Journal of Machine Learning Research 20 (2019) 1-25 Submitted 4/18; Revised 1/19; Published 4/19

Deep Optimal Stopping

Sebastian Becker SEBASTIAN.BECKER@ZENAI.CH
Zenai AG, 8045 Zurich, Switzerland

Patrick Cheridito PATRICK.CHERIDITO@MATH.ETHZ.CH
RiskLab, Department of Mathematics
ETH Zurich, 8092 Zurich, Switzerland

Arnulf Jentzen ARNULF.JENTZEN@SAM.MATH.ETHZ.CH
SAM, Department of Mathematics
ETH Zurich, 8092 Zurich, Switzerland

Editor: Jon McAuliffe

Abstract

In this paper we develop a deep learning method for optimal stopping problems which directly learns the optimal stopping rule from Monte Carlo samples. As such, it is broadly applicable in situations where the underlying randomness can efficiently be simulated. We test the approach on three problems: the pricing of a Bermudan max-call option, the pricing of a callable multi barrier reverse convertible and the problem of optimally stopping a fractional Brownian motion. In all three cases it produces very accurate results in high-dimensional situations with short computing times.

Keywords: optimal stopping, deep learning, Bermudan option, callable multi barrier reverse convertible, fractional Brownian motion

1. Introduction

We consider optimal stopping problems of the form $\sup_\tau \mathbb{E}\, g(\tau, X_\tau)$, where $X = (X_n)_{n=0}^N$ is an $\mathbb{R}^d$-valued discrete-time Markov process and the supremum is over all stopping times τ based on observations of X. Formally, this just covers situations where the stopping decision can only be made at finitely many times. But practically all relevant continuous-time stopping problems can be approximated with time-discretized versions. The Markov assumption means no loss of generality. We make it because it simplifies the presentation and many important problems already are in Markovian form. But every optimal stopping problem can be made Markov by including all relevant information from the past in the current state of X (albeit at the cost of increasing the dimension of the problem).

In theory, optimal stopping problems with finitely many stopping opportunities can be solved exactly. The optimal value is given by the smallest supermartingale that dominates the reward process – the so-called Snell envelope – and the smallest (largest) optimal stopping time is the first time the immediate reward dominates (exceeds) the continuation value; see, e.g., Peskir and Shiryaev (2006) or Lamberton and Lapeyre (2008). However, traditional numerical methods suffer from the curse of dimensionality. For instance, the complexity of standard tree- or lattice-based methods increases exponentially in the dimension. For typical problems they yield good results for up to three dimensions. To treat higher-dimensional problems, various Monte Carlo based have been developed over the last years. A common approach consists in estimating continuation values to either derive stopping rules or recursively approximate the Snell envelope; see, e.g., Tilley (1993), Barraquand and Martineau (1995), Carriere (1996), Longstaff and Schwartz (2001), Tsitsiklis and Van Roy (2001), Boyle et al. (2003), Broadie and Glasserman (2004), Bally et al. (2005), Kolodko and Schoenmakers (2006), Egloff et al. (2007), Berridge and Schumacher (2008), Jain and Oosterlee (2015), Belomestny et al. (2018) or Haugh and Kogan (2004) and Kohler et al. (2010), which use neural networks with one hidden layer to do this. A different strand of the literature has focused on approximating optimal exercise boundaries; see, e.g., Andersen (2000), García (2003) and Belomestny (2011). Based on an idea of Davis and Karatzas (1994), a dual approach was developed by Rogers (2002) and Haugh and Kogan (2004); see Jamshidian (2007) and Chen and Glasserman (2007) for a multiplicative version and Andersen and Broadie (2004), Broadie and Cao (2008), Belomestny et al. (2009), Rogers (2010), Desai et al. (2012), Belomestny (2013), Belomestny et al. (2013) and Lelong (2016) for extensions and primal-dual methods. In Sirignano and Spiliopoulos (2018) optimal stopping problems in continuous time are treated by approximating the solutions of the corresponding free boundary PDEs with deep neural networks.

