Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications xxx (xxxx) xxx

stochastic processes and their applications

www.elsevier.com/locate/spa

Unbiased Optimal Stopping via the MUSE

Zhengqing Zhou[a,1], Guanyang Wang[b,1], Jose H. Blanchet[c,*], Peter W. Glynn[c]

[a] *Department of Mathematics, Stanford University, United States of America*
[b] *Department of Statistics, Rutgers University, United States of America*
[c] *Management Science and Engineering, Stanford University, United States of America*

Received 4 August 2021; received in revised form 1 July 2022; accepted 11 December 2022
Available online xxxx

Abstract

We propose a new unbiased estimator for estimating the utility of the optimal stopping problem. The MUSE, short for 'Multilevel Unbiased Stopping Estimator', constructs the unbiased Multilevel Monte Carlo (MLMC) estimator at every stage of the optimal stopping problem in a backward recursive way. In contrast to traditional sequential methods, the MUSE can be implemented in parallel. We prove the MUSE has finite variance, finite computational complexity, and achieves ε-accuracy with $O(1/\varepsilon^2)$ computational cost under mild conditions. We demonstrate MUSE empirically in an option pricing problem involving a high-dimensional input and the use of many parallel processors.

MSC: 62C05; 60G40; 62L15

Keywords: Multilevel monte carlo; Unbiased estimator; Optimal stopping; Parallel computing

1. Introduction

It is a pleasure to contribute to this special issue in honor of Prof. Larry Shepp, whose work has significantly impacted a wide range of scientific disciplines. This paper focuses on optimal stopping problems, an area of stochastic control in which Prof. Shepp contributed deeply both in terms of theory and applications. in particular in connection to mathematical finance problems; see, for example, [11,14,36,37].

* Corresponding author.
E-mail address: jblanche@stanford.edu (J.H. Blanchet).
[1] Zhengqing Zhou and Guanyang Wang have contributed equally to this paper.

https://doi.org/10.1016/j.spa.2022.12.007
0304-4149/

Our goal in this paper is on designing Monte Carlo methods for solving optimal stopping problems which can be easily parallelized because the estimators that we produce are unbiased, are applicable even in non-Markovian problems and have finite variance. We are not aware of any other Monte Carlo estimators for optimal stopping problems which share these properties.

Monte Carlo methods are ubiquitously used for estimating high dimensional numerical integrals or statistics arising in every computation-related subject. However, vanilla Monte Carlo estimators may produce systematic bias in many practical applications (in particular those involving optimization). The presence of such systematic bias precludes the direct use of parallel computing architectures. Consider the following toy example in a two-stage optimal stopping problem. Suppose one is able to simulate the two-stage process (X_1, X_2) and is interested in estimating the utility

$$U := \mathbb{E}\left[\max\{f(X_1), \mathbb{E}\left[f(X_2) \mid X_1\right]\}\right],$$

where f is some integrable reward function. Then it is not hard to show the vanilla Monte Carlo estimator $\widehat{U}$ will systematically overestimate U by the Jensen's inequality.

To address the issue of bias, the design of unbiased Monte Carlo estimators has recently attracted much attention [7–9,23,25,29,33–35,39–41] in operations research, statistics, and machine learning communities. Many existing debiasing techniques are closely related to the Multilevel Monte Carlo (MLMC) framework developed by Heinrich and Giles [17,18,20,21,24] where a sequence of biased but increasingly accurate estimators are used to estimate the functionals of stochastic processes described by stochastic differential equations. Unbiased estimators have been successfully developed and employed in the context of stochastic approximation, Markov chain Monte Carlo estimation and convergence diagnosis, quantile estimation, and so on. In many settings, unbiased estimators provide a promising direction to efficient parallel implementation and better uncertainty quantification.

In this paper, we study the discrete-time, finite-horizon optimal stopping problem, a fundamental problem that can be found in areas including economics, operations research, and financial engineering. Consider the optimal stopping problem with underlying process $(X_1, \ldots, X_T)$ and reward function f. We are interested in computing the expected utility of the optimal strategy:

$$U_T := \sup_{\tau \in \mathcal{T}_T} \mathbb{E}\left[f\left(X_\tau\right)\right], \tag{1}$$

where $\mathcal{T}_T$ denotes the set of all the stopping times taking values in $\{1, \ldots, T\}$. Following the standard optimal stopping theory, we can define the Snell envelope by

$$U_{T-k} := \sup_{\tau \in \mathcal{T}_{k+1,T}} \mathbb{E}\left[f\left(X_\tau\right) \mid \mathcal{F}_k\right],$$

where $\mathcal{T}_{k+1,T}$ denotes the set of stopping times satisfying $k+1 \le \tau \le T$, $k = 0, \ldots, T-1$ and $\mathcal{F}_k$ is the natural filtration at time k. The dynamical programming can be written as:

$$\begin{cases} U_1 = \mathbb{E}\left[f\left(X_T\right) \mid \mathcal{F}_{T-1}\right], \\ U_{T-k} = \mathbb{E}\left[\max\left\{f(X_{k+1}), U_{T-(k+1)}\right\} \mid \mathcal{F}_k\right], \qquad k = 0, \ldots, T-1. \end{cases}$$

In most practical cases, U_T cannot be analytically solved, and we therefore resort to Monte Carlo methods for an estimation. However, generating unbiased estimation for the utility of the optimal problem is known as a very difficult problem. Suppose one is able to simulate the whole process. Then the above dynamical programming backward recursion suggests a natural Monte Carlo estimator as follows. We sample tree-like paths of the whole process forward in

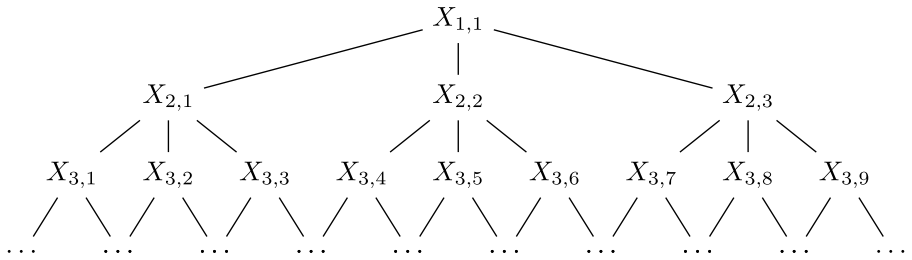

Fig. 1. Tree-like paths for Monte Carlo simulation. Here each node in the first two levels has three children which are *i.i.d.* sampled from the conditional distribution.

time and estimate each U_i backwards in time. The sampling procedure is illustrated in Fig. 1 where one samples many k-ary trees with height T. After sampling enough paths, we aggregate the samples from bottom to top in each layer as estimators of $U_1, \ldots, U_T$ respectively. This estimator is relatively easy to implement but has several limitations. Firstly, the estimator overestimates the utility even in the simplest case $T = 2$, let alone the general case. Secondly, as the estimation error propagates from one time horizon to another, the accuracy relies on a repeated T-limit, which is difficult to quantify. The above approach is the 'high-estimator' suggested in the seminal paper of Broadie and Glasserman [10]. The authors also use the similar idea to construct the 'low-estimator', and a confidence interval that covers the utility by combining the two estimators. In fact, the above authors conjectured that there is no general unbiased estimators for the optimal stopping problem, see page 1326–1327 in [10] for details. There are also regression-based Monte Carlo simulation methods, including the well-celebrated Longstaff–Schwartz [31] and Tsitsiklis–Van Roy [38] algorithms for option pricing, see also [16] for extensions. Both methods approximate the solution of the original problem by solving a sequence of regression problems in linearly parameterized subspaces. Albeit convenient to use, the approximation error and the unavoidable bias still cause concerns for both parallel implementation and uncertainty quantification.

In this paper, we introduce a novel unbiased estimator — the Multilevel Unbiased Stopping Estimator (MUSE[2]) for the optimal stopping problem described in (1). Our estimator is inspired by the randomized Multilevel Monte Carlo estimator described in [8,9]. The MUSE is easy to implement and enjoys both finite variance and finite expected computational complexity. The computational cost to achieve ε-accuracy is $O(1/\varepsilon^2)$, which matches the optimal rate from the Central Limit Theorem (CLT). Our empirical studies suggest that the MUSE scales well with the dimensionality of the underlying process — which is often viewed as a bottleneck of classical regression-based methods. As extra byproducts, we construct confidence intervals for the utility and propose a natural algorithm to determine the optimal stopping time based on the MUSE.

We emphasize that our techniques are of interest in randomized Multilevel Monte Carlo as we relax smoothness assumptions imposed in [8]. In summary, the MUSE can be viewed as a multi-stage extension of the MLMC estimator. In two-stage problems, our estimator (Algorithm 1) has the same expression as the randomized MLMC estimator. However, the general optimal stopping problem is defined in a recursive way, and thus the MLMC estimator

[2] In ancient Greek mythology, the Muses are the nine goddesses (the daughters of Zeus and Mnemosyne), who preside over literature, science, and the arts.

fails to directly apply. Multi-stage MUSE (Algorithm 2) generates data forwardly and then calls for the two-stage MUSE backwardly. Moreover, MUSE relaxes the technical assumptions of the randomized MLMC estimator. In [8], the randomized MLMC estimator is proposed to estimate $g(m_\mu)$ where m_μ is the mean of a probability measure μ and g is a locally twice differentiable function (Section 3, Assumption 2 in [8]), but the function of our interest is a multiple composition of the max function, which is not differentiable everywhere.

There is a large body of literature on solving the optimal stopping problem using the (non-randomized) MLMC methods. For the two-stage optimal stopping problem, a slightly more general version has been considered in [20]. The authors in [20] consider the problem of estimating:

$$\mathbb{E}[\max_d f_d(X, Y)] - \mathbb{E}[\max_d \mathbb{E}[f_d(X, Y) \mid X]] \tag{2}$$

where $\{f_d\}$ is a finite set of functions representing one's possible strategies. The quantity (2) represents the expected value of the partial information. The first term in (2) can be estimated via standard Monte Carlo approach, while the second term is much more complicated. Suppose $d = 2$, $f_1(X, Y) = f(X)$ and $f_2(X, Y) = \mathbb{E}[f(Y) \mid X]$ for the utility function f given at the beginning of our paper, then the second term in (2) is the expected utility of the two-stage optimal stopping problem. The methodology and theory of [20] are closely connected to our paper, though the focus is somewhat different. The main contribution of [20] is applying the (non-randomized) MLMC method to design low-bias estimators which attains ϵ^2 mean-squared error with $\mathcal{O}(\epsilon^2(\log \epsilon)^2)$ or $\mathcal{O}(\epsilon^2)$ computational complexity under different assumptions. In contrast, our effort is mostly on designing completely unbiased estimators with finite variance and finite computational cost. Technically, Assumption 2 in [20] is very close to our Assumption 4 and is posed for similar reasons (see 1.1 for detailed discussions), while our paper has a relatively weaker moment assumption (Assumption 2) than Assumption 1 on [19]. Besides [20], other related works for the two-stage optimal stopping problem include [19,27,28], and the references therein.

Non-randomized MLMC methods have also been used in the general multi-stage optimal stopping problems to obtain low-bias estimators. In Belomestny, Ladkau, and Schoenmakers [4], the authors designed MLMC estimators to improve the computational complexity of the standard Monte Carlo estimator based on their previous work [3,5]. The randomized MLMC idea has been briefly mentioned in the Ph.D. dissertation of Dickmann [13] but not explored in details.

The rest of this paper is organized as follows. After setting up the notation and describing the assumptions in Section 1.1, in Section 2 we introduce the MUSE and prove its theoretical properties. In Section 3 we showcase several applications of the MUSE through numerical examples. In Section 4 we conclude our paper with a discussion on its limitations and potential generalizations. Detailed proofs of Theorems 1 and 2 are deferred to Appendix A.

