Weierstraß-Institut
für Angewandte Analysis und Stochastik
Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 2198-5855

Randomized optimal stopping algorithms and their convergence analysis

Christian Bayer[1], Denis Belomestny[2], Paul Hager[3],

Paolo Pigato[4], John G. M. Schoenmakers[1]

submitted: February 28, 2020

[1] Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: christian.bayer@wias-berlin.de
john.schoenmakers@wias-berlin.de

[2] Duisburg-Essen University
Thea-Leymann-Str. 9
45127 Essen
Germany
E-Mail: denis.belomestny@uni-due.de

[3] Technische Universität Berlin
Straße des 17. Juni 136
10623 Berlin
Germany
E-Mail: phager@math.tu-berlin.de

[4] University of Rome Tor Vergata
Via Columbia 2
00133 Roma
Italy
E-Mail: paolo.pigato@uniroma2.it

No. 2697
Berlin 2020

2010 *Mathematics Subject Classification.* 60G40, 65C05.

Key words and phrases. Randomized optimal stopping, convergence rates, Bermudan options.

CB, JS, PH, and PP gratefully acknowledge financial support from the German science foundation (DFG) via the cluster of excellence MATH+, project AA4-2.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Randomized optimal stopping algorithms and their convergence analysis

Christian Bayer, Denis Belomestny, Paul Hager,
Paolo Pigato, John G. M. Schoenmakers

Abstract

In this paper we study randomized optimal stopping problems and consider corresponding forward and backward Monte Carlo based optimization algorithms. In particular we prove the convergence of the proposed algorithms and derive the corresponding convergence rates.

1 Introduction

Optimal stopping problems play an important role in quantitative finance, as some of the most liquid options are of American or Bermudan type, that is, they allow the holder to exercise the option at any time before some terminal time or on a finite, discrete set of exercise times, respectively. Mathematically, the price of an American or Bermudan option is, hence, given as the solution of the optimal stopping problem

$$\sup_{\tau \in \mathcal{T}} \mathsf{E} Z_\tau,$$

where Z_t denotes the discounted payoff to the holder of the option when exercising at time t, and $\mathcal{T}$ denotes the set of all stopping times taking values in $[0,T]$ in the American case and the set of stopping times taking values in the set of exercise dates $\{t_0,\ldots,t_J\}$ in the Bermudan case. Here E stands for the expectation w.r.t. some risk neutral measure. In this paper, we restrict ourselves to the Bermudan case, either because the option under consideration is of Bermudan type or because we have already discretized the American option in time.

1.1 Background

Due to the fundamental importance of optimal stopping in finance and operations research, numerous numerical methods have been suggested. If the dimension of the underlying driving process is high then deterministic methods become inefficient. As a result most state-of-the-art methods are based on the dynamic programming principle combined with Monte Carlo. This class includes regression methods (local or global) [17, 8], mesh methods [11] and optimization based Monte Carlo algorithms [1, 7]. Other approaches include the quantization method [3] and stochastic policy iteration [16] for example. While the above methods aim at constructing (in general suboptimal) policies, hence lower estimations of the optimal stopping problem, the dual approach independently initiated by [19] and [15] has led to a stream of developments for computing upper bounds of the stopping problem (see, for instance, [10] and the references therein).

In this paper, we revisit optimization-based Monte Carlo (OPMC) algorithms and extend them to the case of randomized stopping times. The idea behind OPMC methods is to maximize a MC estimate of

DOI 10.20347/WIAS.PREPRINT.2697

the associated value function over a family of stopping policies thus approximating the early exercise region associated to the optimal stopping problem rather than the value function. This idea goes back to [1]. For a more general formulation see for instance [20], Ch. 5.3, and [7] for a theoretical analysis. One problem of OPMC algorithms is that the corresponding loss functions are usually very irregular, as was observed in [7] and [4]. In order to obtain smooth optimization problems, the authors in [9] and [4] suggested to relax the problem by randomizing the set of possible stopping times. For example, in the continuous exercise case, it was suggested in [4] to stop at the first jump time of a Poisson process with time and state dependent rate. The advantage of this approach is that the resulting optimization problem becomes smooth. In general the solution of the randomized optimal stopping problem coincides with the solution of the standard optimal stopping problem, as earlier observed in [13].

Let us also mention the recent works [5, 6] that use deep neural networks to solve optimal stopping problems numerically. These papers show very good performance of deep neural networks for solving optimal stopping problems, especially in high dimensions. However a complete convergence analysis of these approaches is still missing. A key issue in [5, 6] is a kind of smoothing of the functional representations of exercise boundaries or policies in order to make them suited for the standard gradient based optimization algorithms in the neural network based framework. In fact, we demonstrate that the randomized stopping provides a nice theoretical framework for such smoothing techniques. As such our results, in particular Corollary 4.8, can be interpreted as a theoretical justification of the neural network based methods in [5, 6].

Summing up, the contribution of this paper is twofold. On the one hand, we propose general OPMC methods that use randomized stopping times, instead of the ordinary ones, thus leading to smooth optimization problems. On the other hand, we provide a thorough convergence analysis of the proposed algorithms that justify the use of randomized stopping times.

The structure of the paper is as follows. In Section 2 we introduce the precise probabilistic setting. In the following Section 3 we introduce the forward and the backward Monte Carlo methods. Convergence rates for both methods are stated and proved in Section 4. In Section 5 we describe a numerical implementation and present some numerical results for the Bermudan max-call. Finally, there is an appendix with technical proofs presented in Section A and with a reminder on the theory of empirical processes in Section B.

2 Randomized optimal stopping problems

Let $(\Omega, \mathcal{F}, (\mathcal{F}_j)_{j\geq 0})$ be a given filtered probability space, and $(\Omega_0, \mathcal{B})$ be some auxiliary space that is “rich enough” in some sense. A randomized stopping time τ is defined as a mapping from $\widetilde{\Omega} := \Omega\times\Omega_0$ to $\mathbb{N}$ (including 0) that is measurable with respect to the σ-field $\widetilde{\mathcal{F}} := \sigma\left\{F \times B : F \in \mathcal{F},\ B \in \mathcal{B}\right\},$ and satisfies

$$\{\tau \leq j\} \in \widetilde{\mathcal{F}}_j := \sigma\left\{F \times B : F \in \mathcal{F}_j,\ B \in \mathcal{B}\right\}, \quad j \in \mathbb{N}.$$

While abusing notation a bit, $\mathcal{F}$ and $\mathcal{F}_j$ are identified with $\sigma\left\{F \times \Omega_0 : F \in \mathcal{F}\right\} \subset \widetilde{\mathcal{F}}$ and $\sigma\left\{F \times \Omega_0 : F \in \mathcal{F}_j\right\} \subset \widetilde{\mathcal{F}}_j$, respectively. Let further P be a given probability measure on $(\Omega, \mathcal{F})$, and $\widetilde{\mathsf{P}}$ be its extension to $(\widetilde{\Omega}, \widetilde{\mathcal{F}})$ in the sense that

$$\widetilde{\mathsf{P}}\left(\Omega_0 \times F\right) = \mathsf{P}\left(F\right) \quad \text{for all} \quad F \in \mathcal{F}.$$

In this setup we may think of P as the measure governing the dynamics of some given adapted nonnegative reward process $(Z_j)_{j\geq 0}$ on $(\Omega, \mathcal{F}, (\mathcal{F}_j)_{j\geq 0})$. We then may write (with E denoting the expectation with respect to the "overall measure" $\widetilde{\mathsf{P}}$)

$$\mathsf{E}\left[Z_\tau\right] = \mathsf{E}\left[\sum_{j=0}^{\infty} Z_j p_j\right] \tag{1}$$

with

$$p_j := \mathsf{E}\left[1_{\{\tau=j\}}\middle|\,\mathcal{F}_j\right] = \widetilde{\mathsf{P}}\left(\tau = j|\,\mathcal{F}_j\right).$$

Hence the sequence of nonnegative random variables $p_0, p_1, \ldots,$ is adapted to $(\mathcal{F}_j)_{j\geq 0}$ and satisfies

$$\sum_{j=0}^{\infty} p_j = \sum_{j=0}^{\infty} \mathsf{E}\left[1_{\{\tau=j\}}\middle|\,\mathcal{F}\right] = 1 \text{ a.s.}$$

In this paper we shall study discrete time optimal stopping problems of the form

$$Y_j^\star = \sup_{\tau\in\widetilde{\mathcal{T}}[j,J]} \mathsf{E}[Z_\tau|\mathcal{F}_j], \quad j = 0,\ldots,J, \tag{2}$$

where $\widetilde{\mathcal{T}}[j,J]$ is the set of randomized stopping times taking values in $\{j,\ldots,J\}$. It is well-known (see [9] and [13]) that there exists a family of ordinary stopping times $\tau_j^\star$, $j = 0,\ldots,J$, solving (2) that satisfies the so-called consistency property $\tau_j^\star > j \Longrightarrow \tau_j^\star = \tau_{j+1}^\star$. That is, at the same time,

$$Y_j^\star = \sup_{\tau\in\mathcal{T}[j,J]} \mathsf{E}[Z_\tau|\mathcal{F}_j], \quad j = 0,\ldots,J$$

where $\mathcal{T}[j,J]$ is the set of the (usual) $\mathcal{F}$-stopping times. Studying (2) over a larger class of stopping times is motivated by the fact that the set of randomized stopping times is convex, see [9].