In this paper we use deep learning to approximate an optimal stopping time. Our approach is related to policy optimization methods used in reinforcement learning (Sutton and Barto, 1998), deep reinforcement learning (Schulman et al., 2015; Mnih et al., 2015; Silver et al., 2016; Lillicrap et al., 2016) and the deep learning method for stochastic control problems proposed by Han and E (2016). However, optimal stopping differs from the typical control problems studied in this literature. The challenge of our approach lies in the implementation of a deep learning method that can efficiently learn optimal stopping times. We do this by decomposing an optimal stopping time into a sequence of 0-1 stopping decisions and approximating them recursively with a sequence of multilayer feedforward neural networks. We show that our neural network policies can approximate optimal stopping times to any degree of desired accuracy. A candidate optimal stopping time $\hat{\tau}$ can be obtained by running a stochastic gradient ascent. The corresponding expectation $\mathbb{E}\, g(\hat{\tau}, X_{\hat{\tau}})$ provides a lower bound for the optimal value $\sup_\tau \mathbb{E}\, g(\tau, X_\tau)$. Using a version of the dual method of Rogers (2002) and Haugh and Kogan (2004), we also derive an upper bound. In all our examples, both bounds can be computed with short run times and lie close together.

The rest of the paper is organized as follows: In Section 2 we introduce the setup and explain our method of approximating optimal stopping times with neural networks. In Section 3 we construct lower bounds, upper bounds, point estimates and confidence intervals for the optimal value. In Section 4 we test the approach on three examples: the pricing of a Bermudan max-call option on different underlying assets, the pricing of a callable multi barrier reverse convertible and the problem of optimally stopping a fractional Brownian motion. In the first two examples, we use a multi-dimensional Black–Scholes model to describe the dynamics of the underlying assets. Then the pricing of a Bermudan max-call option amounts to solving a d-dimensional optimal stopping problem, where d is the number of assets. We provide numerical results for $d = 2, 3, 5, 10, 20, 30, 50, 100, 200$ and 500. In the case of a callable MBRC, it becomes a $d+1$-dimensional stopping problem since one also needs to keep track of the barrier event. We present results for $d = 2, 3, 5, 10, 15$ and 30. In the third example we only consider a one-dimensional fractional Brownian motion. But fractional Brownian motion is not Markov. In fact, all of its increments are correlated. So, to optimally stop it, one has to keep track of all past movements. To make it tractable, we approximate the continuous-time problem with a time-discretized version, which if formulated as a Markovian problem, has as many dimensions as there are time-steps. We compute a solution for 100 time-steps.

2. Deep Learning Optimal Stopping Rules

Let $X = (X_n)_{n=0}^N$ be an $\mathbb{R}^d$-valued discrete-time Markov process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where N and d are positive integers. We denote by $\mathcal{F}_n$ the σ-algebra generated by $X_0, X_1, \dots, X_n$ and call a random variable $\tau \colon \Omega \to \{0, 1, \dots, N\}$ an X-stopping time if the event $\{\tau = n\}$ belongs to $\mathcal{F}_n$ for all $n \in \{0, 1, \dots, N\}$.

Our aim is to develop a deep learning method that can efficiently learn an optimal policy for stopping problems of the form

$$\sup_{\tau \in \mathcal{T}} \mathbb{E}\, g(\tau, X_\tau), \tag{1}$$

where $g \colon \{0, 1, \dots, N\} \times \mathbb{R}^d \to \mathbb{R}$ is a measurable function and $\mathcal{T}$ denotes the set of all X-stopping times. To make sure that problem (1) is well-defined and admits an optimal solution, we assume that g satisfies the integrability condition

$$\mathbb{E}\, |g(n, X_n)| < \infty \quad \text{for all } n \in \{0, 1, \dots, N\}\,; \tag{2}$$

see, e.g., Peskir and Shiryaev (2006) or Lamberton and Lapeyre (2008). To be able to derive confidence intervals for the optimal value (1), we will have to make the slightly stronger assumption

$$\mathbb{E}\big[g(n, X_n)^2\big] < \infty \quad \text{for all } n \in \{0, 1, \dots, N\} \tag{3}$$

in Subsection 3.3 below. This is satisfied in all our examples in Section 4.