1.1. Notations and assumptions

Let $(\Omega, \mathcal{B}(\Omega))$ be a Polish space equipped with Borel σ-algebra $\mathcal{B}(\Omega)$. Let $(\mathcal{X}, \|\cdot\|)$ be a complete separable normed space equipped with Borel σ-algebra $\mathcal{B}(\mathcal{X})$. Given a fixed positive integer T and an adapted $\mathcal{X}-$ valued stochastic process $\{X_i\}_{i=1}^T$ with filtration $\{\mathcal{F}_i\}_{i=1}^T$, we denote by $\pi_{1:T}$ the joint distribution of $(X_1, \ldots, X_T)$. Given $X_i = x_i$ for $1 \le i \le k$, we denote by $\pi_{k+1:T}(\cdot \mid \{x_i\}_{i=1}^k)$ the conditional distribution of $(X_{k+1}, \ldots, X_T)$, and $\pi_{k+1}(\cdot \mid \{x_i\}_{i=1}^k)$ the marginal conditional distribution of X_{k+1}. We will use the convention $\{x_0\} = \varnothing$ and therefore

π_1 denotes the (unconditioned) marginal distribution of X_1. Let $f : \mathcal{X} \to \mathbb{R}$ be an integrable reward function. We denote by U_T the utility of the optimal stopping problem as described in (1) and the conditional utility

$$U_{T-k}(x_1, \ldots, x_k) := \sup_{\tau \in \mathcal{T}_{k+1,T}} \mathbb{E}_{\pi_{k+1:T}} \left[f(X_\tau) \mid \{x_i\}_{i=1}^k \right] \tag{3}$$

which corresponds to the utility function if one starts the $T-k$-stage optimal stopping problem after observing the first k outcomes $\{x_i\}_{i=1}^k$. For simplicity, we write $\pi_{k+1:T}$, π_{k+1}, and U_{T-k} when there is no confusion about the dependency on $\{x_i\}_{i=1}^k$. The geometric distribution taking values in $\{0, 1, \ldots\}$ with parameter r is denoted by $\mathrm{Geo}(r)$. Given a non-negative discrete random variable N, we denote its probability mass function by $p(n) := \mathbb{P}(N = n)$. In particular, $p_r(n)$ denotes $\mathbb{P}(\mathrm{Geo}(r) = n) = r(1-r)^n$.

Before formally describing the MUSE, we introduce the following assumption ensuring that the underlying process can be simulated. This assumption is standard and can be found in almost every Monte Carlo-based optimal stopping algorithm, such as [31,38].

Assumption 1 (*Path Simulation*)**.** Given fixed integers $0 \le k \le T-1, n \ge 1$ and a trajectory $\{x_i\}_{i=1}^k$, we have a simulator $\mathcal{S}$ which takes $\{x_i\}_{i=1}^k$ as inputs and outputs n *i.i.d.* samples $X_{k+1}(1), \ldots, X_{k+1}(n)$ with distribution $\pi_{k+1}(\cdot \mid \{x_i\}_{i=1}^k)$.

Besides the simulation assumption, we also introduce several technical assumptions. The MUSE can always be constructed as long as the simulation assumption is satisfied, but several desired properties such as finite variance are justified under these technical assumptions.

Assumption 2 (*Moment Assumption*)**.** There exists $\delta > 0$, such that $\mathbb{E}\left[\|X_i\|^{2+\delta}\right] < \infty$ for all $1 \le i \le T$.

Assumption 3 (*Linear Growth*)**.** $f : \mathcal{X} \to \mathbb{R}$ satisfies $|f(x)| \le L(1 + \|x\|)$ for some $L > 0$.

Assumption 4 (*Regularity Condition on Conditional Expectation*)**.** There exists a constant $C > 0$, such that

$$\mathbb{P}\left(|U_{T-k}(X_1, \ldots, X_k) - f(X_k)| < \varepsilon\right) < C\varepsilon$$

hold for all $\varepsilon > 0$ and $1 \le k \le T-1$.

Assumption 4 essentially requires the density of $U_{T-k}(X_1, \ldots, X_k) - f(X_k)$ to be bounded at zero. This can be verified directly when $(X_1, \ldots, X_T)$ is an independent process and each X_i has a bounded density. In general, we expect Assumption 4 to hold if the random vector $(X_1, \ldots, X_T)$ has a density, and the reward function f is smooth except for finitely many points.

To be specific, suppose $(X_1, \ldots, X_k)$ follows the joint distribution $\pi_{1:k}$, then in Assumption 4 we require

$$\int \mathbb{1}\left(|U_{T-k}(x_1, \ldots, x_k) - f(x_k)| < \varepsilon\right) \pi_{1:k}(dx_1, \ldots, dx_k) \le C\varepsilon.$$

Note that it is common to assume the underlying information to enjoy certain nice regularity in high-dimensional optimal stopping. For instance, similar regularity assumptions can be found in [4] (Proposition 3.1 and Theorem 3.4 (iv)), and [20] (Assumption 2).

We would like to emphasize that the main objective of our technical assumptions here is to bound the variance and expected computational cost simultaneously, which are crucial properties for the efficiency of any unbiased MLMC estimator. The finite variance enables the Central Limit Theorem, which in turn gives an asymptotically exact confidence interval of the true utility. The computational cost result allows us to control the expected amount of time for simulating one estimator. Putting the two properties together shows our estimator achieves ε^2-expected mean squared error within $O(1/\varepsilon^2)$ expected computational cost, see Corollary 1 for details. We also refer the readers to [8] and [39] for more discussions on the variance, the computational cost and their trade-off for unbiased MLMC methods.

2. Multilevel Unbiased Stopping Estimator (MUSE)

We present our main results in this section. We start with the two-stage MUSE in Section 2.1. The general/multi-stage MUSE is described in Section 2.2 and it is constructed by recursively calling the two-stage MUSE. Two related applications, including the construction of the confidence interval and an algorithm for finding the optimal stopping time, are discussed in Section 2.3.

2.1. MUSE for two-stage optimal stopping problems

Two-stage optimal stopping is a special and simplest non-trivial case among the finite-horizon optimal stopping problems. To build a better intuition for the MUSE, we start with describing the MUSE under this simplified setting, which serves as both a base case and motivation for the general estimator.

Given the bivariate distribution $\pi_{1:2}$ of (X_1, X_2). The utility of the two-stage optimal stopping problem can be written as:

$$U_2 = \mathbb{E}\left[\max\{f(X_1), \mathbb{E}\left[f(X_2) \mid X_1\right]\}\right] = \int_{\Omega} \max\{f(x_1), \mathbb{E}\left[f(X_2) \mid x_1\right]\}\pi_1(dx_1). \tag{4}$$

Therefore, after sampling $x_1 \sim \pi_1$ from the simulator $\mathcal{S}$, it suffices to construct an unbiased estimator of $g_{x_1}(\mathbb{E}\left[f(X_2) \mid x_1\right])$ where $g_{x_1}(a) := \max\{f(x_1), a\}$. It is also clear that a vanilla estimator $\max\{f(x_1), f(x_2)\}$ with $(x_1, x_2) \sim \pi_{1:2}$ will be biased here, as such an estimator has expectation

$$\int_{\Omega\times\Omega} \max\{f(x_1), f(x_2)\}\pi_1(dx_1)\pi_2(dx_2 \mid x_1) \geq \int_{\Omega} \max\{f(x_1), \mathbb{E}\left[f(X_2) \mid x_1\right]\}\pi_1(dx_1),$$

and therefore overestimates the utility. The debiasing strategy follows from the observation of [8,17]. Let $\{n_0, n_1, \ldots\}$ be an increasing sequence of positive integers. For each n_i, let $\overline{f(X_2)}_{n_i}$ be the empirical average of n_i *i.i.d.* samples of $f(X_2)$ with $X_2 \sim \pi_2(\cdot \mid \{x_1\})$. By virtue of the law of large numbers, we have $g_{x_1}(\mathbb{E}\left[f(X_2) \mid x_1\right]) = \lim_{k\to\infty} g_{x_1}(\overline{f(X_2)}_{n_k})$ almost surely. Then we can write $g_{x_1}(\mathbb{E}\left[f(X_2) \mid x_1\right])$ as the following telescoping summation:

$$\begin{aligned} g_{x_1}(\mathbb{E}\left[f(X_2) \mid x_1\right]) &= \lim_{k\to\infty} g_{x_1}\left(\overline{f(X_2)}_{n_k}\right) \\ &= g_{x_1}\left(\overline{f(X_2)}_{n_0}\right) + \sum_{k=1}^{\infty} g_{x_1}\left(\overline{f(X_2)}_{n_k}\right) - g_{x_1}\left(\overline{f(X_2)}_{n_{k-1}}\right). \end{aligned} \tag{5}$$

If one can construct estimator Δ_n with expectation

$$\mathbb{E}\left[g_{x_1}\left(\overline{f(X_2)}_{n_k}\right) - g_{x_1}\left(\overline{f(X_2)}_{n_{k-1}}\right)\right] \tag{6}$$

for $k \geq 1$ and $\mathbb{E}[\Delta_0] = g_{x_1}\left(\overline{f(X_2)}_{n_0}\right)$, a randomized estimator of $g_{x_1}(\mathbb{E}[f(X_2) \mid x_1])$ can be constructed as Δ_N/p_N, where N is a non-negative integer-valued random variable with probability mass function $\mathbb{P}[N = n] = p_n$. The following heuristic calculation explains why one would expect Δ_N/p_N to be unbiased:

$$\mathbb{E}\left[\frac{\Delta_N}{p_N}\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{\Delta_N}{p_N}\middle| N\right]\right] = \mathbb{E}\left[\sum_{k=0}^{\infty}\frac{\Delta_k}{p_k}\cdot p_k\right] = \sum_{k=0}^{\infty}\mathbb{E}[\Delta_k]$$
$$= \lim_{k\to\infty}\mathbb{E}\left[g_{x_1}\left(\overline{f(X_2)}_{n_k}\right)\right] = g_{x_1}(\mathbb{E}[f(X_2) \mid x_1])$$

where the third equality interchanges the order between expectation and (infinite) summation, the last equality interchanges the order between limit and expectation.

The above is the core idea of the unbiased MLMC estimator in [8,9,35]. However, it remains to justify several theoretical issues, such as the validity of the above interchange and the estimator's variance. An extra subtlety is the tradeoff between the sampling complexity and the variance. The expected sampling complexity for generating one estimator Δ_N/p_N is of the order of $\sum_{k=0}^{\infty} p_k n_k$. Clearly, it is desirable that the estimator has both finite variance and finite expected sampling complexity.

Rhee and Glynn [35] show the estimator Δ_N/p_N is unbiased and of finite variance if $\sum_{k=0}^{\infty}\mathbb{E}\left[\Delta_k^2\right]/p_k < \infty$ in a more general context. If one is interested in estimating quantities of the form $g(\mathbb{E}[X])$, Blanchet and Glynn [8] show one can choose $n_k = 2^k$ and $N \sim \text{Geo}(1 - 2^{-3/2})$ provided that X has bounded 6th order moment and g is locally twice differentiable and grows moderately. However, the assumption of [8] is not satisfied even in this simple two-stage case, as the function $g_{x_1}(a) = \max\{f(x_1), a\}$ here is non-differentiable at $f(x_1)$. The absence of smoothness assumptions on the function g_{x_1} causes technical challenges and calls for better theoretical guarantees in analyzing the unbiased MLMC estimator.

Now we are ready to describe the two-stage MUSE and discuss its theoretical properties. Algorithm 1 is referred to as the two-stage MUSE in contrast to the general/multi-stage MUSE described later. Roughly speaking, one first samples $x_1 \sim \pi_1$, then constructs the standard unbiased MLMC estimator Δ_N/p_N for $g_{x_1}(\mathbb{E}[f(X_2) \mid x_1])$ using a geometric random variable N and 2^N *i.i.d.* samples of X_2 with distribution $\pi_2(x_2 \mid \{x_1\})$. The estimator Δ_n described in Step 4 is crucial for theoretical analysis. It is often referred to as the 'antithetical difference' estimator in the literature [8,9]. The intuition is that the antithetic construction reduces the variance. As elaborated later in the theoretical analysis, the estimator Δ_N equals 0 if both $\frac{S^E_{2^{N-1}}}{2^{N-1}}$ and $\frac{S^O_{2^{N-1}}}{2^{N-1}}$ are on the same side of $f(X_1(1))$. This observation turns out to be the key for controlling the expected computational complexity and variance simultaneously. We want to emphasize that the main contribution of the two-stage MUSE is more theoretical rather than the methodological. Algorithmically, the two-stage MUSE is very similar to the unbiased MLMC estimator. Theoretically, two-stage MUSE is the first unbiased estimator with theoretical guarantees for dealing with non-smooth functions.