From now on we consider the Markovian case with $Z_j = G_j(X_j)$, where $(X_j)_{j\geq 0}$ is a Markov chain on $(\Omega,\mathcal{F},(\mathcal{F}_j)_{j=0}^J)$ with values in $\mathbb{R}^d$ and $(G_j)_{j\geq 0}$ is a sequence of $\mathbb{R}^d \to \mathbb{R}_+$ functions. We also shall deal with consistent families of randomized stopping times $(\tau_j)_{j=0}^J$ satisfying $j \leq \tau_j \leq J$ with $\tau_J = J$ and $\tau_j > j \Longrightarrow \tau_j = \tau_{j+1}$. Such a consistent family (τ_k), together with a corresponding family of conditional exercise probabilities

$$p_{k,j} := \mathsf{E}\left[1_{\{\tau_k=j\}}\middle|\,\mathcal{F}_j\right], \quad j = k,\ldots,J, \tag{3}$$

may be constructed by backward induction in the following way. We start with $\tau_J = J$ almost surely and set $p_{J,J} = 1$ reflecting the fact that, since $Z \geq 0$ by assumption, stopping at J is optimal provided one did not stop before J. Assume that τ_k with $k \leq \tau_k \leq J$, and $p_{k,j}$, $j = k,\ldots,J$, satisfying (3), are already defined for some k with $0 < k \leq J$.

Next take some uniformly distributed random variable $\mathcal{U} \sim U[0,1]$, independent from $\mathcal{F}$ and τ_k, and a function $h_{k-1} \in \mathcal{H}$ with $\mathcal{H}$ being a class of functions mapping $\mathbb{R}^d$ to $[0,1]$. We then set

$$\tau_{k-1} = \begin{cases} k-1, & \mathcal{U} < h_{k-1}(X_{k-1}), \\ \tau_k, & \mathcal{U} \geq h_{k-1}(X_{k-1}) \end{cases}$$

and

$$p_{k-1,j} = \begin{cases} h_{k-1}(X_{k-1}) & j = k-1 \\ \left(1 - h_{k-1}(X_{k-1})\right) p_{k,j} & j \geq k. \end{cases} \tag{4}$$

Obviously, we then have for $j \geq k$,

$$\begin{aligned}\mathsf{E}\left[1_{\{\tau_{k-1}=j\}}\middle|\,\mathcal{F}_j\right] &= \mathsf{E}\left[1_{\{\tau_k=j\}}\mathsf{E}\left[1_{\mathcal{U}\geq h_{k-1}(X_{k-1})}\middle|\,\mathcal{F}_j\vee\tau_k\right]|\mathcal{F}_j\right]\\ &= (1-h_{k-1}(X_{k-1}))\mathsf{E}\left[1_{\{\tau_k=j\}}\middle|\,\mathcal{F}_j\right]\ = p_{k-1,j}.\end{aligned}$$

That is, (3) with k replaced by $k-1$ is fulfilled.

It is immediately seen that, by the above construction, the thus obtained family of randomized stopping times $(\tau_k)_{k=0}^J$ is consistent, and that

$$\mathsf{E}[Z_{\tau_k}|\mathcal{F}_k] = \mathsf{E}\left[\sum_{j=k}^{\infty} Z_j p_{k,j}\middle|\,\mathcal{F}_k\right] \tag{5}$$

with $h_J \equiv 1$ by definition, and where

$$p_{k,j} = h_j(X_j)\prod_{l=k}^{j-1}(1-h_l(X_l)), \quad j=k,\ldots,J. \tag{6}$$

Hence each (conditional) probability $p_{k,j}$ is a function of $X_k,\ldots,X_j$ by construction, and so in particular it is measurable with respect to the σ- algebra $\mathcal{F}_j$. In view of (1), (2), and (6), we now consider the following optimization problems

$$\overline{Y}_j = \sup_{\boldsymbol{h}\in\mathcal{H}^{J-j}} \mathsf{E}\left[\sum_{l=j}^{J} Z_l h_l(X_l)\prod_{r=j}^{l-1}(1-h_r(X_r))\middle|\,\mathcal{F}_j\right], \quad j=0,\ldots,J-1, \tag{7}$$

where empty products are equal to 1 by definition, and the supremum is taken over vector functions $\boldsymbol{h}=(h_0,\ldots,h_{J-1})\in\mathcal{H}^{J-j}$. It is well known, that the optimal process (Snell envelope) $(Y_j^\star)$ can be attained by using indicator functions (h_j) of the form $h_j(x)=1_{S_j^\star}(x)$ in (7), where the stopping regions $(S_j^\star)$ have the following characterization

$$S_j^\star = \{x\in\mathbb{R}^d : G_j(x)\geq C_j(x)\}, \quad C_j(x)=\mathsf{E}[Y_{j+1}^\star|X_j=x], \quad j=0,\ldots,J-1,$$

with $S_J^\star=\mathbb{R}^d$ by definition. A family of optimal stopping times $(\tau_j^\star)_{j=0,\ldots,J}$ solving (2) can then be defined as a family of first entry times

$$\tau_j^\star = \min\{j\leq l\leq J : X_l\in S_l^\star\}, \quad j=0,\ldots,J. \tag{8}$$

Note that this definition implies that the family $(\tau_j^\star)_{j=0}^J$ is consistent.

3 Monte Carlo optimization algorithms

We now propose two Monte Carlo optimization algorithms for estimating $Y_0^\star$ in (2). The first one (*forward approach*) is based on simultaneous optimization of a Monte Carlo estimate for (7) over the exercise probability functions $h_0,\ldots,h_J$, whereas in the second approach these functions are estimated step by step backwardly from h_J down to h_0 based on (4). The latter procedure is referred to as the *backward approach.*

3.1 Forward approach

Let us consider the empirical counterpart of the optimization problem (7) at time $j = 0$. To this end we generate a set of independent trajectories of the chain (X_j) :

$$\mathcal{D}_M := \left\{(X_0^{(m)}, \ldots, X_J^{(m)}), \quad m = 1, \ldots, M\right\}$$

and consider the optimization problem

$$\sup_{\boldsymbol{h}\in\mathcal{H}^J} \left\{ \frac{1}{M} \sum_{m=1}^{M} \left[\sum_{l=0}^{J} G_l(X_l^{(m)}) h_l(X_l^{(m)}) \prod_{r=0}^{l-1} (1 - h_r(X_r^{(m)})) \right] \right\}. \tag{9}$$

Let $\boldsymbol{h}_M$ be one of its solutions. Next we generate new N independent paths of the chain $(X_j)_{j=0}^J$ and build an estimate for the optimal value $Y_0^\star$ as

$$Y_{M,N} = \frac{1}{N} \sum_{n=1}^{N} \left[\sum_{l=0}^{J} G_l(X_l^{(n)}) h_{M,l}(X_l^{(n)}) \prod_{r=0}^{l-1} (1 - h_{M,r}(X_r^{(n)})) \right] \tag{10}$$

Note that the estimate $Y_{M,N}$ is low biased, that is, $\mathsf{E}[Y_{M,N}|\mathcal{D}_M] \leq Y_0^\star$. The algorithms based on (9) and (10) are referred to as *forward algorithms* in contrast to the *backward algorithms* described in the next section.

In Section 4.1 we shall study the properties of the estimate $Y_{M,N}$ obtained by the forward approach. In particular we there show that $Y_{M,N}$ converges to $Y_0^\star$ as $N, M \to \infty$, and moreover we derive the corresponding convergence rates.

3.2 Backward approach

The forward approach in the previous sections can be rather costly especially if J is large, as it requires optimization over a product space $\mathcal{H}^J$. In this section we propose an alternative approximative method which is based on a backward recursion. Fix again a class $\mathcal{H}$ of functions $h : \mathbb{R}^d \to [0, 1]$. We construct estimates $\widehat{h}_J, \ldots, \widehat{h}_0$ recursively using a set of trajectories

$$D_M := \left\{ \left(X_0^{(m)}, X_1^{(m)}, \ldots, X_J^{(m)}\right), \, m = 1, \ldots, M \right\}.$$

We start with $\widehat{h}_J \equiv 1$. Suppose that $\widehat{h}_k, \ldots, \widehat{h}_J$ are already constructed, then define

$$\widehat{h}_{k-1} := \operatorname*{arg\,sup}_{h\in\mathcal{H}} \widehat{Q}_{k-1}(h, \widehat{h}_k \ldots, \widehat{h}_J) \tag{11}$$

with

$$\widehat{Q}_{k-1}(h_{k-1}, \ldots, h_J) := \\ = \frac{1}{M} \sum_{m=1}^{M} \left[\sum_{j=k-1}^{J} G_j(X_j^{(m)}) h_j(X_j^{(m)}) \prod_{l=k-1}^{j-1} (1 - h_l(X_l^{(m)})) \right] \tag{12}$$

in view of (7).