2.1. Expressing Stopping Times in Terms of Stopping Decisions

Any X-stopping time can be decomposed into a sequence of 0-1 stopping decisions. In principle, the decision whether to stop the process at time n if it has not been stopped before, can be made based on the whole evolution of X from time 0 until n. But to optimally stop the Markov process X, it is enough to make stopping decisions according to $f_n(X_n)$ for measurable functions $f_n \colon \mathbb{R}^d \to \{0, 1\}$, $n = 0, 1, \dots, N$. Theorem 1 below extends this well-known fact and serves as the theoretical basis of our method.

Consider the auxiliary stopping problems

$$V_n = \sup_{\tau \in \mathcal{T}_n} \mathbb{E}\, g(\tau, X_\tau) \tag{4}$$

for $n = 0, 1, \dots, N$, where $\mathcal{T}_n$ is the set of all X-stopping times satisfying $n \leq \tau \leq N$. Obviously, $\mathcal{T}_N$ consists of the unique element $\tau_N \equiv N$, and one can write $\tau_N = N f_N(X_N)$ for the constant function $f_N \equiv 1$. Moreover, for given $n \in \{0, 1, \dots, N\}$ and a sequence of measurable functions $f_n, f_{n+1}, \dots, f_N \colon \mathbb{R}^d \to \{0, 1\}$ with $f_N \equiv 1$,

$$\tau_n = \sum_{m=n}^{N} m f_m(X_m) \prod_{j=n}^{m-1} (1 - f_j(X_j)) \tag{5}$$

defines[1] a stopping time in $\mathcal{T}_n$. The following result shows that, for our method of recursively computing an approximate solution to the optimal stopping problem (1), it will be sufficient to consider stopping times of the form (5).

Theorem 1 *For a given $n \in \{0, 1, \ldots, N-1\}$, let τ_{n+1} be a stopping time in $\mathcal{T}_{n+1}$ of the form*

$$\tau_{n+1} = \sum_{m=n+1}^{N} m f_m(X_m) \prod_{j=n+1}^{m-1} (1 - f_j(X_j)) \tag{6}$$

for measurable functions $f_{n+1}, \ldots, f_N \colon \mathbb{R}^d \to \{0,1\}$ with $f_N \equiv 1$. Then there exists a measurable function $f_n \colon \mathbb{R}^d \to \{0,1\}$ such that the stopping time $\tau_n \in \mathcal{T}_n$ given by (5) satisfies

$$\mathbb{E}\, g(\tau_n, X_{\tau_n}) \geq V_n - \left(V_{n+1} - \mathbb{E}\, g(\tau_{n+1}, X_{\tau_{n+1}})\right),$$

where V_n and V_{n+1} are the optimal values defined in (4).

Proof Denote $\varepsilon = V_{n+1} - \mathbb{E}\, g(\tau_{n+1}, X_{\tau_{n+1}})$, and consider a stopping time $\tau \in \mathcal{T}_n$. By the Doob–Dynkin lemma (see, e.g., Aliprantis and Border, 2006, Theorem 4.41), there exists a measurable function $h_n \colon \mathbb{R}^d \to \mathbb{R}$ such that $h_n(X_n)$ is a version of the conditional expectation $\mathbb{E}\left[g(\tau_{n+1}, X_{\tau_{n+1}}) \mid X_n\right]$. Moreover, due to the special form (6) of τ_{n+1},