Our main theoretical results on the two-stage MUSE are described in Theorem 1. Notice that the computational cost for Algorithm 1 is a random variable depending on N. If we define the computation time for sampling one random variable and performing one arithmetic operation as 'one unit', then the expected computational complexity is of the order of $\mathbb{E}\left[2^N\right] = \sum_{n=1}^{\infty} 2^n p_r(n)$.

Theorem 1. *Consider a two-stage process* (X_1, X_2). *Suppose Assumptions 1, 2 (with* $\delta < 1/4$*) and 3 hold, and suppose Assumption 4 is satisfied with* $T = 2$, *i.e.,*

$$\mathbb{P}\left(|\mathbb{E}[f(X_2) \mid X_1] - f(X_1)| \leq \varepsilon\right) < C\varepsilon \tag{7}$$

Algorithm 1: Two-stage Multilevel Unbiased Stopping Estimator (Two-stage MUSE)

Input: A simulator $\mathcal{S}$ of a two-stage process (X_1, X_2), parameter $r \in (1/2, 1)$.
Output: An unbiased estimator of $\mathbb{E}[\max\{f(X_1), \mathbb{E}[f(X_2) \mid X_1]\}]$.
Step 1. Sample N from geometric distribution Geo(r).
Step 2. Sample $X_1(1)$. Conditioning on $X_1(1)$, sample 2^N *i.i.d.* $X_2(1), \ldots, X_2(2^N)$.
Step 3. Calculate the following three quantities:

$$S_{2^N} = f(X_2(1)) + \cdots + f\left(X_2\left(2^N\right)\right),$$
$$S^O_{2^{N-1}} = f(X_2(1)) + f(X_2(3)) + \cdots + f\left(X_2\left(2^N - 1\right)\right), \quad \text{(sum over odd indices)}$$
$$S^E_{2^{N-1}} = f(X_2(2)) + f(X_2(4)) + \cdots + f\left(X_2(2^N)\right). \quad \text{(sum over even indices)}$$

Step 4. Calculate (note that $\Delta_0 := \max\{f(X_1(1)), f(X_2(1))\}$)

$$\Delta_N = \max\left\{f(X_1(1)), \frac{S_{2^N}}{2^N}\right\} - \frac{1}{2}\left[\max\left\{f(X_1(1)), \frac{S^O_{2^{N-1}}}{2^{N-1}}\right\} + \max\left\{f(X_1(1)), \frac{S^E_{2^{N-1}}}{2^{N-1}}\right\}\right].$$

Return: $Y := \Delta_N / p_r(N)$.

for all $\varepsilon > 0$. *Let* $r = 1 - 2^{-\frac{2+9\delta/(80+40\delta)}{2+\delta/10}} \in (1/2, 1)$ *in Algorithm 1. Then, the resulting estimator* Y *in Algorithm 1 has the following properties:*

(1) $\mathbb{E}[Y] = \mathbb{E}[\max\{f(X_1), \mathbb{E}[f(X_2) \mid X_1]\}]$.
(2) *The expected computational complexity of* Y *is finite.*
(3) $\mathbb{E}\left[|Y|^{2+\frac{\delta}{10}}\right] \le \widetilde{C} \cdot L^{2+\delta}\left[1 + \mathbb{E}\left[\|X_2\|^{2+\delta}\right]\right]$, *where* $\widetilde{C}$ *is a constant independent of* (X_1, X_2).

The proof of Theorem 1 is deferred to Appendix A. As shown in Theorem 1, the two-stage MUSE is unbiased, has both finite $(2 + \frac{\delta}{10})$th moment (thus finite variance) and finite expected computational complexity. We also want to highlight a seemingly small theoretical improvement that turns out to be crucial in designing the multi-stage MUSE. In the existing literature, such as [8,9], the estimator is guaranteed to have a finite second moment given the original random variable has a higher (say 6th) moment. In our case, we prove the estimator has $(2 + \frac{\delta}{10})$th moment given the original random variable has $(2 + \delta)$th moment, which makes the whole algorithm iterable in the multi-stage case.

2.2. MUSE for general optimal stopping problems

In this section, we propose the multi-stage MUSE algorithm (Algorithm 2) which aims to provide an unbiased estimator for the general optimal stopping problem (1). The multi-stage MUSE, as described in Algorithm 2, can be viewed as a recursive extension of the two-stage MUSE. To get an unbiased estimator of U_T, one feeds $(0; \varnothing; \mathcal{S}, r_1, \ldots, r_{T-1})$ into Algorithm 2. After sampling x_1 from the unconditioned distribution and $N_1 \sim \text{Geo}(r_1)$, it suffices to construct 2^{N_1} unbiased estimators of $U_{T-1}(x_1)$ to build the MLMC estimator. Meanwhile, an unbiased estimator of $U_{T-1}(x_1)$ can be viewed as another optimal stopping problem with horizon $T - 1$ and underlying process $\pi_{2:T}$, and therefore we call the same algorithm recursively after adding x_1 into the trajectory history.

Algorithm 2: Multi-stage Multilevel Unbiased Stopping Estimator (Multi-stage MUSE)

Input: Time index k. Trajectory history $H = \{x_1, \ldots, x_k\}$ or $\varnothing$. A simulator $\mathcal{S}$ of the conditional distribution π_{T-k}, parameters $r_{k+1}, \ldots, r_{T-1} \in (1/2, 1)$.

Output: An unbiased estimator of U_{T-k} in (3).

if $k = T-1$ **then**

Sample one x_T from the conditional distribution of π_T given H.

Return $Y := f(x_T)$.

end

else

Sample x_{k+1} from the condition distribution π_{k+1} given H.

Add x_{k+1} to the trajectory history H.

Sample $N_{k+1} \sim \text{Geo}(r_{k+1})$.

Call Algorithm 2 for $2^{N_{k+1}}$ times with inputs $(H; \mathcal{S}, r_{k+2} \cdots, r_{T-1})$, label the outputs by $Y_{k+1}(1), \ldots, Y_{k+1}(2^{N_{k+1}})$.

Calculate the following three quantities:

$$S_{2^{N_{k+1}}} = Y_{k+1}(1) + \cdots + Y_{k+1}(2^{N_{k+1}}),$$
$$S^O_{2^{N_{k+1}-1}} = Y_{k+1}(1) + Y_{k+1}(3) \cdots + Y_{k+1}(2^{N_{k+1}} - 1),$$
$$S^E_{2^{N_{k+1}-1}} = Y_{k+1}(2) + Y_{k+1}(4) \cdots + Y_{k+1}(2^{N_{k+1}}).$$

Calculate (note that $\Delta_0 := \max\{f(x_{k+1}), Y_{k+2}(1)\}$)

$$\Delta_{N_{k+1}} = \max\left\{f(x_{k+1}), \frac{S_{2^{N_{k+1}}}}{2^{N_{k+1}}}\right\} - \frac{1}{2}\left[\max\left\{f(x_{k+1}), \frac{S^O_{2^{N_{k+1}-1}}}{2^{N_{k+1}-1}}\right\} + \max\left\{f(x_{k+1}), \frac{S^E_{2^{N_{k+1}-1}}}{2^{N_{k+1}-1}}\right\}\right].$$

Return: $Y := \Delta_{N_{k+1}} / p_{r_{k+1}}(N_{k+1})$.

end

The next theorem studies the theoretical properties of the multi-stage MUSE. The computational complexity of Algorithm 2 comes from the sampling complexity, which is of the order of $\mathbb{E}\left[\prod_{k=1}^{T-1} 2^{N_k}\right]$.

Theorem 2. *With Assumptions 1, 2, 3, and 4, consider the input*

$$(0; \varnothing; \mathcal{S}, r_1, \ldots, r_{T-1})$$

in Algorithm 2, where

$$r_i = 1 - 2^{-\frac{2+9\left(\delta \cdot 10^{i+1-T}\right)/\left(80+40\left(\delta \cdot 10^{i+1-T}\right)\right)}{2+\delta \cdot 10^{i-T}}} \in (1/2, 1)$$

for $1 \le i \le T-1$. *Then, the resulting estimator* Y *in Algorithm 2 has the following properties:*

(1) $\mathbb{E}[Y] = U_T$.

(2) *The expected computational complexity is* $O\left(10^{T^2}\right)$.

(3) $\mathsf{Var}(Y) = O\left(10^{T^2}\right)$.

To illustrate the iterative structure of multi-stage MUSE, we sketch the proof of Theorem 2 in below. The detailed proof is deferred to Appendix A. By the standard dynamical programming for optimal stopping, we have

$$\begin{cases} U_1(X_{1:T-1}) = \mathbb{E}\left[f(X_T) \mid X_{1:T-1}\right], \\ U_{T-k}(X_{1:k}) = \mathbb{E}\left[\max\left\{f\left(X_{k+1}\right), U_{T-(k+1)}(X_{1:k+1})\right\} \mid X_{1:k}\right], \quad 0 \le k \le T-2. \end{cases}$$

Here, for a generic d-tuple $(v_1, \ldots, v_d)$, let $v_{i:j} := (v_i, \ldots, v_j)$ for $1 \le i \le j \le d$ for notational convenience. By applying the techniques in the proof of Theorem 1, one can show that for each stage, the output Y_{T-k} always has a moment of order greater than 2, and is unbiased for U_{T-k}. The proof of the moment bounds is iterable because of the careful technical analysis in Theorem 1. Moreover, by a proper choice of the parameters $r_1, \ldots, r_{T-1}$, the expected sampling complexity for each stage is bounded. As a result, the total expected sampling complexity is also bounded.

Corollary 1. *Let Assumptions 2 and 4 hold. For any $\varepsilon > 0$, and a fixed time horizon T, we can construct an estimator Y that satisfies the following properties:*

- *The expected computational complexity for constructing Y is $O(1/\varepsilon^2)$.*
- *The expected mean squared error between Y and the true utility is bounded by ϵ^2, i.e., $\mathbb{E}\left[(Y - U_T)^2\right] \le \varepsilon^2$.*

Proof of Corollary 1. We fix a positive integer n. Calling Algorithm 2 n times yields n *i.i.d.* unbiased estimators $Y_1, \ldots, Y_n$ of U_T. Then,

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} Y_i - U_T\right)^2\right] = \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}(Y_i - \mathbb{E}[Y_i])\right)^2\right] = \frac{1}{n}\mathsf{Var}(Y_1).$$

Taking $n = \mathsf{Var}(Y_1)/\varepsilon^2$ (note that $\mathsf{Var}(Y_1) < \infty$ by Theorem 2) and define $Y := \frac{1}{n}\sum_{i=1}^{n} Y_i$. It follows from the above calculation that $\mathbb{E}\left[(Y - U_T)^2\right] \le \varepsilon^2$. Moreover, since sampling each Y_i has expected computational complexity $O(1)$, the expected computational complexity for constructing Y is $O(1/\varepsilon^2)$, as desired. □

Finally we comment on some practical issues when implementing the MUSE for multi-stage optimal stopping problems. One drawback of our algorithm is that the computational complexity (the constant hidden in $\mathcal{O}(1/\epsilon^2)$ in Corollary 1) grows exponentially with time horizon T. Therefore, our algorithm is prohibitively slow when T becomes large. We believe this is expected due to the comprehensive multi-stage structure of the optimal stopping problem (1). In fact, the same phenomenon happens in the Monte Carlo-based methods, including the popular algorithms of Broadie and Glasserman [10], Longstaff and Schwartz [31] and Tsitsiklis and Van Roy [38]. It is known in Glasserman and Yu [22] that the number of sample paths required for the regression coefficients to converge grows exponentially in the degree of basis functions under the worst-case scenario. Zanger [42] proved the expected L^2 error has an $O((\log^{1/2} N)N^{-1/2})$ convergence rate (N is the number of sample paths) given the approximation architecture has a finite Vapnik–Chervonenkis (VC) dimension. Their error bound also scales exponentially with respect to the time horizon, see Theorem 3.3 of [42]. Meanwhile, we emphasize that besides Assumption 4, there is currently no specific distributional assumption on the underlying process. Therefore, there is a potential for designing computationally efficient estimators given additional distribution assumptions. Moreover, if the

unbiased requirement can be relaxed, then it is possible to design fast algorithms while retaining the $\mathcal{O}(1/\epsilon^2)$ complexity. One can potentially either choose a fixed level based on the standard MLMC approaches [18], or use a truncated geometric random variable to replace the geometric random variable N in Algorithm 2 as described in the recent work [1]. These ideas open up exciting possibilities for new algorithms, but are already beyond the target of our paper.