Remark 3.1. *Note that the optimal functions $h_j^\star(x) = 1_{\{x \in S_j^\star\}}, j = 0, \ldots, J-1,$ can be sequentially constructed via relations*

$$h_{k-1}^\star := \arg\sup_{h \in \mathcal{H}} [Q_{k-1}(h, h_k^\star, \ldots, h_J^\star)], \quad h_J^\star \equiv 1,$$

where

$$Q_{k-1}(h_{k-1}, h_k, \ldots, h_J) := \mathsf{E}\left[\sum_{j=k-1}^{J} Z_j h_j(X_j) \prod_{l=k-1}^{j-1} (1 - h_l(X_l))\right], \tag{13}$$

(see also (5) and (6)), provided that $h_1^\star, \ldots, h_J^\star \in \mathcal{H}$. This fact was used in [6] to construct approximations for $h_j^\star, j = 0, \ldots, J,$ via neural networks. Although it might seem appealing to consider classes of functions $\mathcal{H} : \mathbb{R}^d \to \{0, 1\}$, this may lead to nonsmooth and nonconvex optimization problems. Here we present general framework allowing us to balance between smoothness of the class $\mathcal{H}$ and its ability to approximate $h_j^\star, j = 0, \ldots, J,$ see Section 4.2.

Working all the way back we thus end up with a sequence $\widehat{h}_J, \ldots, \widehat{h}_0$ and, similar to (10), may next obtain a low-biased approximation $\widehat{Y}_{M,N}$ via an independent re-simulation with sample size N. By writing

$$\begin{aligned} &\widehat{Q}_{k-1}(h_{k-1}, \ldots, h_J) = \\ &\quad \frac{1}{M} \sum_{m=1}^{M} \left(G_{k-1}(X_{k-1}^{(m)}) - \sum_{j=k}^{J} G_j(X_j^{(m)}) h_j(X_j^{(m)}) \prod_{l=k}^{j-1} (1 - h_l(X_l^{(m)})) \right) h_{k-1}(X_{k-1}^{(m)}) \\ &\qquad + \frac{1}{M} \sum_{m=1}^{M} \sum_{j=k}^{J} G_j(X_j^{(m)}) h_j(X_j^{(m)}) \prod_{l=k}^{j-1} (1 - h_l(X_l^{(m)})) \end{aligned}$$

we see that the backward step (11)-(12) is equivalent to

$$\begin{aligned} \widehat{h}_{k-1} &= \arg\sup_{h \in \mathcal{H}} \widehat{Q}_{k-1}(h, \widehat{h}_k \ldots, \widehat{h}_J) \\ &= \arg\sup_{h \in \mathcal{H}} \sum_{m=1}^{M} h(X_{k-1}^{(m)}) \\ &\quad \times \left(G_{k-1}(X_{k-1}^{(m)}) - \sum_{j=k}^{J} G_j(X_j^{(m)}) \widehat{h}_j(X_j^{(m)}) \prod_{l=k}^{j-1} (1 - \widehat{h}_l(X_l^{(m)})) \right) \\ &=: \arg\sup_{h \in \mathcal{H}} \sum_{m=1}^{M} \xi_{k-1}^{(m)} h(X_{k-1}^{(m)}) \end{aligned} \tag{14}$$

Advantage of the backward algorithm is its computational efficiency: in each step of the algorithm we need to perform optimization over a space $\mathcal{H}$ and not over the product space $\mathcal{H}^J$ as in the forward approach.

4 Convergence analysis

In this section we provide a convergence analysis of the procedures introduced in Section 3.1 and Section 3.2 respectively.

4.1 Convergence analysis of the forward approach

We make the following assumptions.

(H) Denote for any $\boldsymbol{h}_1, \boldsymbol{h}_2 \in \mathcal{H}^J$,

$$d_X(\boldsymbol{h}_1, \boldsymbol{h}_2) := \sqrt{\mathsf{E}\left[\left|\sum_{j=0}^{J-1} |h_{1,j}(X_j) - h_{2,j}(X_j)| \prod_{l=0}^{j-1}(1 - h_{2,l}(X_l))\right|^2\right]}$$

Assume that the class of functions $\mathcal{H}$ is such that

$$\log[\mathcal{N}(\delta, \mathcal{H}^J, d_X)] \leq A\delta^{-\rho} \tag{15}$$

for some constant $A > 0$, any $0 < \delta < 1$ and some $\rho \in (0, 2)$, where $\mathcal{N}$ is the minimal number (covering number) of closed balls of radius δ (w.r.t. d_X) needed to cover the class $\mathcal{H}$.

(G) Assume that all functions G_j, $j = 0, \ldots, J$, are uniformly bounded by a constant C_Z.

(B) Assume that the inequalities

$$\mathsf{P}\big(|G_j(X_j) - C_j(X_j)]| \leq \delta\big) \leq A_{0,j}\delta^{\alpha}, \quad \delta < \delta_0 \tag{16}$$

hold for some $\alpha > 0$, $A_{0,j} > 0$, $j = 1, \ldots, J-1$, and $\delta_0 > 0$.

Remark 4.1. *Note that*

$$d_X(\boldsymbol{h}_1, \boldsymbol{h}_2) \leq \sum_{j=0}^{J-1} \|h_{1,j} - h_{2,j}\|_{L_2(P_{X_j})},$$

where P_{X_i} stands for the distribution of X_i. Hence (15) *holds if*

$$\max_{j=0,\ldots,J-1} \log[\mathcal{N}(\delta, \mathcal{H}, L_2(P_{X_j}))] \leq \left(\frac{A'}{\delta}\right)^{\rho} \tag{17}$$

for some constant $A' > 0$.

Theorem 4.2. *Assume that assumptions (**H**), (**G**) and (**B**) hold. Then for any $U > U_0$ and $M > M_0$ it holds with probability at least $1 - \delta$,*

$$0 \leq Y_0^\star - \mathsf{E}[Y_{M,N}|\mathcal{D}_M] \leq C\left(\frac{\log^2(1/\delta)}{M}\right)^{\frac{1+\alpha}{2+\alpha(1+\rho)}} \tag{18}$$

with some constants $U_0 > 0$, $M_0 > 0$ and $C > 0$, provided that

$$0 \leq Y_0^\star - \overline{Y}_0 \leq DM^{-1/(1+\rho/2)}, \tag{19}$$

for a constant $D > 0$, where $\overline{Y}_0$ is defined in (7)*.*

Proof. Denote

$$\mathcal{Q}(\boldsymbol{h}) := \mathsf{E}\left[\sum_{j=0}^{J} Z_j p_j(\boldsymbol{h})\right], \quad \Delta(\boldsymbol{h}) := \mathcal{Q}(\boldsymbol{h}^\star) - \mathcal{Q}(\boldsymbol{h})$$

with

$$p_j(\boldsymbol{h}) := h_j(X_j)\prod_{l=0}^{j-1}(1-h_l(X_l)), \quad j = 0,\ldots,J.$$

Define also $\Delta_M(\boldsymbol{h}) := \sqrt{M}(\mathcal{Q}_M(\boldsymbol{h}) - \mathcal{Q}(\boldsymbol{h}))$ with

$$\mathcal{Q}_M(\boldsymbol{h}) := \frac{1}{M}\sum_{m=1}^{M}\left[\sum_{j=0}^{J} Z_j^{(m)} h_j(X_j^{(m)})\prod_{l=0}^{j-1}(1-h_l(X_l^{(m)}))\right]$$

and put $\Delta_M(\boldsymbol{h}', \boldsymbol{h}) := \Delta_M(\boldsymbol{h}') - \Delta_M(\boldsymbol{h})$. Let $\overline{\boldsymbol{h}}$ be one of the solutions of the optimization problem $\sup_{\boldsymbol{h}\in\mathcal{H}^J}\mathcal{Q}(\boldsymbol{h})$. Since $\mathcal{Q}_M(\boldsymbol{h}_M) \geq \mathcal{Q}_M(\overline{\boldsymbol{h}})$, we obviously have

$$\Delta(\boldsymbol{h}_M) \leq \Delta(\overline{\boldsymbol{h}}) + \frac{\left[\Delta_M(\boldsymbol{h}^\star, \overline{\boldsymbol{h}}) + \Delta_M(\boldsymbol{h}_M, \boldsymbol{h}^\star)\right]}{\sqrt{M}}. \tag{20}$$