$$g(\tau_{n+1}, X_{\tau_{n+1}}) = \sum_{m=n+1}^{N} g(m, X_m) 1_{\{\tau_{n+1} = m\}} = \sum_{m=n+1}^{N} g(m, X_m) 1_{\{f_m(X_m) \prod_{j=n+1}^{m-1}(1 - f_j(X_j)) = 1\}}$$

is a measurable function of $X_{n+1}, \ldots, X_N$. So it follows from the Markov property of X that $h_n(X_n)$ is also a version of the conditional expectation $\mathbb{E}\left[g(\tau_{n+1}, X_{\tau_{n+1}}) \mid \mathcal{F}_n\right]$. Since the events

$$D = \{g(n, X_n) \geq h_n(X_n)\} \quad \text{and} \quad E = \{\tau = n\}$$

are in $\mathcal{F}_n$, $\tau_n = n 1_D + \tau_{n+1} 1_{D^c}$ belongs to $\mathcal{T}_n$ and $\tilde{\tau} = \tau_{n+1} 1_E + \tau 1_{E^c}$ to $\mathcal{T}_{n+1}$. It follows from the definitions of V_{n+1} and ε that $\mathbb{E}\, g(\tau_{n+1}, X_{\tau_{n+1}}) = V_{n+1} - \varepsilon \geq \mathbb{E}\, g(\tilde{\tau}, X_{\tilde{\tau}}) - \varepsilon$. Hence,

$$\mathbb{E}\left[g(\tau_{n+1}, X_{\tau_{n+1}}) 1_{E^c}\right] \geq \mathbb{E}[g(\tilde{\tau}, X_{\tilde{\tau}}) 1_{E^c}] - \varepsilon = \mathbb{E}[g(\tau, X_\tau) 1_{E^c}] - \varepsilon,$$

from which one obtains

$$\begin{aligned}
\mathbb{E}\, g(\tau_n, X_{\tau_n}) &= \mathbb{E}\left[g(n, X_n) I_D + g(\tau_{n+1}, X_{\tau_{n+1}}) I_{D^c}\right] = \mathbb{E}[g(n, X_n) I_D + h_n(X_n) I_{D^c}] \\
&\geq \mathbb{E}[g(n, X_n) I_E + h_n(X_n) I_{E^c}] = \mathbb{E}\left[g(n, X_n) I_E + g(\tau_{n+1}, X_{\tau_{n+1}}) I_{E^c}\right] \\
&\geq \mathbb{E}[g(n, X_n) I_E + g(\tau, X_\tau) I_{E^c}] - \varepsilon = \mathbb{E}\, g(\tau, X_\tau) - \varepsilon.
\end{aligned}$$

Since $\tau \in \mathcal{T}_n$ was arbitrary, this shows that $\mathbb{E}\, g(\tau_n, X_{\tau_n}) \geq V_n - \varepsilon$. Moreover, one has $1_D = f_n(X_n)$ for the function $f_n \colon \mathbb{R}^d \to \{0,1\}$ given by

$$f_n(x) = \begin{cases} 1 & \text{if } g(n,x) \geq h_n(x) \\ 0 & \text{if } g(n,x) < h_n(x) \end{cases}.$$

1. In expressions of the form (5), we understand the empty product $\prod_{j=n}^{n-1}(1 - f_j(X_j))$ as 1.

Therefore,

$$\tau_n = nf_n(X_n) + \tau_{n+1}(1 - f_n(X_n)) = \sum_{m=n}^{N} mf_m(X_m) \prod_{j=n}^{m-1} (1 - f_j(X_j)),$$

which concludes the proof. ■

Remark 2 *Since for $f_N \equiv 1$, the stopping time $\tau_N = f_N(X_N)$ is optimal in $\mathcal{T}_N$, Theorem 1 inductively yields measurable functions $f_n \colon \mathbb{R}^d \to \{0,1\}$ such that for all $n \in \{0,1,\dots,N-1\}$, the stopping time τ_n given by* (5) *is optimal among $\mathcal{T}_n$. In particular,*

$$\tau = \sum_{n=1}^{N} nf_n(X_n) \prod_{j=0}^{n-1} (1 - f_j(X_j)) \tag{7}$$

is an optimal stopping time for problem (1).