2.3. Confidence interval and optimal stopping time

A confidence interval (CI) is crucial if one is not merely interested in getting a point estimate, but also expects to assess the quality of such estimation. Fortunately, since many $i.i.d.$ estimators of U_T can be constructed by repeatedly calling Algorithm 2, the $1-\alpha$ confidence interval (CI) of the utility can be constructed as follows: Let $Y_1, \ldots, Y_n$ be n unbiased estimators of U_T generated by the MUSE. Let $\overline{Y}$ be their empirical mean and s the standard deviation. Then, two types of CIs can be built via

- (CLT) $\left[\overline{Y} - z_{\alpha/2} \cdot s/\sqrt{n}, \overline{Y} + z_{\alpha/2} \cdot s/\sqrt{n}\right]$, where $z_{\alpha/2}$ is the $(1-\alpha/2)$th quantile of $\mathcal{N}(0,1)$.
- (Bootstrap [15]) $[\overline{Y}^{\star}_{\alpha/2}, \overline{Y}^{\star}_{1-\alpha/2}]$, where $\overline{Y}^{\star}_{\alpha/2}$, $\overline{Y}^{\star}_{1-\alpha/2}$, are the $\alpha/2$th and $(1-\alpha/2)$th empirical quantile of the bootstrap averages.

In principle, both methods are valid as the number of simulated estimators goes to infinity. The first CI is based on the Central Limit Theorem, and the convergence rate depends on the higher-order cumulants. The second CI uses the empirical distribution to approximate the actual underlying distribution, which is non-parametric and is (monotone) transformation-respecting ([15], Chapter 12). It is known that the percentile bootstrap may not work well when the data has a significant skewed distribution. In these cases one may consider alternative methods such as the $\mathrm{BC_a}$ (bias corrected accelerated) bootstrap [12].

Besides estimating U_T, we are also interested in finding the optimal stopping time τ^* such that $\mathbb{E}[f(X_{\tau^*})] = U_T$. By standard dynamical programming,

$$\tau^* = \inf\{k \geq 1 : U_{T-k}(X_1, \ldots, X_k) \leq f(X_k)\}.$$

Though U_{T-k} is not analytically available, the MUSE provides us with powerful tools for estimating U_{T-k} at each round. The algorithm for the optimal stopping time is as follows:

Algorithm 3: Optimal Stopping Time via MUSE

Input: Simulator of the process $(X_1, \ldots, X_T)$, tolerance level ε.
Output: A stopping time $\widehat{\tau}$.
Sample $X_1 = x_1$.
for $k \leftarrow 1$ **to** $T-1$ **do**
  Call Algorithm 2 with history $(x_1, \ldots, x_k)$ n times to get $i.i.d.$ unbiased estimators $Y_1, \ldots, Y_n$ of U_{T-k}.
  if $f(x_k) > \overline{Y} - \varepsilon$ **return** k. **else** Sample $X_{k+1} = x_{k+1}$.
  if $k+1 = T$ **return** T.
end

There are multiple ways of choosing ε, which clearly depend on the decision maker's risk sensitivity. One promising option would be to choose ε adaptively, according to the CIs derived by the MUSE.

3. Numerical experiments

3.1. Optimal stopping of independent random variables

The optimal stopping problem for independent random variables has been extensively studied in the literature. In this example, we consider the case where $X_1, \ldots, X_T$ are $i.i.d.$ $\mathcal{N}(0, 1)$ random variables with reward $f(x) = x$. Standard calculation yields $U_1 = 0$, $U_2 = \mathbb{E}[|X_1|]/2$ and $U_k = \mathbb{E}[\max\{X_1, U_{k-1}\}]$ so that the utility can be solved numerically. With each fixed time horizon, three estimators — MUSE and two vanilla Monte Carlo estimators MC1 and MC2 are implemented. MC1 is a naive Monte Carlo estimator. For each T, it samples 10^7 paths and estimates U_T by the average of the maximum in each path, which is clearly biased. MC2 is a refinement of MC1 but still biased. It samples tree-like paths as described in Section 1, Fig. 1. In our case, the simulated data forms a forest that consists of 1000 complete 5-ary trees of depth T. MC2 estimates the utility using the dynamical programming formula $U_T = \mathbb{E}[\max\{X_1, U_{T-1}(X_1)\}]$ in a backward recursive way. Given the history $X_1, \ldots, X_{n-1}$, the utility of $U_1(X_1, \ldots, X_{n-1})$ can be easily estimated by averaging the samples in the last layer. Similarly, we can use the formula $U_2(X_1, \ldots, X_{n-2}) = \mathbb{E}[\max\{X_{n-1}, U_1(X_1, \ldots, X_{n-1})\}]$ to estimate U_2 after replacing the quantity $U_1(X_1, \ldots, X_{n-1})$ by its estimator described above. Then we estimate $U_3, U_4, \ldots$ and finally U_T. Formally, the final estimator $\hat{U}_T = (\sum_{i=1}^{1000} \hat{U}_T^{(i)})/1000$ is the average of the 1000 estimators from each tree. For each i, the estimator $\hat{U}_T^{(i)}$ of tree i is of the form $\hat{U}_T^{(i)} := (\sum_{j=1}^{5} \max\{X_{1,1}^{(i)}, \hat{U}_{T-1}^{(i)}\})/5$, where the number 5 comes from the 5-ary tree design, $X_{1,1}^{(i)}$ is the root node of the ith tree, and $\hat{U}_{T-1}^{(i)}$ is the estimator for $U_{T-1}(X_{1,1}^{(i)})$ using the dynamical programming procedure mentioned above.

The only hyperparameter for the MUSE is the success probability r for the geometric distribution. Larger r leads to shorter computational time but larger variance, and vice versa. We implement a simple experiment to determine r. For each r in $\{0.51, 0.52, \ldots, 0.7\}$, we run 10^6 MUSEs for horizon $T = 3$ and examine their empirical performances. Our results are summarized in Fig. 2. The cost of time decays significantly when r increases. Furthermore, we also calculate the self-normalized variance [8] as a measure of efficiency. The self-normalized variance is defined as the product between the expected time and the variance for every single estimator. It is clear from the right subplot of Fig. 2 that the self-normalized variance initially decays and then increases as r increases, with a minimum at around 0.6, therefore we choose $r = 0.6$ in the numerical experiments henceforth.

After setting up the hyperparameter, we implement the three methods for $T \in \{2, \ldots, 7\}$. Our results are presented in Fig. 3. Both MC1 (red curve) and MC2 (green curve) systematically overestimate the true utility (black dotted line), as expected. The accuracy of MC1 is poor while MC2 has much better accuracy, sometimes comparable with the MUSE. The MUSE (blue curve) uses parameters $r_i = 0.6$ for each stage,[3] and averages of 10^6 estimators for each T. It typically has the most accurate result among all three methods. To better understand the empirical convergence behavior of the MUSE, we also show the traceplot for the running average of the MUSE for each horizon in the right subplot of Fig. 3. It is clear from the traceplot that the CIs typically covers the ground truth, though the convergence becomes much slower when T is increasing.

[3] To ease the computation burden, the parameters chosen here do not strictly follow Theorem 2.

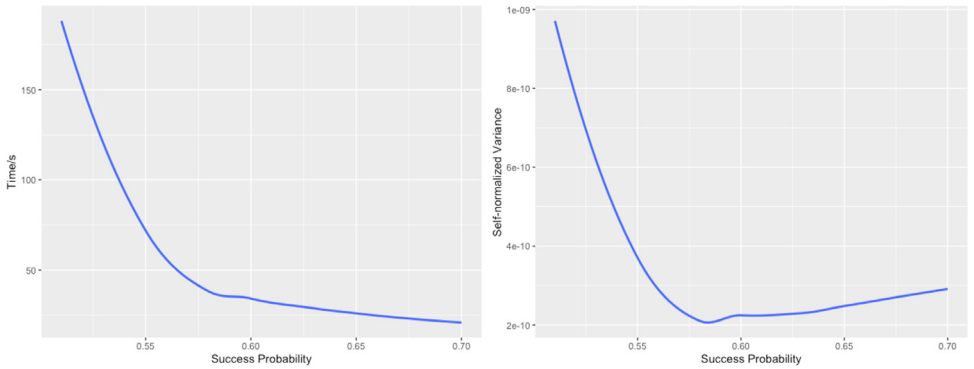

Fig. 2. Left: The cost of time for generating 10^6 MUSEs with different success probabilities. Right: The self-normalized variance of the MUSEs with different success probabilities.

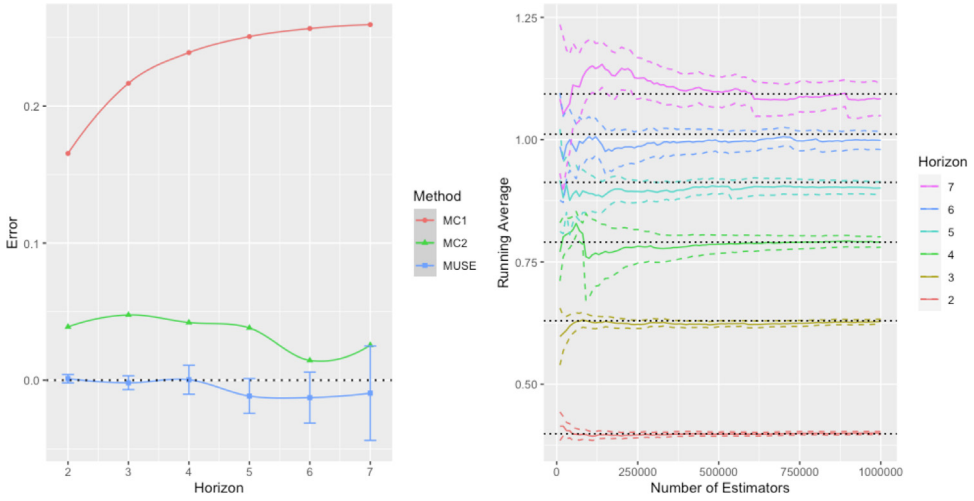

Fig. 3. Left: Comparison between the errors of the MUSE (blue), MC1 (red), and MC2 (green) for estimating the utility for $i.i.d.$ standard Gaussian random variables. Blue error bars stand for the 95% confidence intervals of the MUSE. Black dotted line stands for the ground truth (error = 0). Right: The traceplot of the running averages of the MUSE with different horizons. Black dotted line stands for the ground truth. Colored dashed lines stand for the running 95% CIs of the MUSE. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.2. Pricing the Bermudan options with high-dimensional inputs on a computer cluster

In this section we consider a more challenging setup, where the underlying process $\mathbf{X}_t := \left(X_t^{(1)}, \ldots, X_t^{(d)}\right)$ takes values in a high-dimensional space $\mathbb{R}^d$. The example we are considering here is a standard one — pricing the high-dimensional Bermudan-basket put options. The underlying process is a d-dimensional independent geometric Brownian motion with drift $\gamma - \delta$

Table 1
Comparison between different methods when $d = 5$. SGM and BKS stand for results reported by [30] and [6] respectively. LSM stands for Longstaff–Schwartz method, reported by [30].