Set $\varepsilon_M = M^{-1/(2+\rho)}$ and derive

$$\begin{aligned}\Delta(\boldsymbol{h}_M) \leq \Delta(\overline{\boldsymbol{h}}) &+ \frac{2}{\sqrt{M}}\sup_{\boldsymbol{h}\in\mathcal{H}^J:\,\Delta_X(\boldsymbol{h}^\star,\boldsymbol{h})\leq\varepsilon_M}|\Delta_M(\boldsymbol{h}^\star,\boldsymbol{h})| \\ &+ 2\frac{\Delta_X^{(1-\rho/2)}(\boldsymbol{h}^\star,\boldsymbol{h}_M)}{\sqrt{M}}\sup_{\boldsymbol{h}\in\mathcal{H}^J:\,\Delta_X(\boldsymbol{h}^\star,\boldsymbol{h})>\varepsilon_M}\left[\frac{|\Delta_M(\boldsymbol{h}^\star,\boldsymbol{h})|}{\Delta_X^{(1-\rho/2)}(\boldsymbol{h}^\star,\boldsymbol{h})}\right], \end{aligned} \tag{21}$$

where

$$\Delta_X(\boldsymbol{h}_1, \boldsymbol{h}_2) := \sqrt{\mathsf{E}\left[\left|\sum_{j=0}^{J} Z_j p_j(\boldsymbol{h}_1) - \sum_{j=0}^{J} Z_j p_j(\boldsymbol{h}_2)\right|^2\right]}, \quad \boldsymbol{h}_1, \boldsymbol{h}_2 \in \mathcal{H}^J.$$

The reason behind splitting the r. h. s. of (20) into two parts is that the behavior of the empirical process $\Delta_M(\boldsymbol{h}^*, \boldsymbol{h})$ is different on the sets $\{\boldsymbol{h}\in\mathcal{H}^J : \Delta_X(\boldsymbol{h}^*,\boldsymbol{h}) > \varepsilon_M\}$ and $\{\boldsymbol{h}\in\mathcal{H}^J : \Delta_X(\boldsymbol{h}^*,\boldsymbol{h}) \leq \varepsilon_M\}$. The following lemma holds.

Lemma 4.3. *It holds*

$$\Delta_X(\boldsymbol{h}_1, \boldsymbol{h}_2) \leq C_Z d_X(\boldsymbol{h}_1, \boldsymbol{h}_2)$$

for any $\boldsymbol{h}_1, \boldsymbol{h}_2 \in \mathcal{H}^J$.

Lemma 4.3 immediately implies that

$$\log\big(\mathcal{N}(\delta, \mathcal{H}^J, \Delta_X)\big) \leq J\log\big(\mathcal{N}(\delta, \mathcal{H}, d_X)\big) \tag{22}$$

where $\mathcal{N}(\delta,\mathcal{S},d)$ is the covering number of a class $\mathcal{S}$ w.r.t. the pseudo-distance d. Hence due to the assumption (**H**) we derive from [22] that for any $\boldsymbol{h}\in\mathcal{H}^J$ and any $U>U_0$,

$$\mathsf{P}\left(\sup_{\boldsymbol{h}'\in\mathcal{H}^J,\,\Delta_X(\boldsymbol{h},\boldsymbol{h}')\leq\varepsilon_M}|\Delta_M(\boldsymbol{h},\boldsymbol{h}')|>U\varepsilon_M^{1-\rho/2}\right)\leq C\exp\left(-\frac{U}{\varepsilon_M^{\rho}C^2}\right) \tag{23}$$

and

$$\mathsf{P}\left(\sup_{\boldsymbol{h}'\in\mathcal{H}^J,\,\Delta_X(\boldsymbol{h},\boldsymbol{h}')>\varepsilon_M}\frac{|\Delta_M(\boldsymbol{h},\boldsymbol{h}')|}{\Delta_X^{1-\rho/2}(\boldsymbol{h},\boldsymbol{h}')}>U\right)\leq C\exp(-U/C^2), \tag{24}$$

$$\mathsf{P}\left(\sup_{\boldsymbol{h}'\in\mathcal{H}^J}|\Delta_M(\boldsymbol{h},\boldsymbol{h}')|>z\sqrt{M}\right)\leq C\exp(-Mz^2/C^2B) \tag{25}$$

with some constants $C>0,\ B>0$ and $U_0>0$. To simplify notations denote

$$\mathcal{W}_{1,M}:=\sup_{\boldsymbol{h}\in\mathcal{H}^J:\,\Delta_X(\boldsymbol{h}^*,\boldsymbol{h})\leq\varepsilon_M}|\Delta_M(\boldsymbol{h}^*,\boldsymbol{h})|,$$
$$\mathcal{W}_{2,M}:=\sup_{\boldsymbol{h}\in\mathcal{H}^J:\,\Delta_X(\boldsymbol{h}^*,\boldsymbol{h})>\varepsilon_M}\frac{|\Delta_M(\boldsymbol{h}^*,\boldsymbol{h})|}{\Delta_X^{(1-\rho/2)}(\boldsymbol{h}^*,\boldsymbol{h})}$$

and set $\mathcal{A}_0:=\{\mathcal{W}_{1,M}\leq U\varepsilon_M^{1-\rho/2}\}$ for some $U>U_0$. Then the inequality (23) leads to the estimate

$$\mathsf{P}(\bar{\mathcal{A}}_0)\leq C\exp(-U\varepsilon_M^{-\rho}/C^2).$$

Furthermore, since $\Delta(\overline{\boldsymbol{h}})\leq DM^{-1/(1+\rho/2)}$ and $\varepsilon_M^{1-\rho/2}/\sqrt{M}=M^{-1/(1+\rho/2)}$, we get on $\mathcal{A}_0$

$$\Delta(\boldsymbol{h}_M)\leq C_0M^{-1/(1+\rho/2)}+2\frac{\Delta_X^{(1-\rho/2)}(\boldsymbol{h}^*,\boldsymbol{h}_M)}{\sqrt{M}}\,\mathcal{W}_{2,M} \tag{26}$$

with $C_0=D+2U$. Now we need to find a bound for $\Delta_X(\boldsymbol{h}^*,\boldsymbol{h}_M)$ in terms of $\Delta(\boldsymbol{h}_M)$. This is exactly the place, where the condition (16) is used. The following proposition holds.

Proposition 4.4. *Suppose that the assumption (**BA**) holds for $\delta>0$, then there exists a constant $c_\alpha>0$ not depending on J such that for all $\boldsymbol{h}\in\mathcal{H}^J$ it holds*

$$\Delta_X(\boldsymbol{h}^\star,\boldsymbol{h})\leq c_\alpha\sqrt{J}\Delta^{\alpha/(2(1+\alpha))}(\boldsymbol{h}),$$

where c_α depends on α only.

Let us introduce a set

$$\mathcal{A}_1:=\left\{\Delta(\boldsymbol{h}_M)>C_0(1-\varkappa)^{-1}M^{-1/(1+\rho/2)}\right\}$$

for some $0<\varkappa<1$. Thus we get on $\mathcal{A}_0\cap\mathcal{A}_1$

$$\Delta(\boldsymbol{h}_M)\leq C_1\frac{\Delta^{\alpha(1-\rho/2)/(2(1+\alpha))}(\boldsymbol{h}_M)}{\varkappa\sqrt{M}}\,\mathcal{W}_{2,M},$$

where the constant C_1 depends on α but not on ρ. Therefore

$$\Delta(\boldsymbol{h}_M)\leq(\varkappa/C_1)^{-\nu}M^{-\nu/2}\mathcal{W}_{2,M}^{\nu}$$

with $\nu := \frac{2(1+\alpha)}{2+\alpha(1+\rho/2)}$. Applying inequality (24) to $\mathcal{W}^{\nu}_{2,M}$ and using the fact that $\nu/2 \le 1/(1+\rho/2)$ for all $0 < \rho \le 2$, we finally obtain the desired bound for $\Delta(\boldsymbol{h}_M)$,

$$\begin{aligned}
\mathsf{P}\left(\left\{\Delta(\boldsymbol{h}_M) > (V/M)^{\nu/2}\right\} \cap \mathcal{A}_1\right) \le \\
\mathsf{P}\left(\left\{\Delta(\boldsymbol{h}_M) > (V/M)^{\nu/2}\right\} \cap \mathcal{A}_0 \cap \mathcal{A}_1\right) + \mathsf{P}(\bar{\mathcal{A}}_0) \\
\le C\exp(-\sqrt{V}/B) + C\exp\left(-UM^{\rho/(2+\rho)}/C^2\right)
\end{aligned}$$

which holds for all $V > V_0$ and $M > M_0$ with some constant B depending on ρ and α. □

4.2 Convergence analysis of the backward approach

In this section we study the properties of the backward algorithm and prove its convergence. The following result holds.

Proposition 4.5. *For any $k > 1$ and any $h_{k-1}, \ldots, h_J \in \mathcal{H}$, one has that*

$$\begin{aligned}
0 \le Q_{k-1}(h^\star_{k-1}, \ldots, h^\star_J) - Q_{k-1}(h_{k-1}, \ldots, h_J) \le Q_k(h^\star_k, \ldots, h^\star_J) - Q_k(h_k, \ldots, h_J) \\
+ \mathsf{E}\left[\left(Z_{k-1} - C^\star_{k-1}\right)\left(h^\star_{k-1}(X_{k-1}) - h_{k-1}(X_{k-1})\right)\right].
\end{aligned}$$

Note that $Z_{k-1} - C^\star_{k-1} \ge 0$ if $h^\star_{k-1}(X_{k-1}) = 1$ and $Z_{k-1} - C^\star_{k-1} < 0$ if $h^\star_{k-1}(X_{k-1}) = 0$ due to the dynamic programming principle.