Remark 3 *In many applications, the Markov process X starts from a deterministic initial value $x_0 \in \mathbb{R}^d$. Then the function f_0 enters the representation* (7) *only through the value $f_0(x_0) \in \{0,1\}$; that is, at time* 0, *only a constant and not a whole function has to be learned.*

2.2. Neural Network Approximation

Our numerical method for problem (1) consists in iteratively approximating optimal stopping decisions $f_n \colon \mathbb{R}^d \to \{0,1\}$, $n = 0,1,\dots,N-1$, by a neural network $f^\theta \colon \mathbb{R}^d \to \{0,1\}$ with parameter $\theta \in \mathbb{R}^q$. We do this by starting with the terminal stopping decision $f_N \equiv 1$ and proceeding by backward induction. More precisely, let $n \in \{0,1,\dots,N-1\}$, and assume parameter values $\theta_{n+1}, \theta_{n+2}, \dots, \theta_N \in \mathbb{R}^q$ have been found such that $f^{\theta_N} \equiv 1$ and the stopping time

$$\tau_{n+1} = \sum_{m=n+1}^{N} mf^{\theta_m}(X_m) \prod_{j=n+1}^{m-1} (1 - f^{\theta_j}(X_j))$$

produces an expected value $\mathbb{E}\, g(\tau_{n+1}, X_{\tau_{n+1}})$ close to the optimum V_{n+1}. Since f^θ takes values in $\{0,1\}$, it does not directly lend itself to a gradient-based optimization method. So, as an intermediate step, we introduce a feedforward neural network $F^\theta \colon \mathbb{R}^d \to (0,1)$ of the form

$$F^\theta = \psi \circ a_I^\theta \circ \varphi_{q_{I-1}} \circ a_{I-1}^\theta \circ \dots \circ \varphi_{q_1} \circ a_1^\theta,$$

where

- $I, q_1, q_2, \dots, q_{I-1}$ are positive integers specifying the depth of the network and the number of nodes in the hidden layers (if there are any),
- $a_1^\theta \colon \mathbb{R}^d \to \mathbb{R}^{q_1}, \dots, a_{I-1}^\theta \colon \mathbb{R}^{q_{I-2}} \to \mathbb{R}^{q_{I-1}}$ and $a_I^\theta \colon \mathbb{R}^{q_{I-1}} \to \mathbb{R}$ are affine functions,
- for $j \in \mathbb{N}$, $\varphi_j \colon \mathbb{R}^j \to \mathbb{R}^j$ is the component-wise ReLU activation function given by $\varphi_j(x_1, \dots, x_j) = (x_1^+, \dots, x_j^+)$

- $\psi \colon \mathbb{R} \to (0,1)$ is the standard logistic function $\psi(x) = e^x/(1+e^x) = 1/(1+e^{-x})$.

The components of the parameter $\theta \in \mathbb{R}^q$ of F^θ consist of the entries of the matrices $A_1 \in \mathbb{R}^{q_1 \times d}, \dots, A_{I-1} \in \mathbb{R}^{q_{I-1} \times q_{I-2}}, A_I \in \mathbb{R}^{1 \times q_{I-1}}$ and the vectors $b_1 \in \mathbb{R}^{q_1}, \dots, b_{I-1} \in \mathbb{R}^{q_{I-1}}, b_I \in \mathbb{R}$ given by the representation of the affine functions

$$a_i^\theta(x) = A_i x + b_i, \quad i = 1, \dots, I.$$

So the dimension of the parameter space is

$$q = \begin{cases} d+1 & \text{if } I = 1 \\ 1 + q_1 + \dots + q_{I-1} + d q_1 + \dots + q_{I-2} q_{I-1} + q_{I-1} & \text{if } I \geq 2, \end{cases}$$