Method	LSM (s.e.)	SGM direct (s.e.)	SGM LB (s.e.)	BKS (95% CI)	MUSE (s.e.)
$d = 5$	2.163(0.001)	2.141(0.008)	2.134(0.012)	[2.154, 2.164]	**2.161(0.004)**

Table 2
Results of the MUSE under different dimensions. The second column reports the means and standard errors of the MUSEs. The third column reports the average computation time over the 500 processors.

d	MUSE (s.e.)	Average time (s) per processor
5	2.161 (0.004)	15.922
10	0.985 (0.002)	14.787
20	0.355 (0.001)	16.004
100	0.0043 ($< 10^{-4}$)	18.271
1000	0(0)	32.191

and volatility σ where all parameters will be specified later. Bermudan-basket option has utility $f(t, \mathbf{X}_t) = e^{-\gamma t} \max\{0, K - \sum_{i=1}^{d} X_t^{(i)}/d\}$ at each t, where K is the strike price and $e^{-\gamma}$ is often referred to as the discounting factor. Bermudan option is only exercisable in a discrete set of times, which transforms the pricing problem to solving the optimal stopping problem: $U_T := \sup_{\tau \in \{T_1, \ldots, T_k\}} \mathbb{E}[f(\tau, \mathbf{X}_\tau)]$, where $0 \le T_1 \le \cdots \le T_k \le T$ are all the exercisable dates. It has been observed [26] that the computational cost for standard regression-based methods typically scales superlinearly with dimension d, which discourages their uses in the high-dimensional setups. Existing experiments on Bermudan options often assume $d \le 20$, though it can be as large as 5000 in practice [2].

In our experiment we adopt the standard parameters in [6,30] where $T = 3$ (years), $\sigma = 0.2, \gamma = 0.05, \delta = 0, K = X_0^{(i)} = 100$ for every i. Owners can exercise the option at the initial time or after 1, 2, 3 years. We first benchmark our result with the results reported in [6,30] when $d = 5$, next we present our results for $d \in \{10, 20, 100, 1000\}$. For each d, we use 10^7 MUSEs generated by a 500-core CPU-based computer cluster, where the parameters r_i are set to be 0.6 for each stage. The results when $d = 5$ is presented in Table 1, the MUSE matches the results from other methods while preserving unbiasedness and having a relatively small standard error.

Table 2 records the estimates and the standard errors of the MUSE when d is increasing. There are no existing benchmark results for large d thus we are not able to compare with the ground truth. But the law of large numbers shows the utility should converge to 0 as d goes to infinity, which matches our result here. We also record the average computing time for every processor in the last column of Table 2, the computation time scales sublinearly with the dimensionality d, which may be benefited from the use of vectorization in simulating the d-dimensional geometric Brownian motion. We also plot the histogram of the computing time among 500 cores when $d = 100$ in Fig. 4. It is clear from Fig. 4 that the computing times are relatively short (less than 15 seconds) for most clusters even in this high-dimensional regime. There are a small proportion of clusters that uses much longer time. This fact indicates the MUSE has a high variance in its computational complexity, which is in line with our theoretical

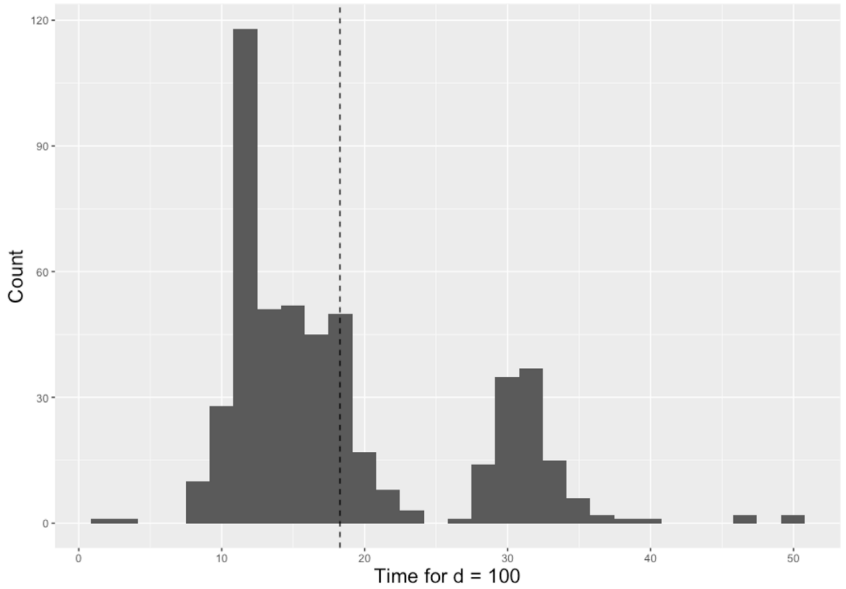

Fig. 4. Histogram of computational times among 500 processors when $d = 100$. The black dotted line is the average computation time, which is 18.271 s.

intuitions. Finally, it seems the MUSE scales well with d, which may be another appealing feature besides parallel computing.

4. Conclusion and future work

Optimal stopping problems play an important role in modern decision-making processes. However, existing simulation algorithms introduce unavoidable bias in estimating the utility. In this paper, an unbiased estimator, the MUSE, is proposed and analyzed. Our estimator is easy to implement and enjoys unbiasedness, finite variance, and finite computational complexity after choosing the parameters appropriately. A key ingredient of the general MUSE is the iterative use of the two-stage MUSE, which preserves unbiasedness at every stage by the multilevel approach.

In the theoretical part of this paper, we focus on bounding the variance and computational complexity of the MUSE. Though finite variance and finite complexity are guaranteed, these upper bounds may be too crude to shed light on practical applications. Moreover, theoretical guarantees on the applications described in Section 2.3, such as regret bounds for Algorithm 3, are worth investigating.

In the numerical studies, experiments in Section 3 suggest the MUSE is able to provide accurate estimation for the utilities, especially when T is small or moderate. The MUSE also seems to scale well with the dimensionality of the underlying process, as shown in Section 3.2. On the other hand, our estimator's variance and computational complexity grow significantly with the horizon length. Running Algorithm 2 can easily be prohibitive with large horizons. It remains a key challenge to design scalable algorithms while maintaining unbiasedness, or at

least controlling bias at a negligible level under the large-horizon regime. As a closing remark, unbiasedness is undoubtedly an appealing property in parallel computation, but it could come with higher computational cost or lower statistical accuracy. Therefore, studying the trade-offs between unbiasedness, computational budget constraints, and accuracy may be of paramount interest to both theorists and practitioners. We hope future studies will provide much-needed insight toward achieving practical unbiasedness with sustainable cost and high accuracy.

While our paper focus on the optimal stopping problem, we believe our technique can potentially be extended to more general setups. The optimal stopping is a subfamily of the stochastic control problems, where one picks the optimal time to stop. One natural extension is the case where the one needs to choose one out of K possible actions at each stage. Moreover, our paper considers the optimal problem with finite horizon. Another natural extension is to consider the infinite horizon stopping problem with discounted reward.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Material in this paper is based upon work supported by the Air Force Office of Scientific Research, United States of America under award number FA9550-20-1-0397. Additional support is gratefully acknowledged from National Science Foundation, United States of America grants 1915967, 1820942, and 1838576. Guanyang Wang would like to sincerely thank Changpeng Lu for her patient guidance on configuring the computer cluster, and David Sichen Wu for helpful suggestions on improving this paper. We would like to thank two anonymous referees for their time, effort, and constructive suggestions.

Appendix A. Auxiliary results

Lemma 1 (*[32] Marcinkiewicz–Zygmund Inequality*)**.** *If $X_1, \ldots, X_n$ are independent random variables with $\mathbb{E}[X_i] = 0$ and $\mathbb{E}\left[|X_i|^p\right] < \infty$ for some $p > 2$. Then,*

$$\mathbb{E}\left[\left|\sum_{i=1}^{n} X_i\right|^p\right] \leq C'_p \mathbb{E}\left[\left(\sum_{i=1}^{n} |X_i|^2\right)^{p/2}\right],$$

where C'_p is a constant that only depends on p. If we further assume that $X_1, \ldots, X_n$ are i.i.d. Then,

$$\mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right|^p\right] \leq C'_p \mathbb{E}\left[\frac{1}{n^{p/2}}\left[\frac{1}{n}\sum_{i=1}^{n} |X_i|^2\right]^{p/2}\right] \leq C'_p \cdot \frac{\mathbb{E}|X_1|^p}{n^{p/2}}.$$

Corollary 2. *Let (Z_1, Z_2) be a $2-$stage stochastic process, and there exists $p > 2$, such that $\sup_{i=1,2} \mathbb{E}\left[|Z_i|^p\right] < \infty$. Conditioning on Z_1, sample i.i.d. $Z_2(1), \ldots, Z_2(n)$. Then,*

$$\mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n} Z_2(i) - \mathbb{E}[Z_2 \mid Z_1]\right|^p\right] \leq C_p \cdot \frac{\mathbb{E}\left[|Z_2|^p\right]}{n^{p/2}},$$

where C_p is a universal constant only depends on p.

Proof of Corollary 2. Since $\mathbb{E}\left[|Z_2|^p\right] < \infty$, we have that $\mathbb{E}\left[|Z_2|^p \mid Z_1 = z_1\right]$ exists almost surely. Let $\pi_{1:2}$ be the joint measure of (Z_1, Z_2), applying Lemma 1 to the conditional distribution $\pi_2(\cdot \mid z_1)$ yields

$$\begin{aligned}
&\mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n} Z_2(i) - \mathbb{E}[Z_2 \mid Z_1]\right|^p\right] \\
&= \int_{\Omega} \mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n} Z_2(i) - \mathbb{E}[Z_2 \mid Z_1]\right|^p \;\middle|\; Z_1 = z_1\right] \pi_1(dz_1) \\
&\le \int_{\Omega} \frac{C'_p}{n^{p/2}} \mathbb{E}\left[|Z_2 - \mathbb{E}[Z_2 \mid Z_1]|^p \mid Z_1 = z_1\right] \pi_1(dz_1) \\
&\le \int_{\Omega} \frac{C'_p 2^{p-1}}{n^{p/2}} \left[\mathbb{E}\left[|Z_2|^p | Z_1 = z_1\right] + |\mathbb{E}[Z_2 | Z_1 = z_1]|^p\right] \pi_1(dz_1) \\
&\le \int_{\Omega} \frac{C'_p 2^{p}}{n^{p/2}} \mathbb{E}\left[|Z_2|^p \mid Z_1 = z_1\right] \pi_1(dz_1) = \frac{C'_p 2^p}{n^{p/2}} \mathbb{E}[|Z_2|^p]. \quad \square
\end{aligned}$$

Appendix B. Proofs of main theorems

We first present the proof of Theorem 1.

Proof of Theorem 1. We first show that $\mathbb{E}[Y] = \mathbb{E}\left[\max\left\{f(X_1), \mathbb{E}\left[f(X_2) \mid X_1\right]\right\}\right]$. Note that the X_1, X_2 are integrable,

$$\begin{aligned}
& \mathbb{E}[Y] \\
= \;& \mathbb{E}\left[\mathbb{E}\left[Y \mid N\right]\right] && \text{(B.1)} \\
= \;& \sum_{n=0}^{\infty} \mathbb{E}[\Delta_n] && \text{(B.2)} \\
= \;& \sum_{n=1}^{\infty} \left(\mathbb{E}\left[\max\left\{f\left(X_1(1)\right), \frac{S_{2^n}}{2^n}\right\}\right] - \mathbb{E}\left[\max\left\{f\left(X_1(1)\right), \frac{S_{2^{n-1}}}{2^{n-1}}\right\}\right] \right) + \\
& \quad \mathbb{E}\left[\max\left\{f\left(X_1(1)\right), f\left(X_2(1)\right)\right\}\right] \\
= \;& \lim_{n\to\infty} \mathbb{E}\left[\max\left\{f\left(X_1(1)\right), \frac{S_{2^n}}{2^n}\right\}\right] - \mathbb{E}\left[f\left(X_1(1)\right), S_1\right] \\
& + \mathbb{E}\left[\max\left\{f\left(X_1(1)\right), f\left(X_2(1)\right)\right\}\right] \\
= \;& \mathbb{E}\left[\max\left\{f\left(X_1(1)\right), \lim_{n\to\infty} \frac{S_{2^n}}{2^n}\right\}\right] && \text{(B.3)} \\
= \;& \mathbb{E}\left[\max\left\{f\left(X_1(1)\right), \mathbb{E}[f(X_2) \mid X_1(1)]\right\}\right]. && \text{(B.4)}
\end{aligned}$$