This implies the desired convergence.

Theorem 4.6. *Assume (**G**) and suppose that*

$$\max_{j=0,\ldots,J-1} \mathcal{N}(\delta, \mathcal{H}, L_2(P_{X_j})) \le \left(\frac{\mathcal{A}}{\delta}\right)^V \tag{27}$$

holds for some $V > 0$ and $\mathcal{A} > 0$. Then with probability at least $1-\delta$, and $k = 1, \ldots, J$,

$$\begin{aligned}
0 \le Q_{k-1}(h^\star_{k-1}, \ldots, h^\star_J) - Q_{k-1}(\widehat{h}_{k-1}, \ldots, \widehat{h}_J) \lesssim J\sqrt{\frac{V\log(J\mathcal{A})}{M}} + \\
+ J\sqrt{\frac{\log(1/\delta)}{M}} + \sum_{l=k-1}^{J-1} \inf_{h\in\mathcal{H}} \mathsf{E}\left[\left(Z_l - C^\star_l\right)\left(h^\star_l(X_l) - h(X_l)\right)\right],
\end{aligned} \tag{28}$$

where $\lesssim$ stands for inequality up to a constant depending on C_Z.

Remark 4.7. *A simple inequality*

$$\sum_{l=k-1}^{J-1} \inf_{h\in\mathcal{H}} \mathsf{E}\left[\left(Z_l - C^\star_l\right)\left(h^\star_l(X_l) - h(X_l)\right)\right] \le C_Z \sum_{l=k-1}^{J-1} \inf_{h\in\mathcal{H}} \|h_l - h^\star_l\|_{L_2(P_{X_l})}$$

shows that we can chose class $\mathcal{H}$ to minimise $\inf_{h\in\mathcal{H}} \|h - h^\star_j\|_{L_2(P_{X_j})}$. Consider classes $\mathcal{H}$ of the form:

$$\mathcal{H}_{n,r}(R) := \left\{\sum_{i=1}^n a_i\psi(A_i x + b),\ A_i \in \mathbb{R}^{r\times d},\ a_i \in \mathbb{R},\ b \in \mathbb{R}^r,\ \sum_{i=1}^n |a_i| \le R\right\}$$

where $\psi : \mathbb{R}^r \to \mathbb{R}$ is infinitely many times differentiable in some open sphere in $\mathbb{R}^r$ and $r \leq d$. Then according to Corollary B.3,

$$\inf_{h\in\mathcal{H}_{n,r}(R)} \|h - h_j^\star\|_{L_2(P_{X_j})} \leq C(d,\delta)\, n^{-\frac{1}{2d}+\delta}$$

for arbitrary small $\delta > 0$ and large enough $R > 0$, provided that each measure P_{X_j} is regular in the sense that

$$P_{X_j}\left(S_j^\star \setminus (S_j^\star)^\varepsilon\right) \leq a_d\, \varepsilon, \tag{29}$$

where for any set A in $\mathbb{R}^d$ we denote by A^ε the set $A^\varepsilon := \{x \in \mathbb{R}^d : \operatorname{dist}(x, A) \leq \varepsilon\}$. Moreover, it is not difficult to see that (27) *holds for $\mathcal{H}_{n,r}$ with V proportional to n and $\mathcal{A}$ proportional to R, see Lemma 16.6 in [14].*

Corollary 4.8. *Remark 4.7 implies that the second term in* (28) *converges to 0 at the rate $O(n^{-1/(2d)+\delta})$ as $n \to \infty$, where n is the number of neurons in the approximating neural network. On the other hand, the constant $\mathcal{A}$ in* (27) *is proportional to n, so that we get*

$$Q_{k-1}(h_{k-1}^\star, \ldots, h_J^\star) - Q_{k-1}(\widehat{h}_{k-1}, \ldots, \widehat{h}_J) \lesssim J\sqrt{\frac{nJ\log(\mathcal{A})}{M}} + \\ + J\sqrt{\frac{\log(1/\delta)}{M}} + \frac{J}{n^{1/(2d)}}.$$

By balancing two errors we arrive at the bound

$$0 \leq Q_{k-1}(h_{k-1}^\star, \ldots, h_J^\star) - Q_{k-1}(\widehat{h}_{k-1}, \ldots, \widehat{h}_J) \lesssim J\left(\sqrt{\frac{J}{M^{1+1/d}}} + \sqrt{\frac{\log(1/\delta)}{M}}\right)$$

with probability $1 - \delta$. In fact, this gives the overall error bounds for the case of one layer neural networks based approximations.

Proof. Using (11), (12), (13), we have

$$\begin{aligned}
0 &\leq Q_{k-1}(h_{k-1}^\star, \ldots, h_J^\star) - Q_{k-1}(\widehat{h}_{k-1}, \ldots, \widehat{h}_J) \\
&= \inf_{h_{k-1}\in\mathcal{H}} \Big(Q_{k-1}(h_{k-1}^\star, \ldots, h_J^\star) - \widehat{Q}_{k-1}(h_{k-1}, \widehat{h}_k, \ldots, \widehat{h}_J)\Big) \\
&+ \widehat{Q}_{k-1}(\widehat{h}_{k-1}, \ldots, \widehat{h}_J) - Q_{k-1}(\widehat{h}_{k-1}, \ldots, \widehat{h}_J) \\
&\leq \inf_{h_{k-1}\in\mathcal{H}} \Big(Q_{k-1}(h_{k-1}^\star, \ldots, h_J^\star) - Q_{k-1}(h_{k-1}, \widehat{h}_k, \ldots, \widehat{h}_J)\Big) \\
&+ 2 \sup_{h_{k-1},\ldots,h_J\in\mathcal{H}} \Big|Q_{k-1}(h_{k-1}, h_k, \ldots, h_J) - \widehat{Q}_{k-1}(h_{k-1}, h_k, \ldots, h_J)\Big| \\
&\leq Q_k(h_k^\star, \ldots, h_J^\star) - Q_k(\widehat{h}_k, \ldots, \widehat{h}_J) \\
&+ \inf_{h_{k-1}\in\mathcal{H}} \mathsf{E}\left[\left(Z_{k-1} - C_{k-1}^\star\right)\left(h_{k-1}^\star(X_{k-1}) - h_{k-1}(X_{k-1})\right)\right] &(30)\\
&+ 2 \sup_{h_{k-1},\ldots,h_J\in\mathcal{H}} \Big|Q_{k-1}(h_{k-1}, h_k, \ldots, h_J) - \widehat{Q}_{k-1}(h_{k-1}, h_k, \ldots, h_J)\Big|, &(31)
\end{aligned}$$

where the last step follows from Proposition 4.5. Set

$$g_{\mathbf{h}}(X_{k-1}, \ldots, X_J) = \sum_{j=k-1}^{J} Z_j h_j(X_j) \prod_{l=k-1}^{j-1} (1 - h_l(X_l)), \quad \mathbf{h} = (h_{k-1}, \ldots, h_J),$$

then

$$\widehat{Q}_{k-1}(h_{k-1},\ldots,h_J)-Q_{k-1}(h_{k-1},\ldots,h_J) = \frac{1}{M}\sum_{m=1}^{M}\left\{g_{\mathbf{h}}(X^{(m)}_{k-1},\ldots,X^{(m)}_J)-\mathsf{E}[g_{\mathbf{h}}(X^{(m)}_{k-1},\ldots,X^{(m)}_J)]\right\}.$$

Now consider the class $\mathcal{G}:=\{g_{\mathbf{h}},\,\mathbf{h}\in\mathcal{H}^{(J-k+1)}\}$ of uniformly bounded functions on $\mathbb{R}^{d(J-k+1)}$. Indeed we have $|g_{\mathbf{h}}|\le C_Z$. Moreover

$$\mathcal{N}(\delta,\mathcal{G},L_2(P))\le(\mathcal{A}/\delta)^{JV} \tag{32}$$

under (27). Denote

$$Z=\sqrt{M}\sup_{h_{k-1},\ldots,h_J\in\mathcal{H}}\left|Q_{k-1}(h_{k-1},h_k,\ldots,h_J)-\widehat{Q}_{k-1}(h_{k-1},h_k,\ldots,h_J)\right|,$$

then the Talagrand inequality (see [21] and [12]) yields

$$\mathsf{P}(Z\ge\mathsf{E}[Z]+\sqrt{x(4C_Z\mathsf{E}[Z]+M)}+C_Zx/3)\le e^{-x},$$

where

$$\mathsf{E}[Z]\le\sqrt{MJ\log(\mathcal{A}C_Z)}$$

provided (32) holds. □

5 Implementation of the Bermudan max-call

In this section we implement the pricing of the Bermudan max-call, a benchmark example in the literature [2]. As underlying we take a d-dimensional Black & Scholes model, with log-price dynamics given by

$$dX^i_t=\sigma dW^i_t+(r-\delta-\frac{\sigma^2}{2})dt,\quad i=1,\ldots,d,$$

where $X^i_0=0$ and W^i, $i=1,\ldots,d$, are independent Brownian Motions. Parameters σ,r,δ represent respectively volatility, interest, and dividend rate. The corresponding stock prices are given by $S^i_t=S^i_0\exp(X^i_t)$, $t\in[0,T]$. Our goal is the price of a Bermudan max-call option, given by the following representation,

$$\sup_{\tau\in\mathcal{T}}E[e^{-r\tau}\max_{i=1,\ldots,d}(S^i_\tau-K)_+],$$

where $\mathcal{T}$ is the set of stopping times in $\{t_0=0,t_1,t_2,\ldots,t_J=T\}$ adapted to the process $X=(X^1,\ldots,X^d)$, and $(\cdot)_+$ denotes the positive part.

