh_l} \{\ell(Z_{kh_l}) - \ell(Z_s)\} ds \right)^4 \right]^{1/4} + Ch_l \right]^2$$

where we have used the fact that ℓ is bounded. Then applying Jensen with standard results in diffusions, yields $T_1 \leq Ch_l$. \square

A.4 Proof of Proposition A.1

Proof of Proposition A.1. We make the defintions:

$$\begin{aligned} a^N &:= \frac{1}{N_l} \sum_{i=1}^{N_l} \varphi_l(z_{h_l:M_l h_l}^i(l)) \check{H}^{l,1}(z_{h_l:1}^i(l), z_{h_{l-1}:1}^i(l-1)) \\ b^N &:= \frac{1}{N_l} \sum_{i=1}^{N_l} \check{H}^{l,1}(z_{h_l:1}^i(l), z_{h_{l-1}:1}^i(l-1)) \\ c^N &:= \frac{1}{N_l} \sum_{i=1}^{N_l} \varphi_{l-1}(z_{h_{l-1}:M_{l-1} h_{l-1}}^i(l-1)) \check{H}^{l,2}(z_{h_l:1}^i(l), z_{h_{l-1}:1}^i(l-1)) \\ d^N &:= \frac{1}{N_l} \sum_{i=1}^{N_l} \check{H}^{l,2}(z_{h_l:1}^i(l), z_{h_{l-1}:1}^i(l-1)) \\ a &:= \mathbb{E}_{\pi^{l,l-1}}[\varphi_l(Z_{h_l:M_l h_l}(l)) \check{H}^{l,1}(Z_{h_l:1}(l), Z_{h_{l-1}:1}(l-1))] \\ b &:= \mathbb{E}_{\pi^{l,l-1}}[\check{H}^{l,1}(Z_{h_l:1}(l), Z_{h_{l-1}:1}(l-1))] \\ c &:= \mathbb{E}_{\pi^{l,l-1}}[\varphi_{l-1}(Z_{h_{l-1}:M_{l-1} h_{l-1}}(l-1)) \check{H}^{l,2}(Z_{h_l:1}(l), Z_{h_{l-1}:1}(l-1))] \\ d &:= \mathbb{E}_{\pi^{l,l-1}}[\check{H}^{l,2}(Z_{h_l:1}(l), Z_{h_{l-1}:1}(l-1))]. \end{aligned}$$

Then component-wise (from here, our calculations should be considered component-by-component):

$$r\left(u^{l,N_l}(0, x_0) - u^{l-1,N_l}(0, x_0) - \{u^l(0, x_0) - u^{l-1}(0, x_0)\}\right) = \frac{a^N}{b^N} - \frac{c^N}{d^N} - \left(\frac{a}{b} - \frac{c}{d}\right).$$

By [12, Lemma D.5]

$$\begin{aligned} &\frac{a^N}{b^N} - \frac{c^N}{d^N} - \left(\frac{a}{b} - \frac{c}{d}\right) = \\ &\frac{1}{b^N}(a^N - c^N - (a - c)) - \frac{c^N}{b^N d^N}(b^N - d^N - (b - d)) + \frac{1}{b^N b}(b - b^N)(a - c) + \\ &\frac{c}{d^N b d}(d^N - d)(b - d) - \frac{1}{b^N d^N}(c^N - c)(b - d) + \frac{c}{b^N d^N b}(b^N - b)(b - d). \end{aligned}$$

The first two terms on the R.H.S. and the last four terms on the R.H.S. can be treated using similar calculations, so we only consider

$$T_1 := \frac{1}{b^N}(a^N - c^N - (a - c)) \quad \text{and} \quad T_2 := \frac{1}{b^N b}(b - b^N)(a - c).$$

Term: T_1 . We have

$$T_1 = \left(\frac{1}{b^N} - \frac{1}{b}\right)(a^N - c^N - (a - c)) + \frac{1}{b}(a^N - c^N - (a - c)).$$