Here the law of large number is applied to guarantee the equality between (B.3) and (B.4), and the equality of (B.1), and (B.2) is established by the interchanging the order of summation and expectation, which is legitimate due to the fact that

$$\sum_{n=0}^{n} \mathbb{E}\left[|\Delta_n|\right] < \infty. \tag{B.5}$$

To verify the inequality (B.5), note that $\max\{x, a\}$ is a 1-Lipschitz function of x for any fixed a, we have

$$\begin{aligned}|\Delta_n| &\le \frac{1}{2}\left|\max\left\{f\left(X_1(1)\right), \frac{S_{2^n}}{2^n}\right\} - \max\left\{f\left(X_1(1)\right), \frac{S^O_{2^{n-1}}}{2^{n-1}}\right\}\right| \\ &\quad + \frac{1}{2}\left|\max\left\{f\left(X_1(1)\right), \frac{S_{2^n}}{2^n}\right\} - \max\left\{f\left(X_1(1)\right), \frac{S^E_{2^{n-1}}}{2^{n-1}}\right\}\right| \\ &\le \frac{1}{2}\left|S^O_{2^{n-1}}/2^{n-1} - S^E_{2^{n-1}}/2^{n-1}\right|.\end{aligned}$$

By Corollary 2, and note that $|f(x)| \le L(1 + \|x\|)$, we have

$$\begin{aligned}&\sum_{n=0}^{\infty}\mathbb{E}\left[|\Delta_n|\right] \\ &\le 2L^{2+\delta}\left[1 + \sup_{i=1,2}\left[\mathbb{E}\left[\|X_i\|^{2+\delta}\right]\right]^{\frac{1}{2+\delta}}\right] + \frac{1}{2}\sum_{n=1}^{\infty}\left[\mathbb{E}\left[\left[\left|\frac{S^O_{2^{n-1}}}{2^{n-1}} - \frac{S^E_{2^{n-1}}}{2^{n-1}}\right|^{2+\delta}\right]\right]\right]^{\frac{1}{2+\delta}} \\ &\le 2L^{2+\delta}\left[1 + \sup_{i=1,2}\left[\mathbb{E}\left[\|X_i\|^{2+\delta}\right]\right]^{\frac{1}{2+\delta}}\right] + \frac{1}{2}\sum_{n=1}^{\infty}\left[\frac{C_{2+\delta}2^{2+\delta}\mathbb{E}\left[|f(X_2)|^{2+\delta}\right]}{2^{(n-1)(2+\delta)/2}}\right]^{\frac{1}{2+\delta}} \\ &= 2L^{2+\delta}\left[1 + \sup_{i=1,2}\left[\mathbb{E}\left[\|X_i\|^{2+\delta}\right]\right]^{\frac{1}{2+\delta}}\right] + C_{2+\delta}^{\frac{1}{2+\delta}}L2^{\frac{1+\delta}{2+\delta}}\left[1 + \mathbb{E}\|X_2\|^{2+\delta}\right]^{\frac{1}{2+\delta}}\sum_{n=1}^{\infty}\frac{\sqrt{2}}{2^{\frac{n}{2}}} \\ &< \infty.\end{aligned}$$

Next, we show that Y satisfies the properties (2) and (3) in Theorem 1. Namely, finite expected sampling complexity and bounded $2+\delta/10$ moment. In order to bound the $2+\delta/10$ moment of Y, we introduce the following events:

$$\begin{aligned}E_1 &:= \{|\mathbb{E}[f(X_2) \mid X_1(1)] - f(X_1(1))| < \varepsilon\}, \\ E_2 &:= \left\{\left|S^O_{2^{n-1}}/2^{n-1} - \mathbb{E}[f(X_2) \mid X_1(1)]\right| \ge \varepsilon/2\right\}, \\ E_3 &:= \left\{\left|S^O_{2^{n-1}}/2^{n-1} - S^E_{2^{n-1}}/2^{n-1}\right| \ge \varepsilon/2\right\}.\end{aligned}$$

Observe that

$$\mathbb{E}\left[|\Delta_n|^{2+\delta/10}\right] = \mathbb{E}\left[|\Delta_n|^{2+\delta/10}\mathbb{1}(E_1^c \cap E_2^c \cap E_3^c)\right] + \mathbb{E}\left[|\Delta_n|^{2+\delta/10}\mathbb{1}(E_1 \cup E_2 \cup E_3)\right].$$

On the event $E_1^c \cap E_2^c \cap E_3^c$, we have

$$\begin{aligned}|\mathbb{E}[f(X_2) \mid X_1(1)] - f(X_1(1))| &\ge \varepsilon, \\ \left|S^O_{2^{n-1}}/2^{n-1} - \mathbb{E}[f(X_2) \mid X_1(1)]\right| &\le \varepsilon/2, \\ \left|S^O_{2^{n-1}}/2^{n-1} - S^E_{2^{n-1}}/2^{n-1}\right| &\le \varepsilon/2.\end{aligned}$$

Thus, both $S^O_{2^{n-1}}/2^{n-1}$ and $S^E_{2^{n-1}}/2^{n-1}$ are on the same side of $f(X_1(1))$. Since $S_{2^n} = S^O_{2^{n-1}} + S^E_{2^{n-1}}$, we get

$$\begin{aligned}\Delta_n &= \max\left\{f(X_1(1)), \frac{S_{2^n}}{2^n}\right\} - \frac{1}{2}\left[\max\left\{f(X_1(1)), \frac{S^O_{2^{n-1}}}{2^{n-1}}\right\} + \max\left\{f(X_1(1)), \frac{S^E_{2^{n-1}}}{2^{n-1}}\right\}\right] \\ &= 0.\end{aligned}$$

In other words,

$$\mathbb{E}\left[|\Delta_n|^{2+\delta/10}\mathbb{1}(E_1^c \cap E_2^c \cap E_3^c)\right] = 0. \tag{B.6}$$

Next, we bound the term $\mathbb{E}\left[|\Delta_n|^{2+\delta/10}\mathbb{1}(E_1 \cup E_2 \cup E_3)\right]$. By Hölder's inequality (with parameter $p = (2+\delta)/(2+\delta/10)$, and $q = (20+10\delta)/(9\delta)$. It is straight forward to verify that $1/p + 1/q = 1$),

$$\begin{aligned}
&\mathbb{E}\left[|\Delta_n|^{2+\delta/10}\mathbb{1}(E_1 \cup E_2 \cup E_3)\right] \\
&\le \left[\mathbb{E}\left[|\Delta_n|^{(2+\delta/10)\cdot\frac{2+\delta}{2+\delta/10}}\right]\right]^{\frac{2+\delta/10}{2+\delta}} \cdot \mathbb{E}[\mathbb{1}(E_1 \cup E_2 \cup E_3)]^{\frac{9\delta}{20+10\delta}} \\
&\le \left[\mathbb{E}\left[|\Delta_n|^{2+\delta}\right]\right]^{\frac{2+\delta/10}{2+\delta}} \cdot (\mathbb{P}(E_1) + \mathbb{P}(E_2) + \mathbb{P}(E_3))^{\frac{9\delta}{20+10\delta}}.
\end{aligned}$$

Take $\varepsilon = \frac{1}{2^{n/4}}$, by the assumption in (7),

$$\mathbb{P}(E_1) \le C\varepsilon.$$

Now, we bound the probabilities $\mathbb{P}(E_2)$ and $\mathbb{P}(E_3)$. By Corollary 2, there exists a universal constant $C_{2+\delta} > 0$, such that

$$\begin{aligned}
\mathbb{P}(E_2) &\le \frac{1}{(\varepsilon/2)^{2+\delta}}\mathbb{E}\left[\left|S^O_{2^{n-1}} - \mathbb{E}[f(X_2) \mid X_1(1)]\right|^{2+\delta}\right] \\
&\le \frac{1}{(\varepsilon/2)^{2+\delta}} \cdot \frac{C_{2+\delta}2^{1+\delta}L^{2+\delta}\left[1 + \mathbb{E}\left[\|X_2\|^{2+\delta}\right]\right]}{2^{(n-1)(2+\delta)/2}} \\
&= C_{2+\delta}2^{2+3\delta/2}L^{2+\delta} \cdot \left[1 + \mathbb{E}\left[\|X_2\|^{2+\delta}\right]\right]\varepsilon \cdot \frac{1}{2^{(\frac{1}{4}+\frac{\delta}{4})n-(2+\delta)}} \\
&\le C_{2+\delta}2^{2+3\delta/2}L^{2+\delta} \cdot \left[1 + \mathbb{E}\left[\|X_2\|^{2+\delta}\right]\right]\varepsilon.
\end{aligned}$$

Similarly,

$$\mathbb{P}(E_3) \le C_{2+\delta}2^{4+5\delta/2}L^{2+\delta}\mathbb{E}\left[1 + \mathbb{E}\left[\|X_2\|^{2+\delta}\right]\right]\varepsilon.$$

Moreover, recall that

$$|\Delta_n| \le \frac{1}{2}\left|S^O_{2^{n-1}}/2^{n-1} - S^E_{2^{n-1}}/2^{n-1}\right|,$$

by Corollary 2 again, we have

$$\begin{aligned}
\left[\mathbb{E}\left[|\Delta_n|^{2+\delta}\right]\right]^{\frac{2+\delta/10}{2+\delta}} &\le \left(C_{2+\delta}L^{2+\delta}\left[1 + \mathbb{E}\left[\|X_2\|^{2+\delta}\right]\right]\right)^{\frac{2+\delta/10}{2+\delta}} \cdot \frac{1}{(2^n)^{\frac{2+\delta/10}{2}}} \\
&\le \left(C_{2+\delta}L^{2+\delta}\left[1 + \mathbb{E}\left[\|X_2\|^{2+\delta}\right]\right]\right)^{\frac{2+\delta/10}{2+\delta}} \cdot \frac{1}{2^n}.
\end{aligned}$$

Thus, there exists universal constant $C' > 0$, such that

$$\begin{aligned}
&\mathbb{E}\left[|\Delta_n|^{2+\delta/10}\mathbb{1}(E_1 \cup E_2 \cup E_3)\right] \\
&\le C'\left(L^{2+\delta}\left[1 + \mathbb{E}\left[\|X_2\|^{2+\delta}\right]\right]\right)^{\frac{2+\delta/10}{2+\delta} + \frac{9\delta}{20+10\delta}} \cdot \frac{1}{2^{\left(1+\frac{9\delta}{80+40\delta}\right)n}} \\
&\le C'L^{2+\delta}\left[1 + \mathbb{E}\left[\|X_2\|^{2+\delta}\right]\right]\frac{1}{2^{\left(1+\frac{9\delta}{80+40\delta}\right)n}}.
\end{aligned} \tag{B.7}$$

Now, combining (B.6) and (B.7),

$$\begin{aligned}\mathbb{E}\left[|Y|^{2+\delta/10}\right] &\le 2^{1+\delta/10}\sum_{n=0}^{\infty}\frac{\mathbb{E}\left[|\Delta_n|^{2+\delta/10}\right]}{p_n^{1+\delta/10}}\\ &\le \frac{C'}{r}L^{2+\delta}\left[1+\mathbb{E}\left[\|X_2\|^{2+\delta}\right]\right]\sum_{n=0}^{\infty}\frac{1}{2^{\left(1+\frac{9\delta}{80+40\delta}\right)n}2^{-\frac{2+9\delta/(80+40\delta)}{2+\delta/10}n\cdot(1+\delta/10)}}\\ &= \frac{C'}{r}L^{2+\delta}\left[1+\mathbb{E}\left[\|X_2\|^{2+\delta}\right]\right]\sum_{n=0}^{\infty}\frac{1}{\left(2^{\frac{9\delta/(80+40\delta)-\delta/10}{2+\delta/10}}\right)^n}\\ &\le \widetilde{C}L^{2+\delta}\left[1+\mathbb{E}\left[\|X_2\|^{2+\delta}\right]\right]<\infty,\end{aligned}$$

where $\widetilde{C}$ is a universal constant independent of the process (X_1, X_2), and we have also used the fact that $\frac{9\delta}{80+40\delta}-\frac{\delta}{10}>0$ when $\delta<1/4$. Finally, the sampling complexity of Y is

$$\mathbb{E}\left[2^N\right]=\sum_{n=0}^{\infty}2^n p_n \lesssim \sum_{n=0}^{\infty}\frac{1}{\left(2^{\frac{9\delta/(80+40\delta)-\delta/10}{2+\delta/10}}\right)^n}<\infty.$$

Thus, the expected computational cost of Y is also finite. In sum, our estimator Y satisfies all the desired properties in Theorem 1. □

Next, we provide the full proof of Theorem 2.