5.1 Backward approach

In order to implement a randomized stopping strategy we need some suitable parametric choice for h. For example we may take $h_l(x)$ to be the composition of a polynomial in x with the logistic function $\frac{e^x}{1+e^x}$ (cf. [6, Framework 3.1] and [5, Section 2.2]), i.e.

$$h_l(x)=h_{\theta_l}(x)=\frac{e^{pol^g_{\theta_l}(x)}}{1+e^{pol^g_{\theta_l}(x)}},\quad l=0,\ldots,J, \tag{33}$$

where $pol^g_\theta(x)$ is a polynomial of degree g in x and θ is the vector of its coefficients. As another example we may compose polynomials in x with a Gumble type distribution function $\beta(x) := 1 - \exp(-\exp(x))$, i.e.

$$h_l(x) = h_{\theta_l}(x) = 1 - \exp(-\exp(pol^g_{\theta_l}(x))), \quad l = 0, ..., J. \tag{34}$$

Both functions are smooth approximations of the indicator function. Let us choose the latter and carry out the backward procedure in Section 3.2. Assume that we have already determined parameters $\widehat{\theta}_k, \ldots, \widehat{\theta}_{J-1}$ (hence the corresponding functions $\widehat{h}_k, \ldots, \widehat{h}_{J-1}$, $\widehat{h}_J = h^\star_J = 1$). Now, according to (14), the estimate of $\widehat{\theta}_{k-1}$ is given by the maximization over θ of the function

$$\widehat{L}_{k-1}(h_\theta, \widehat{h}_k, \ldots, \widehat{h}_J) = \sum_{m=1}^{M} \xi^{(m)}_{k-1} h_\theta(X^{(m)}_{k-1}), \tag{35}$$

where $\xi^{(m)}_{k-1}$ is as in (14). The corresponding gradient is given by

$$\nabla_\theta \widehat{L}_{k-1}(h_\theta, \widehat{h}_k, \ldots, \widehat{h}_J) = \sum_{m=1}^{M} \xi^{(m)}_{k-1} \nabla_\theta h_\theta(X^{(m)}_{k-1}). \tag{36}$$

The parametric choice (34) allows for the following representation of the θ-gradient. We may write straightforwardly

$$\nabla_\theta h_\theta(x) = (1 - h_\theta(x)) \exp\left(pol^g_\theta(x)\right)\, \nabla_\theta pol^g_\theta(x) \tag{37}$$

and since θ is the vector of the coefficients of pol^g_θ, the gradient $\nabla_\theta pol_\theta$ is the vector of monomials in x of degree less or equal than g. Injecting this representation in (36) we get an explicit expression for the gradient of the objective function that we can feed into the optimization algorithm. In fact, the catch of the randomization is the smoothness of h_θ in (34) with respect to θ. This in turn allows for gradient based optimization procedures with explicitly given objective function and gradient. However, a non-trivial issue is how to find the global maximum, at each step, of the function $\widehat{L}$. This is also a well know issue in machine learning, see for instance [6, Section 2.6] or [5, Section 2.3]. We do not dig into this question in the present paper and just refer to [5, 6] for relevant literature.

5.2 Forward approach

We can alternatively write $h_j(X_j) = h(X_j, t_j)$ with $h(x,t) = h_\theta(x,t)$ a function depending on a parameter θ to be optimized. In this case we use the forward approach, since the backward induction cannot be used with this type of parametrization.

As an example (analogous to (34)), we consider

$$h_\theta(x,t) = 1 - \exp(-\exp(pol^g_\theta(x,t))), \tag{38}$$

with pol^g_θ polynomial of degree g in x and t, from which

$$\begin{aligned}\nabla_\theta h(x,t) &= \exp(-\exp(pol^g_\theta(x,t))) \exp(pol^g_\theta(x,t)) \nabla_\theta pol^g_\theta(x,t)\\ &= (1 - h_\theta(x,t)) \exp(pol^g_\theta(x,t)) \nabla_\theta pol^g_\theta(x,t)\end{aligned}$$

As before, we want to maximize over θ the payoff

$$\mathcal{P} = E\Big[\sum_{j=1}^{J} Z_j(X_j) p^\theta_j(X_1, \ldots, X_j)\Big].$$

S_0	K	Backward, $g=3$	Forward, $g=4$	NN price in [6]	95% CI in [2]
90	100	8.072	8.055	8.072	$[8.053, 8.082]$
100	100	13.728	13.882	13.899	$[13.892, 13.934]$

Table 1: Bermudan max-call prices for Black-Scholes model, with $d = 2, T = 3, J = 9$ and $r = 0.05, \delta = 0.1, \sigma = 0.2$.

We have

$$\nabla_\theta \mathcal{P} = E\Big[\sum_{j=1}^{J} Z_j(X_j)\nabla_\theta p_j^\theta(X_1,\dots,X_j)\Big]$$

with $p_j^\theta(X_1,\dots,X_j)$ as in (6). Explicit computations give now

$$\nabla_\theta p_j^\theta(X_1,\dots,X_j) = p_j^\theta(X_1,\dots,X_j)$$
$$\left(\frac{1}{h_\theta(X_j,t_j)}\exp(pol_\theta^g(X_j,t_j))\nabla_\theta pol_\theta^g(X_j,t_j) - \sum_{l=1}^{j}\exp(pol_\theta^g(X_l,t_l))\nabla_\theta pol_\theta^g(X_l,t_l)\right)$$

for $j = 1,\dots,J$. We can compute $\nabla_\theta \mathcal{P}$ and use in the optimization this explicit expression for the gradient of the loss function.

5.3 Numerical results

We take parameters $r = 0.05, \delta = 0.1, \sigma = 0.2$ (as in [2, 6]). We first compute the stopping functions h in (34) using the backward method with $M = 10^7$ trajectories. The price is then re-computed using 10^7 independent trajectories. We compute each step in the backward optimization as described in the previous section, using polynomials of degree three in the case of two stocks (ten parameters for each time in $t_0, t_1, \dots, t_J$). We take $J = 9$ and $T = 3$, with $t_i = i/3,\ i = 1,\dots,9$.

Then, we also implement the time-dependent stopping function in (38) and optimize it using the forward method on the same example, this time using polynomials of degree four. Results and relative benchmark are given in Table 5.3. We report results obtained in [6] using neural networks (NN) and the confidence intervals (CI) given in [2].

The experiments were also repeated with the alternative parametrization (33), with comparable numerical results.

A Proofs of auxiliary results

A.1 Proof of Lemma 4.3

Set

$$T = \sum_{j=0}^{J} Z_j h_{1,j}(X_j)\prod_{l=0}^{j-1}(1-h_{1,l}(X_l)) - \sum_{j=0}^{J} Z_j h_{2,j}(X_j)\prod_{l=0}^{j-1}(1-h_{2,l}(X_l)).$$

We have

$$T = \sum_{j=0}^{J} Z_j \left(h_{1,j}(X_j) - h_{2,j}(X_j)\right) \prod_{l=0}^{j-1}(1 - h_{2,l}(X_l))$$
$$+ \sum_{j=0}^{J} Z_j h_{1,j}(X_j) \left[\prod_{l=0}^{j-1}(1 - h_{1,l}(X_l)) - \prod_{l=0}^{j-1}(1 - h_{2,l}(X_l))\right].$$

Due to the simple identity

$$\prod_{k=1}^{K} a_k - \prod_{k=1}^{K} b_k = \sum_{k=1}^{K}(a_k - b_k) \prod_{l=1}^{k-1} a_l \prod_{r=k+1}^{K} b_r,$$

we derive

$$\prod_{l=0}^{j-1}(1 - h_{1,l}(X_l)) - \prod_{l=0}^{j-1}(1 - h_{2,l}(X_l)) =$$
$$\sum_{l=0}^{j-1} \left(h_{2,l}(X_l) - h_{1,l}(X_l)\right) \prod_{s=0}^{l-1}(1 - h_{2,s}(X_s)) \prod_{m=l+1}^{j-1} (1 - h_{1,m}(X_m)).$$