On taking expectations w.r.t. the law of the simulated Markov chain, we have

$$\mathbb{E}[T_1] = \mathbb{E}\left[\left(\frac{1}{b^N} - \frac{1}{b}\right)(a^N - c^N - (a - c))\right]$$

as the chain is started in stationarity. Applying the Cauchy-Schwarz inequality:

$$|\mathbb{E}[T_1]| \leq \mathbb{E}\left[\left(\frac{1}{b^N} - \frac{1}{b}\right)^2\right]^{1/2} \mathbb{E}[(a^N - c^N - (a - c))^2]^{1/2}.$$

By using a similar result to [13, Proposition A.1] and by Proposition A.2

$$\mathbb{E}[(a^N - c^N - (a - c))^2]^{1/2} \leq \frac{Ch_l^{1/2}}{N_l^{1/2}}.$$

By standard results for uniformly ergodic Markov chains (e.g. the result in [13, Proposition A.1]), along with Lemma A.1

$$\mathbb{E}\left[\left(\frac{1}{b^N} - \frac{1}{b}\right)^2\right]^{1/2} \leq \frac{C}{N_l^{1/2}}$$

and hence

$$|\mathbb{E}[T_1]| \leq \frac{Ch_l^{1/2}}{N_l}.$$

Term: T_2 . We have

$$T_2 = \left(\frac{1}{b^N b} - \frac{1}{b^2}\right)(b - b^N)(a - c) + \frac{1}{b^2}(b - b^N)(a - c)$$

and taking expectations as for T_1

$$\mathbb{E}[T_2] = \mathbb{E}\left[\left(\frac{1}{b^N b} - \frac{1}{b^2}\right)(b - b^N)(a - c)\right].$$

Hence, by Cauchy-Schwarz

$$|\mathbb{E}[T_2]| \leq |a - c| \mathbb{E}\left[\left(\frac{1}{b^N b} - \frac{1}{b^2}\right)^2\right]^{1/2} \mathbb{E}[(b - b^N)^2]^{1/2}.$$

By Proposition A.2 $|a - c| \leq Ch_l^{1/2}$ and again, standard results for uniformly ergodic Markov chains, along with Lemma A.1

$$\mathbb{E}\left[\left(\frac{1}{b^N b} - \frac{1}{b^2}\right)^2\right]^{1/2} \mathbb{E}[(b - b^N)^2]^{1/2} \leq \frac{C}{N_l}$$

and hence

$$|\mathbb{E}[T_2]| \leq \frac{Ch_l^{1/2}}{N_l}.$$

□

References

- [1] ANDRIEU, C., DOUCET, A. & HOLENSTEIN, R. (2010). Particle Markov chain Monte Carlo methods (with discussion). *J. R. Statist. Soc. Ser. B*, **72**, 269–342.
- [2] ARNOLD, L. (1974). *Stochastic Differential Equations: Theory and Applications*. Wiley: New York.
- [3] BERTOLI, F. & BISHOP, A. N. (2018). Nonlinear stochastic receding horizon control: Stability, robustness and Monte Carlo methods for control approximation. *Int'l J. Cont.*, **91**, 2387–2402.
- [4] BERTOLI, F. & BISHOP, A. N. (2018). An error analysis in the limit approximation in path integral control. Technical Report.
- [5] FLEMING, W. H. & MITTER, S. K. (1982). Optimal control and nonlinear filtering for nondegenerate diffusion processes. *Stochastics*, **8**(1), 63–77.
- [6] FLEMING, W. H. & SONER, H. M. (2006). *Controlled Markov Processes and Viscosity Solutions*. 2nd edition, Springer: New York.
- [7] FRANKS, J., JASRA, A., LAW, K. J. H. & VIHOLA, M. (2018). Unbiased inference for discretely observed hidden Markov model diffusions. arXiv:1807.10259.
- [8] GILES, M. B. (2008). Multilevel Monte Carlo path simulation. *Op. Res.*, **56**, 607–617.
- [9] GILES, M. B. (2015). Multilevel Monte Carlo methods. *Acta Numerica* **24**, 259–328.