Proof of Theorem 2. By the standard dynamical programming for optimal stopping, we have

$$\begin{cases}U_1(X_{1:T-1})=\mathbb{E}\left[f(X_T)\mid X_{1:T-1}\right],\\ U_{T-k}(X_{1:k})=\mathbb{E}\left[\max\left\{f\left(X_{k+1}\right),U_{T-(k+1)}(X_{1:k+1})\right\}\mid X_{1:k}\right],\quad 0\le k\le T-2.\end{cases}$$

Let $Y_{T-k}(x_{1:k})$ denote the output (which is a random variable) of Algorithm 2 given the input history $x_{1:k}$ which is sampled from $X_{1:k}$. For simplicity, let $\delta_k := \delta\cdot 10^{k+1-T}$ for $0\le k\le T-1$. We will prove by a backward induction to show that:

(a) $\mathbb{E}_{\pi_{k+1:T}}[Y_{T-k}(x_{1:k})]=U_{T-k}(x_{1:k}),\quad 0\le k\le T-1.$
(b) The expected sampling complexity $=\prod_{i=k+1}^{T-1}C_i<\infty$, $0\le k\le T-2$. As a result, the expected computational complexity is also finite.
(c) $\mathbb{E}_{\pi_{1:T}}\left[|Y_{T-k}(x_{1:k})|^{2+\delta_k}\right]<\left(\prod_{i=k+1}^{T-1}\widetilde{C}_i\right)\cdot L^{2+\delta}\left[1+\mathbb{E}\left[\|X_T\|^{2+\delta}\right]\right],\quad$ for all $0\le k\le T-2.$

Here $C_i, \widetilde{C}_i$ $(1\le i\le T-1)$ are some positive constants independent of the underlying process.

When $k=T-1$, we have $Y_{T-k}(x_{1:T-1})=f(X_T)$ with X_T sampled from $\pi_T(\cdot\mid\{x_i\}_{i=1}^{T-1})$, thus (a) holds by definition. When $k=T-2$, we have (a), (b) and (c) are guaranteed exactly by Theorem 1.

Suppose that (a), (b) and (c) are held for $k+1$, where $0\le k\le T-3$. Conditioning on the input history $x_{1:k}$ (sampled from $X_{1:k}$), let us sample x_{k+1} from $\pi_{k+1}\left(\cdot\mid\{x_i\}_{i=1}^{k}\right)$. Then, we sample $N_{k+1}\sim\text{Geo}(r_{k+1})$, and get $i.i.d.$

$$Y_{T-(k+1)}(x_{1:k+1})(1),\ldots,Y_{T-(k+1)}(x_{1:k+1})\left(2^{N_{k+1}}\right).$$

Adapting the same notations as before, we define

$$S_{2^{N_{k+1}}} = \sum_{i=1}^{2^{N_{k+1}}} Y_{T-(k+1)}\left(x_{1:k+1}\right)(i),$$
$$S^O_{2^{N_{k+1}-1}} = \sum_{i=1}^{2^{N_{k+1}-1}} Y_{T-(k+1)}\left(x_{1:k+1}\right)(2i-1),$$
$$S^E_{2^{N_{k+1}-1}} = \sum_{i=1}^{2^{N_{k+1}-1}} Y_{T-(k+1)}\left(x_{1:k+1}\right)(2i).$$

Then, $Y_{T-k}(x_{1:k}) = \Delta_{N_{k+1}}/p_{r_{k+1}}(N_{k+1})$, where $\Delta_{N_{k+1}}$ is defined in Algorithm 2. Note that by the induction hypothesis, we have

$$\mathbb{E}_{\pi_{k+2:T}}\left[Y_{T-(k+1)}(x_{1:k+1})\right] = U_{T-(k+1)}(x_{1:k+1}),$$

and

$$\mathbb{E}_{\pi_{1:T}}\left[|Y_{T-k+1}(x_{1:k+1})|^{2+\delta_{k+1}}\right] < \left(\prod_{i=k+2}^{T-1}\widetilde{C}_i\right)\cdot L^{2+\delta}\left[1+\mathbb{E}\left[\|X_T\|^{2+\delta}\right]\right].$$

We first show that $Y_{T-k}(x_{1:k})$ is an unbiased estimator of $U_{T-k}(x_{1:k})$.

$$\begin{aligned}
&\mathbb{E}_{\pi_{k+1:T}}\left[Y_{T-k}(x_{1:k})\right]\\
&= \sum_{n=0}^{\infty}\mathbb{E}_{\pi_{k+1:T}}[\Delta_n]\\
&= \sum_{n=1}^{\infty}\left(\mathbb{E}_{\pi_{k+1:T}}\left[\max\left\{f\left(x_{k+1}\right), \frac{S_{2^n}}{2^n}\right\}\right] - \mathbb{E}_{\pi_{k+1:T}}\left[\max\left\{f\left(x_{k+1}\right), \frac{S_{2^{n-1}}}{2^{n-1}}\right\}\right]\right) +\\
&\quad \mathbb{E}_{\pi_{k+1:T}}\left[\max\left\{f\left(x_{k+1}\right), Y_{T-(k+1)}(x_{1:k+1})(1)\right\}\right]\\
&= \mathbb{E}_{\pi_{k+1:T}}\left[\max\left\{f\left(x_{k+1}\right), \lim_{n\to\infty}\frac{S_{2^n}}{2^n}\right\}\right]\\
&= \mathbb{E}_{\pi_{k+1:T}}\left[\max\left\{f\left(x_{k+1}\right), U_{T-(k+1)}(x_{1:k+1})\right\}\right] = U_{T-k}(x_{1:k}).
\end{aligned}$$

Next, we bound the expected value $\mathbb{E}_{\pi_{1:T}}\left[|Y_{T-k}(x_{1:k})|^{2+\delta_k}\right]$. For simplicity, in the following proof, we use $\mathbb{P}$ and $\mathbb{E}$ as abbreviations of $\mathbb{P}_{\pi_{1:T}}$ and $\mathbb{E}_{\pi_{1:T}}$.

Following the same idea in the proof of Theorem 1, we introduce three events:

$$E'_1 := \left\{\left|U_{T-(k+1)}(x_{1:k+1}) - f(x_{k+1})\right| < \varepsilon\right\},$$
$$E'_2 := \left\{\left|S^O_{2^{n-1}}/2^{n-1} - U_{T-(k+1)}(x_{1:k+1})\right| \geq \varepsilon/2\right\},$$
$$E'_3 := \left\{\left|S^O_{2^{n-1}}/2^{n-1} - S^E_{2^{n-1}}/2^{n-1}\right| \geq \varepsilon/2\right\}.$$

We start with bounding the probability of event E'_1. Take $\varepsilon = \frac{1}{2^{n/4}}$ (the same as Theorem 1), by Assumption 4, we have

$$\mathbb{P}(E'_1) = \int \mathbb{1}\left(\left\{\left|U_{T-(k+1)}(x_{1:k+1}) - f(x_{k+1})\right| < \varepsilon\right\}\right)\pi_{1:k+1}(dx_1,\ldots,dx_{k+1}) \leq C\varepsilon.$$

Similar to the proof of Theorem 1, by conditioning on $x_{1:k}$, we can apply Lemma 1 to get

$$\begin{aligned}\mathbb{P}_{\pi_{k+1:T}}(E_2') &\le \frac{1}{(\varepsilon/2)^{2+\delta_{k+1}}} \cdot \frac{C_{2+\delta_{k+1}} 2^{1+\delta_{k+1}} \left[1+\mathbb{E}\left[\left|Y_{T-(k+1)}\right|^{2+\delta_{k+1}} \mid x_{1:k}\right]\right]}{2^{(n-1)(2+\delta_{k+1})/2}} \\ &\le C_{2+\delta_{k+1}} 2^{2+3\delta_{k+1}/2}\left[1+\mathbb{E}\left[\left|Y_{T-(k+1)}(x_{1:k+1})\right|^{2+\delta_{k+1}} \mid x_{1:k}\right]\right]\varepsilon.\end{aligned}$$

Thus, we can bound the probability of event E_2' under $\pi_{1:T}$ by

$$\begin{aligned}\mathbb{P}(E_2') &= \int \mathbb{P}_{\pi_{k+1:T}}(E_2')\pi_{1:k}(dx_{1:k}) \\ &\le C_{2+\delta_{k+1}} 2^{2+3\delta_{k+1}/2}\left[1+\int \mathbb{E}\left[\left|Y_{T-(k+1)}(x_{1:k+1})\right|^{2+\delta_{k+1}} \mid x_{1:k}\right]\pi_{1:k}(dx_{1:k})\right]\varepsilon \\ &= C_{2+\delta_{k+1}} 2^{2+3\delta_{k+1}/2}\left[1+\mathbb{E}\left[\left|Y_{T-(k+1)}(x_{1:k+1})\right|^{2+\delta_{k+1}}\right]\right]\varepsilon.\end{aligned}$$

Similarly,

$$\mathbb{P}(E_3') \le C_{2+\delta_{k+1}} 2^{4+5\delta_{k+1}/2}\left[1+\mathbb{E}\left[\left|Y_{T-(k+1)}(x_{1:k+1})\right|^{2+\delta_{k+1}}\right]\right]\varepsilon.$$

Moreover,

$$\begin{aligned}\left[\mathbb{E}\left[|\Delta_n|^{2+\delta_{k+1}}\right]\right]^{\frac{2+\delta_k}{2+\delta_{k+1}}} &\le \left(C_{2+\delta_{k+1}}\left[1+\mathbb{E}\left[\left|Y_{T-(k+1)}(x_{1:k+1})\right|^{2+\delta_{k+1}}\right]\right]\right)^{\frac{2+\delta_k}{2+\delta_{k+1}}} \cdot \frac{1}{(2^n)^{\frac{2+\delta_k}{2}}} \\ &\le \left(C_{2+\delta_{k+1}}\left[1+\mathbb{E}\left[\left|Y_{T-(k+1)}(x_{1:k+1})\right|^{2+\delta_{k+1}}\right]\right]\right)^{\frac{2+\delta_k}{2+\delta_{k+1}}} \cdot \frac{1}{2^n}.\end{aligned}$$

Following the same technique in the proof of Theorem 1, there exists universal constant $C' > 0$ independent of the underlying process and T, such that

$$\begin{aligned}&\mathbb{E}\left[|\Delta_n|^{2+\delta_k}\mathbb{1}(E_1' \cup E_2' \cup E_3')\right] \\ \le\ & C'\left(\left[1+\mathbb{E}\left[\left|Y_{T-(k+1)}(x_{1:k+1})\right|^{2+\delta_{k+1}}\right]\right]\right)^{\frac{2+\delta_k}{2+\delta_{k+1}}+\frac{9\delta_{k+1}}{20+10\delta_{k+1}}} \cdot \frac{1}{2^{\left(1+\frac{9\delta_{k+1}}{80+40\delta_{k+1}}\right)n}} \\ \le\ & C'\left[1+\mathbb{E}\left[\left|Y_{T-(k+1)}(x_{1:k+1})\right|^{2+\delta_{k+1}}\right]\right]\frac{1}{2^{\left(1+\frac{9\delta_{k+1}}{80+40\delta_{k+1}}\right)n}}.\end{aligned}$$