Hence

$$T = \sum_{j=0}^{J} Z_j \left(h_{1,j}(X_j) - h_{2,j}(X_j)\right) \prod_{l=0}^{j-1}(1 - h_{2,l}(X_l))$$
$$+ \sum_{j=0}^{J} Z_j h_{1,j}(X_j) \sum_{l=r}^{j-1} \left(h_{2,l}(X_l) - h_{1,l}(X_l)\right) \prod_{s=0}^{l-1}(1 - h_{1,s}(X_s)) \prod_{m=l+1}^{j-1} (1 - h_{2,m}(X_m))$$
$$= \sum_{j=0}^{J} Z_j \left(h_{1,j}(X_j) - h_{2,j}(X_j)\right) \prod_{l=0}^{j-1}(1 - h_{2,l}(X_l))$$
$$+ \sum_{l=0}^{J-1} \left(h_{2,l}(X_l) - h_{1,l}(X_l)\right) \prod_{s=0}^{l-1}(1 - h_{2,s}(X_s)) \left\{ \sum_{j=l+1}^{J} Z_j h_{1,j}(X_j) \prod_{m=l+1}^{j-1} (1 - h_{1,m}(X_m)) \right\}$$

and

$$|T| \leq C_Z \sum_{j=0}^{J-1} |h_{1,j}(X_j) - h_{2,j}(X_j)| \prod_{l=r}^{j-1}(1 - h_{2,l}(X_l)).$$

A.2 Proof of Proposition 4.4

Lemma A.1. *Let $(\tau_{1,j})$ and $(\tau_{2,j})$ be two consistent families of randomized stopping times, then*

$$\mathsf{E}_r\left[Z_{\tau_{1,r}} - Z_{\tau_{2,r}}\right] = \mathsf{E}_r\left[\sum_{l=r}^{J-1} \left\{Z_l - \mathsf{E}_l\left[Z_{\tau_{1,l+1}}\right]\right\} (q_{1,l} - q_{2,l}) \prod_{k=r}^{l-1} (1 - q_{2,k})\right]$$

with $q_{i,j} = \widetilde{\mathsf{P}}\left(\tau_{i,j} = j\right), i = 1, 2,$ and $\mathsf{E}_r[\cdot] = \mathsf{E}[\cdot|\mathcal{F}_r]$.

Proof. We have

$$\begin{aligned}\mathsf{E}_r\left[Z_{\tau_{1,r}}-Z_{\tau_{2,r}}\right] &= \left\{Z_r-\mathsf{E}_r\left[Z_{\tau_{1,r+1}}\right]\right\}\left(\widetilde{\mathsf{P}}\left(\tau_{1,r}=r\right)-\widetilde{\mathsf{P}}\left(\tau_{2,r}=r\right)\right)\\ &\quad+\mathsf{E}_r\left[\left(\mathsf{E}_{r+1}\left[Z_{\tau_{1,r+1}}-Z_{\tau_{2,r+1}}\right]\right)\widetilde{\mathsf{P}}\left(\tau_{2,r}>r\right)\right]\\ &=\left\{Z_r-\mathsf{E}_r\left[Z_{\tau_{1,r+1}}\right]\right\}\left(q_{1,r}-q_{2,r}\right)\\ &\quad+\mathsf{E}_r\left[\left(\mathsf{E}_{r+1}\left[Z_{\tau_{1,r+1}}-Z_{\tau_{2,r+1}}\right]\right)\left(1-q_{2,r}\right)\right].\end{aligned}$$

By denoting $\Delta_r=\mathsf{E}_r\left[Z_{\tau_{1,r}}-Z_{\tau_{2,r}}\right]$, we derive the following recurrent relation

$$\Delta_r=\left\{Z_r-\mathsf{E}_r\left[Z_{\tau_{1,r+1}}\right]\right\}\left(q_{1,r}-q_{2,r}\right)+\mathsf{E}_r\left[\Delta_{r+1}\right]\left(1-q_{2,r}\right),$$

where all quantities with index r are $\mathcal{F}_r-$measurable. □

Using the property (8) we derive an important corollary.

Corollary A.2. *It holds for any consistent family $(\tau_r)_{r=0}^{J}$ of randomized stopping times,*

$$\mathsf{E}\left[Z_{\tau_r^\star}-Z_{\tau_r}\right]=\mathsf{E}\left[\sum_{l=0}^{J-1}\left|Z_l-\mathsf{E}_l\left[Z_{\tau_{l+1}^\star}\right]\right|\left|q_l^\star-q_l\right|\prod_{k=r}^{l-1}\left(1-q_k\right)\right],$$

where $q_l=\widetilde{\mathsf{P}}\left(\tau_l=l\right)$ and $q_l^\star=1_{\{\tau_l^\star=l\}}$.

Denote

$$\mathcal{A}_l:=\left\{\left|G_l(X_l)-C_l^\star(X_l)\right|>\delta\right\},\quad l=0,\ldots,J-1,$$

then

$$\begin{aligned}\Delta(\boldsymbol{h})&\geq\mathsf{E}\left[\sum_{l=0}^{J-1}1_{\mathcal{A}_l}\left|Z_l-\mathsf{E}_l\left[Z_{\tau_{l+1}^\star}\right]\right|\left|q_l^\star-q_l\right|\prod_{k=0}^{l-1}\left(1-q_k\right)\right]\\&=\delta\left\{\mathsf{E}\left[\sum_{l=0}^{J-1}\left|h_l^\star(X_l)-h_l(X_l)\right|\prod_{k=0}^{l-1}\left(1-h_k(X_k)\right)\right]-\sum_{l=0}^{J-1}\mathsf{P}\left(\overline{\mathcal{A}_l}\right)\right\}.\end{aligned}$$

Due to (**B**)

$$\sum_{l=0}^{J-1}\mathsf{P}\left(\overline{\mathcal{A}_l}\right)\leq A_0\delta^\alpha,\quad A_0=\sum_{l=0}^{J-1}A_{0,l}$$

and hence

$$\Delta(\boldsymbol{h})\geq\delta\left\{\frac{d_X^2(\boldsymbol{h}^\star,\boldsymbol{h})}{J}-A_0\delta^\alpha\right\}\geq\delta\left\{\frac{\Delta_X^2(\boldsymbol{h}^\star,\boldsymbol{h})}{JC_Z^2}-A_0\delta^\alpha\right\}$$

due to Lemma 4.3. Taking maximum of the right hand side in δ, we get

$$\Delta_X(\boldsymbol{h}^\star,\boldsymbol{h})\leq c_\alpha\sqrt{J}\Delta^{\alpha/(2(1+\alpha))}(\boldsymbol{h})$$

for some constant c_α depending on α only.

A.3 Proof of Proposition 4.5

By (13) we may write,

$$
\begin{aligned}
&0 \leq Q_{k-1}(h^\star_{k-1}, h^\star_k, \ldots, h^\star_J) - Q_{k-1}(h_{k-1}, h_k, \ldots, h_J) \\
&= \mathsf{E}\left[Z_{k-1}h^\star_{k-1}(X_{k-1}) - Z_{k-1}h_{k-1}(X_{k-1})\right] \\
&+ \mathsf{E}\left[(1-h^\star_{k-1}(X_{k-1}))\mathsf{E}\left[\sum_{j=k}^{J} Z_j h^\star_j(X_j)\prod_{l=k}^{j-1}(1-h^\star_l(X_l))\middle|\mathcal{F}_{k-1}\right]\right] \\
&- \mathsf{E}\left[(1-h_{k-1}(X_{k-1}))\mathsf{E}\left[\sum_{j=k}^{J} Z_j h_j(X_j)\prod_{l=k}^{j-1}(1-h_l(X_l))\middle|\mathcal{F}_{k-1}\right]\right] \\
&= \mathsf{E}\left[\left(Z_{k-1}-C^\star_{k-1}\right)h^\star_{k-1}(X_{k-1}) + C^\star_{k-1} - Z_{k-1}h_{k-1}(X_{k-1})\right] - Q_k(h_k, \ldots, h_J) \\
&+ \mathsf{E}\left[h_{k-1}(X_{k-1})\mathsf{E}\left[\sum_{j=k}^{J} Z_j h_j(X_j)\prod_{l=k}^{j-1}(1-h_l(X_l))\middle|\mathcal{F}_{k-1}\right]\right] \\
&= Q_k(h^\star_k, h^\star_k, \ldots, h^\star_J) - Q_k(h_k, \ldots, h_J) \\
&+ \mathsf{E}\left[\left(Z_{k-1}-C^\star_{k-1}\right)\left(h^\star_{k-1}(X_{k-1}) - h_{k-1}(X_{k-1})\right)\right] \\
&+ \mathsf{E}\left[h_{k-1}(X_{k-1})\left(\mathsf{E}\left[\sum_{j=k}^{J} Z_j h_j(X_j)\prod_{l=k}^{j-1}(1-h_l(X_l))\middle|\mathcal{F}_{k-1}\right] - C^\star_{k-1}\right)\right] \\
&\leq Q_k(h^\star_k, h^\star_k, \ldots, h^\star_J) - Q_k(h_k, \ldots, h_J) \\
&+ \mathsf{E}\left[\left(Z_{k-1}-C^\star_{k-1}\right)\left(h^\star_{k-1}(X_{k-1}) - h_{k-1}(X_{k-1})\right)\right].
\end{aligned}
$$