- [10] HEINRICH, S. (2001). Multilevel Monte Carlo methods. In *Large-Scale Scientific Computing*, (eds. S. Margenov, J. Wasniewski & P. Yalamov), Springer: Berlin.
- [11] JASRA, A. & DOUCET, A. (2009). Sequential Monte Carlo for diffusion processes, *Proc. Roy. Soc. A.* **465**, 3709–3727.
- [12] JASRA, A., KAMATANI, K., LAW K. J. H. & ZHOU, Y. (2017). Multilevel particle filters. *SIAM J. Numer. Anal.*, **55**, 3068–3096.
- [13] JASRA, A., KAMATANI, K., LAW, K. & ZHOU, Y. (2018). Bayesian static parameter estimation for partially observed diffusions via multilevel Monte Carlo. *SIAM J. Sci. Comp.*, **40**, A887–A902.
- [14] KAPPEN, H. J. (2005). Path integrals and symmetry breaking for optimal control theory. *J. Stat. Mech.*, **11**, 11011.
- [15] KAPPEN, H. J. (2005). Linear theory for control of nonlinear stochastic systems. *Physical Review Letters*, **95**(20), 200201.
- [16] KAPPEN, H. J., GÓMEZ, V., & OPPER, M. (2012). Optimal control as a graphical model inference problem. *Machine learning*, **87**(2), 159–182.
- [17] KAPPEN, H. J. & RUIZ, H. C. (2016). Adaptive importance sampling for control and inference. *Journal of Statistical Physics*, **162**(5), 1244–1266.
- [18] KRYLOV, N. V. (1972). Control of a solution of a stochastic integral equation. *Theory Probab. Appl.*, **17**, 114–130.
- [19] KRYLOV, N. V. (2008). *Controlled Diffusion Processes*. Springer: New York.
- [20] KURTZ, T., & PROTTER, P. (1991). Wong-Zakai corrections, random evolutions and simulation schemes for SDEs. In *Stochastic Analysis*, (Meyer-Wolf, Merzbach & Schwarz, Eds). Academic Press: New York.
- [21] MENCHÓN, S. A., & KAPPEN, H. J. (2018). Learning effective state-feedback controllers through efficient multilevel importance samplers. *International Journal of Control*, 1–8.
- [22] RUIZ, H. C. & KAPPEN, H. J. (2017). Particle smoothing for hidden diffusion processes: Adaptive path integral smoother. *IEEE Transactions on Signal Processing*, **65**(12), 3191–3203.
- [23] THEODOROU, E. A., BUCHLI, J., & SCHAAL, S. (2010). A generalized path integral control approach to reinforcement learning. *J. Mach. Learn. Res.*, **11**, 3137–3181.
- [24] THEODOROU, E. A. & TODOROV E. (2012). Relative entropy and free energy dualities: Connections to path integral and KL control. In *Proceedings 51st IEEE Conference on Decision and Control (CDC)*, 1466–1473.
- [25] THIJSSSEN, S. & KAPPEN, H. J. (2015). Path integral control and state-dependent feedback. *Phys. Rev. E*, **91**, 032104.
- [26] TORNATORE, E., VETRO, P., & BUCCELLATO, S. M. (2014). SIVR Epidemic Model with Stochastic Perturbation. *Neural Comp. Appl.* **24**(2), 309–315.
- [27] TOUZI, N. (2012). *Optimal Stochastic Control, Stochastic Target Problems and Backward Stochastic Differential Equations*. Springer: New York.
- [28] WITBOOI, P. J., MULLER, G. E., & VAN SCHALKWYK, G. J. (2015). Vaccination Control in a Stochastic SVIR Epidemic Model. *Comp. Math. Meth. Med.*, Article ID 271654.