Here we have used the fact that $\frac{2+\delta_k}{2+\delta_{k+1}} + \frac{9\delta_{k+1}}{20+10\delta_{k+1}} = 1$. Noticing that

$$\mathbb{E}\left[|\Delta_n|^{2+\delta_k}\mathbb{1}(E_1'^c \cap E_2'^c \cap E_3'^c)\right] = 0,$$

we get

$$\begin{aligned}&\mathbb{E}\left[|Y_{T-k}(x_{1:k})|^{2+\delta_k}\right] \\ \le\ & 2^{1+\delta_k}\sum_{n=0}^{\infty}\frac{\mathbb{E}\left[|\Delta_n|^{2+\delta_k}\right]}{\left[p_{r_{k+1}}(n)\right]^{1+\delta_k}} \\ =\ & \frac{C'}{r_{k+1}}\left[1+\mathbb{E}\left[\left|Y_{T-(k+1)}(x_{1:k+1})\right|^{2+\delta_{k+1}}\right]\right]\sum_{n=0}^{\infty}\frac{1}{\left(2^{\frac{9\delta_{k+1}/(80+40\delta_{k+1})-\delta_k}{2+\delta_k}}\right)^n} \\ \le\ & \widetilde{C}_{k+1}\left(\prod_{i=k+2}^{T-1}\widetilde{C}_i\right)L^{2+\delta}\left[1+\mathbb{E}\left[\|X_T\|^{2+\delta}\right]\right] = \left(\prod_{i=k+1}^{T-1}\widetilde{C}_i\right)L^{2+\delta}\left[1+\mathbb{E}\left[\|X_T\|^{2+\delta}\right]\right],\end{aligned}$$

where in the last inequality we have applied the induction hypothesis (c). Note that $9\delta_{k+1}/(80+40\delta_{k+1})-\delta_k>0$ when $\delta<1/4$, we have

$$\begin{aligned}
\widetilde{C}_{k+1} :&= \frac{C'}{r_{k+1}}\sum_{n=0}^{\infty}\frac{1}{\left(2^{\frac{9\delta_{k+1}/(80+40\delta_{k+1})-\delta_k}{2+\delta_k}}\right)^n}\\
&= \frac{C'}{r_{k+1}}\cdot\frac{1}{2^{\frac{9\delta_{k+1}/(80+40\delta_{k+1})-\delta_k}{2+\delta_k}}-1}\\
&\le \frac{C'}{r_{k+1}}\cdot\frac{1}{\frac{9\delta_{k+1}/(80+40\delta_{k+1})-\delta_k}{2+\delta_k}} \quad \text{(note that } 2^{\alpha}\ge 1+\alpha \text{ for } \alpha>0)\\
&= \frac{C'}{r_{k+1}}\cdot\frac{(2+\delta_{k+1}/10)(80+40\delta_{k+1})}{9\delta_{k+1}-(8+4\delta_{k+1})\delta_{k+1}} \quad (\delta_k=\delta_{k+1}/10)\\
&\le \frac{C'}{r_{k+1}}\cdot\frac{3\cdot 90}{1-4\delta}\cdot\frac{1}{\delta_{k+1}} \quad (\delta_{k+1}<\delta<1/4)\\
&= \frac{C'}{r_{k+1}}\cdot\frac{3}{\delta(1-4\delta)}\cdot 10^{T-k} \quad (\delta_{k+1}=\delta\cdot 10^{k+2-T})\\
&< \infty
\end{aligned} \tag{B.8}$$

is a constant independent of underlying process. Finally, since we have called Algorithm 2 for $2^{N_{k+1}}$ times to construct $Y_{T-k}(x_{1:k})(i)$ $(1\le i\le 2^{N_{k+1}})$, the expected sampling complexity of computing $Y_{T-k}(x_{1:k})$ is

$$\mathbb{E}[2^{N_{k+1}}]\cdot\prod_{i=k+2}^{T-1}C_i=\prod_{i=k+1}^{T-1}C_i<\infty,$$

where

$$C_{k+1}:=\mathbb{E}[2^{N_{k+1}}]\lesssim\sum_{n=0}^{\infty}\frac{1}{\left(2^{\frac{9\delta_{k+1}/(80+40\delta_{k+1})-\delta_k}{2+\delta_k}}\right)^n}\le\frac{3}{\delta(1-4\delta)}\cdot 10^{T-k}. \tag{B.9}$$

As a result, the expected computational complexity is also finite. To sum up, (a), (b) and (c) are satisfied for k. Thus, the proof by induction is completed. In particular, together with (B.8) and (B.9), there exist universal constant $D>0$ independent of the underlying process and T, such that the resulting estimator Y_T in Algorithm 2 satisfying:

(1) $\mathbb{E}[Y_T]=U_T$
(2) Expected computational complexity is $\prod_{i=1}^{T}C_i<D\cdot 10^{T^2}$ (by (B.9)).
(3) The variance of Y_T is bounded by

$$\left(\prod_{i=k+1}^{T-1}\widetilde{C}_i\right)\cdot L^{2+\delta}\left[1+\mathbb{E}\left[\|X_T\|^{2+\delta}\right]\right]<D\cdot 10^{T^2}\cdot L^{2+\delta}\left[1+\mathbb{E}\left[\|X_T\|^{2+\delta}\right]\right]. \quad \square$$

References

[1] H. Asi, Y. Carmon, A. Jambulapati, Y. Jin, A. Sidford, Stochastic bias-reduced gradient methods, Adv. Neural Inf. Process. Syst. 34 (2021).
[2] S. Becker, P. Cheridito, A. Jentzen, T. Welti, Solving high-dimensional optimal stopping problems using deep learning, 2019, arXiv preprint arXiv:1908.01602.

[3] D. Belomestny, M. Ladkau, J. Schoenmakers, Tight bounds for American options via multilevel Monte Carlo, in: Proceedings of the 2012 Winter Simulation Conference, WSC, IEEE, 2012, pp. 1–8.
[4] D. Belomestny, M. Ladkau, J. Schoenmakers, Multilevel simulation based policy iteration for optimal stopping–convergence and complexity, SIAM/ASA J. Uncertain. Quantif. 3 (1) (2015) 460–483.
[5] D. Belomestny, J. Schoenmakers, F. Dickmann, Multilevel dual approach for pricing American style derivatives, Finance Stoch. 17 (4) (2013) 717–742.
[6] C. Bender, A. Kolodko, J. Schoenmakers, Policy iteration for American options: Overview, Monte Carlo Methods Appl. (2006).
[7] N. Biswas, P.E. Jacob, P. Vanetti, Estimating convergence of Markov chains with L-lag couplings, in: Advances in Neural Information Processing Systems, Vol. 32, 2019.
[8] J.H. Blanchet, P.W. Glynn, Unbiased Monte Carlo for optimization and functions of expectations via multi-level randomization, in: 2015 Winter Simulation Conference, WSC, 2015, pp. 3656–3667.
[9] J.H. Blanchet, P.W. Glynn, Y. Pei, Unbiased multilevel Monte Carlo: Stochastic optimization, steady-state simulation, quantiles, and other applications, 2019, arXiv preprint arXiv:1904.09929.
[10] M. Broadie, P. Glasserman, Pricing American-style securities using simulation, J. Econom. Dynam. Control 21 (8–9) (1997) 1323–1352.
[11] R.W. Chen, B. Rosenberg, L.A. Shepp, A secretary problem with two decision makers, J. Appl. Probab. 34 (4) (1997) 1068–1074.
[12] T.J. DiCiccio, B. Efron, Bootstrap confidence intervals, Statist. Sci. 11 (3) (1996) 189–228.
[13] F. Dickmann, Multilevel Approach for Bermudan Option Pricing (Ph.D. thesis), Duisburg, Essen, 2014.
[14] L.E. Dubins, L.A. Shepp, A.N. Shiryaev, Optimal stopping rules and maximal inequalities for Bessel processes, Theory Probab. Appl. 38 (2) (1994) 226–261.
[15] B. Efron, R.J. Tibshirani, An Introduction to the Bootstrap, CRC Press, 1994.
[16] D. Egloff, Monte Carlo algorithms for optimal stopping and statistical learning, Ann. Appl. Probab. 15 (2) (2005) 1396–1432.
[17] M.B. Giles, Multilevel Monte Carlo path simulation, Oper. Res. 56 (3) (2008) 607–617.
[18] M.B. Giles, Multilevel Monte Carlo methods, Acta Numer. 24 (2015) 259–328.
[19] M.B. Giles, MLMC for nested expectations, in: Contemporary Computational Mathematics- A Celebration of the 80th Birthday of Ian Sloan, Springer, 2018, pp. 425–442.
[20] M.B. Giles, T. Goda, Decision-making under uncertainty: Using MLMC for efficient estimation of EVPPI, Stat. Comput. 29 (4) (2019) 739–751.
[21] M.B. Giles, L. Szpruch, Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation, Ann. Appl. Probab. 24 (4) (2014) 1585–1620.
[22] P. Glasserman, B. Yu, Number of paths versus number of basis functions in American option pricing, Ann. Appl. Probab. 14 (4) (2004) 2090–2119.
[23] P.W. Glynn, C.-h. Rhee, Exact estimation for Markov chain equilibrium expectations, J. Appl. Probab. 51 (A) (2014) 377–389.
[24] S. Heinrich, Multilevel Monte Carlo methods, in: International Conference on Large-Scale Scientific Computing, Springer, 2001, pp. 58–67.
[25] J. Heng, P.E. Jacob, Unbiased Hamiltonian Monte Carlo with couplings, Biometrika 106 (2) (2019) 287–302.
[26] C. Herrera, F. Krach, P. Ruyssen, J. Teichmann, Optimal stopping via randomized neural networks, 2021, arXiv preprint arXiv:2104.13669.
[27] T. Hironaka, M.B. Giles, T. Goda, H. Thom, Multilevel Monte Carlo estimation of the expected value of sample information, SIAM/ASA J. Uncertain. Quantif. 8 (3) (2020) 1236–1259.
[28] L.J. Hong, S. Juneja, Estimating the mean of a non-linear function of conditional expectation, in: Proceedings of the 2009 Winter Simulation Conference, WSC, IEEE, 2009, pp. 1223–1236.
[29] P.E. Jacob, J. O'Leary, Y.F. Atchadé, Unbiased Markov chain Monte Carlo methods with couplings, J. R. Stat. Soc. Ser. B Stat. Methodol. 82 (3) (2020) 543–600.
[30] S. Jain, C.W. Oosterlee, Pricing high-dimensional Bermudan options using the stochastic grid method, Int. J. Comput. Math. 89 (9) (2012) 1186–1211.
[31] F.A. Longstaff, E.S. Schwartz, Valuing American options by simulation: A simple least-squares approach, Rev. Financ. Stud. 14 (1) (2001) 113–147.
[32] J. Marcinkiewicz, A. Zygmund, Quelques théoremes sur les fonctions indépendantes, Fund. Math. 29 (1937) 60–90.
[33] D. McLeish, A general method for debiasing a Monte Carlo estimator, Monte Carlo Methods Appl. 17 (4) (2011) 301–315.

[34] L. Middleton, G. Deligiannidis, A. Doucet, P.E. Jacob, Unbiased Markov chain Monte Carlo for intractable target distributions, Electron. J. Stat. 14 (2) (2020) 2842–2891.
[35] C.-h. Rhee, P.W. Glynn, Unbiased estimation with square root convergence for SDE models, Oper. Res. 63 (5) (2015) 1026–1043.
[36] L. Shepp, A.N. Shiryaev, The Russian option: Reduced regret, Ann. Appl. Probab. 3 (3) (1993) 631–640.
[37] L.A. Shepp, A.N. Shiryaev, A dual Russian option for selling short, Probab. Theory Math. Stat. (1996) 209–218.
[38] J.N. Tsitsiklis, B. Van Roy, Regression methods for pricing complex American-style options, IEEE Trans. Neural Netw. 12 (4) (2001) 694–703.
[39] M. Vihola, Unbiased estimators and multilevel Monte Carlo, Oper. Res. 66 (2) (2018) 448–462.
[40] G. Wang, J. O'Leary, P. Jacob, Maximal couplings of the Metropolis-Hastings Algorithm, in: International Conference on Artificial Intelligence and Statistics, PMLR, 2021, pp. 1225–1233.
[41] G. Wang, T. Wang, Unbiased Multilevel Monte Carlo methods for intractable distributions: MLMC meets MCMC, 2022, arXiv preprint arXiv:2204.04808.
[42] D.Z. Zanger, Quantitative error estimates for a least-squares Monte Carlo algorithm for American option pricing, Finance Stoch. 17 (3) (2013) 503–534.