B Some auxiliary results

Let $\mathcal{X} \subset \mathbb{R}^d$ and let π be a probability measure on $\mathcal{X}$. We denote by $C(\mathcal{X})$ a set of all continuous (possibly piecewise) functions on $\mathcal{X}$ and by $C^s(\mathcal{X})$ the set of all s-times continuously differentiable (possibly piecewise) functions on $\mathcal{X}$. For a real-valued function h on $\mathcal{X} \subset \mathbb{R}^d$ we write $\|h\|_{L^p(\pi)} = (\int_{\mathcal{X}} |h(x)|^p \pi(x)dx)^{1/p}$ with $1 \leq p < \infty$. The set of all functions h with $\|h\|_{L^p(\pi)} < \infty$ is denoted by $L^p(\pi)$. If λ is the Lebesgue measure, we write shortly L^p instead of $L^p(\lambda)$. The (real) Sobolev space is denoted by $W^{s,p}(\mathcal{X})$, i.e.,

$$
W^{s,p}(\mathcal{X}) := \{u \in L^p : D^\alpha u \in L^p, \quad \forall |\alpha| \leqslant s\}, \tag{39}
$$

where $\alpha = (\alpha_1, \ldots, \alpha_d)$ is a multi-index with $|\alpha| = \alpha_1 + \ldots + \alpha_d$ and D^α stands for differential operator of the form

$$
D^\alpha = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}}. \tag{40}
$$

Here all derivatives are understood in the weak sense. The Sobolev norm is defined as

$$
\|u\|_{W^{s,p}(\mathcal{X})} = \sum_{|\alpha| \leq r} \|D^\alpha u\|_{L^p}.
$$

Theorem B.1 (Theorem 2.1 in [18]). *Let $1 \leq r \leq d, p \geq 1, n \geq 1$ be integers, $\psi : \mathbb{R}^r \to \mathbb{R}$ be infinitely many times differentiable in some open sphere in $\mathbb{R}^r$ and moreover, there is $b \in \mathbb{R}^r$ in this* *sphere such that $D^\alpha\psi(b) \neq 0$ for all α. Then there are $r \times d$ real matrices $\{A_j\}_{j=1}^n$ with the following property. For any $f \in W^{s,p}(\mathcal{X})$ with $s \geq 1$ there are coefficients $a_j(f)$*

$$\left\| f - \sum_{i=1}^n a_i(f)\psi(A_i(\cdot) + b) \right\|_{L^p(\mathcal{X})} \leq \frac{c\|f\|_{W^{s,p}(\mathcal{X})}}{n^{s/d}},$$

where c is an absolute constant. Moreover, a_j are continuous linear functionals on $W^{s,p}(\mathcal{X})$ with $\sum_{j=1}^n |a_j| \leq C$ and C depending on $\|f\|_{L^p(\mathcal{X})}$.

For any set $A \subset \mathbb{R}^d$ let

$$A^\varepsilon = \{x \in \mathbb{R}^d : \rho_A(x) \leq \varepsilon\}, \quad \rho_A(x) = \inf_{y \in A} |x - y|.$$

Lemma B.2. *Let a set $A \subset \mathbb{R}^d$ be convex. Then for any $\varepsilon > 0$ there exists a infinitely differentiable function φ_A with $0 \leq \phi \leq 1$, such that*

$$\varphi_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \in \mathbb{R}^d \setminus A^\varepsilon \end{cases}$$

and for any multiindex $\alpha = (\alpha_1, \ldots, \alpha_d)$

$$|D^\alpha \varphi_A(x)| \leq \frac{C_\alpha}{\varepsilon^{|\alpha|}}, \quad x \in \mathbb{R}^d$$

with a constant $C_\alpha > 0$.

Corollary B.3. *Let $S \subseteq \mathbb{R}^d$ be convex and let function ψ satisfy the conditions of Theorem B.1. Then for any fixed $s > d$, there are $r \times d$ real matrices $\{A_j\}_{j=1}^n$ and $b \in \mathbb{R}^r$ with the following property*

$$\left\| 1_S(\cdot) - \sum_{i=1}^n a_i(S)\psi(A_i(\cdot) + b) \right\|_{L_2(\pi)} \leq \frac{C_s}{\varepsilon^s n^{s/d}} + \sqrt{\pi(S \setminus S^\varepsilon)}$$

for some constant $C_s > 0$ and some real numbers $a_1, \ldots, a_n$ depending on S such that $\sum_{i=1}^n |a_i| \leq Q$ where Q is an absolute constant.

Proof. Due to Lemma B.2, there is an infinitely smooth function ϕ_S such that

$$\sup_{x \in \mathbb{R}^d} |1_S(x) - \phi_S(x)| \leq \pi(S \setminus S^\varepsilon).$$

According to Theorem B.1, we have with $p = \infty$

$$\sup_{x \in \mathbb{R}^d} \left| \varphi_S(x) - \sum_{i=1}^n a_i(S)\psi(A_i x + b) \right| \leq \frac{C_s}{\varepsilon^s n^{s/d}}$$

for some matrices $r \times d$ real matrices $\{A_j\}_{j=1}^n$ and real numbers $a_1, \ldots, a_n$ depending on S. □

References

[1] Leif Andersen, *A Simple Approach to the Pricing of Bermudan Swaptions in the multi-factor LIBOR Market Model*, Journal of Computational Finance **3** (1999), 5–32.

[2] Leif Andersen and Mark Broadie, *Primal-dual simulation algorithm for pricing multidimensional American options*, Management Science 50 (2004).

[3] Vlad Bally, Gilles Pagès, and Jacques Printems, *A quantization tree method for pricing and hedging multidimensional American options*, Math. Finance **15** (2005), no. 1, 119–168.

[4] Christian Bayer, Raúl Tempone, and Sören Wolfers, *Pricing American options by exercise rate optimization*, arXiv preprint arXiv:1809.07300 (2018).

[5] Sebastian Becker, Patrick Cheridito, and Arnulf Jentzen, *Deep optimal stopping*, arXiv preprint arXiv:1804.05394 (2018).

[6] Sebastian Becker, Patrick Cheridito, Arnulf Jentzen, and Timo Welti, *Solving high-dimensional optimal stopping problems using deep learning*, 2019.

[7] Denis Belomestny, *On the rates of convergence of simulation-based optimization algorithms for optimal stopping problems*, The Annals of Applied Probability **21** (2011), no. 1, 215–239.

[8] ———, *Pricing Bermudan options by nonparametric regression: optimal rates of convergence for lower estimates*, Finance Stoch. **15** (2011), no. 4, 655–683. MR 2863638

[9] Denis Belomestny and Volker Krätschmer, *Optimal stopping under model uncertainty: randomized stopping times approach*, The Annals of Applied Probability **26** (2016), no. 2, 1260–1295.

[10] Denis Belomestny and John Schoenmakers, *Advanced simulation-based methods for optimal stopping and control*, Palgrave Macmillan, London, 2018, With applications in finance. MR 3752620

[11] Mark Broadie and Paul Glasserman, *Pricing American-style securities using simulation*, Journal of Economic Dynamics and Control **21** (1997), no. 8, 1323–1352.

[12] Evarist Giné and Armelle Guillou, *Rates of strong uniform consistency for multivariate kernel density estimators*, Annales de l'Institut Henri Poincare (B) Probability and Statistics, vol. 38, Elsevier, 2002, pp. 907–921.

[13] István Gyöngy and David Šiška, *On randomized stopping*, Bernoulli **14** (2008), no. 2, 352–361.

[14] László Györfi, Michael Kohler, Adam Krzyżak, and Harro Walk, *A distribution-free theory of nonparametric regression*, Springer Series in Statistics, Springer-Verlag, New York, 2002. MR 1920390 (2003g:62006)

[15] Martin Haugh and Leonid Kogan, *Pricing American options: A duality approach.*, Oper. Res. **52** (2004), no. 2, 258–270.

[16] Anastasia Kolodko and John Schoenmakers, *Iterative construction of the optimal Bermudan stopping time.*, Finance Stoch. **10** (2006), no. 1, 27–49.

[17] Francis A. Longstaff and Eduardo S. Schwartz, *Valuing American options by simulation: a simple least-squares approach*, Review of Financial Studies **14** (2001), no. 1, 113–147.

[18] Hrushikesh N. Mhaskar, *Neural Networks for Optimal Approximation of Smooth and Analytic Functions*, Neural Computation **8** (1996), no. 1, 164–177.

[19] Leonard C. G. Rogers, *Monte Carlo valuation of American options*, Mathematical Finance **12** (2002), no. 3, 271–286.

[20] John Schoenmakers, *Robust Libor modelling and pricing of derivative products*, Chapman & Hall/CRC Financial Mathematics Series, Chapman & Hall/CRC, Boca Raton, FL, 2005. MR 2138789 (2005m:91005)

[21] Michel Talagrand, *Sharper bounds for Gaussian and empirical processes*, Ann. Probab. **22** (1994), no. 1, 28–76. MR 1258865

[22] Sara A Van de Geer, *Applications of empirical process theory*, vol. 91, Cambridge University Press Cambridge, 2000.