and for given $x \in \mathbb{R}^d$, $F^\theta(x)$ is continuous as well as almost everywhere smooth in θ. Our aim is to determine $\theta_n \in \mathbb{R}^q$ so that

$$\mathbb{E}\Big[g(n, X_n) F^{\theta_n}(X_n) + g(\tau_{n+1}, X_{\tau_{n+1}})(1 - F^{\theta_n}(X_n))\Big]$$

is close to the supremum $\sup_{\theta \in \mathbb{R}^q} \mathbb{E}\big[g(n, X_n) F^\theta(X_n) + g(\tau_{n+1}, X_{\tau_{n+1}})(1 - F^\theta(X_n))\big]$. Once this has been achieved, we define the function $f^{\theta_n} \colon \mathbb{R}^d \to \{0,1\}$ by

$$f^{\theta_n} = 1_{[0,\infty)} \circ a_I^{\theta_n} \circ \varphi_{q_{I-1}} \circ a_{I-1}^{\theta_n} \circ \dots \circ \varphi_{q_1} \circ a_1^{\theta_n}, \tag{8}$$

where $1_{[0,\infty)} \colon \mathbb{R} \to \{0,1\}$ is the indicator function of $[0,\infty)$. The only difference between F^{θ_n} and f^{θ_n} is the final nonlinearity. While F^{θ_n} produces a stopping probability in $(0,1)$, the output of f^{θ_n} is a hard stopping decision given by 0 or 1, depending on whether F^{θ_n} takes a value below or above 1/2.

The following result shows that for any depth $I \geq 2$, a neural network of the form (8) is flexible enough to make almost optimal stopping decisions provided it has sufficiently many nodes.

Proposition 4 *Let $n \in \{0, 1, \dots, N-1\}$ and fix a stopping time $\tau_{n+1} \in \mathcal{T}_{n+1}$. Then, for every depth $I \geq 2$ and constant $\varepsilon > 0$, there exist positive integers $q_1, \dots, q_{I-1}$ such that*

$$\begin{aligned} &\sup_{\theta \in \mathbb{R}^q} \mathbb{E}\Big[g(n, X_n) f^\theta(X_n) + g(\tau_{n+1}, X_{\tau_{n+1}})(1 - f^\theta(X_n))\Big] \\ &\geq \sup_{f \in \mathcal{D}} \mathbb{E}\big[g(n, X_n) f(X_n) + g(\tau_{n+1}, X_{\tau_{n+1}})(1 - f(X_n))\big] - \varepsilon, \end{aligned}$$

where $\mathcal{D}$ is the set of all measurable functions $f \colon \mathbb{R}^d \to \{0,1\}$.

Proof Fix $\varepsilon > 0$. It follows from the integrability condition (2) that there exists a measurable function $\tilde{f} \colon \mathbb{R}^d \to \{0,1\}$ such that

$$\begin{aligned} &\mathbb{E}\Big[g(n, X_n) \tilde{f}(X_n) + g(\tau_{n+1}, X_{\tau_{n+1}})(1 - \tilde{f}(X_n))\Big] \\ &\geq \sup_{f \in \mathcal{D}} \mathbb{E}\big[g(n, X_n) f(X_n) + g(\tau_{n+1}, X_{\tau_{n+1}})(1 - f(X_n))\big] - \varepsilon/4. \end{aligned} \tag{9}$$

$\tilde{f}$ can be written as $\tilde{f} = 1_A$ for the Borel set $A = \{x \in \mathbb{R}^d : \tilde{f}(x) = 1\}$. Moreover, by (2),

$$B \mapsto \mathbb{E}[|g(n, X_n)| 1_B(X_n)] \quad \text{and} \quad B \mapsto \mathbb{E}\big[|g(\tau_{n+1}, X_{\tau_{n+1}})| 1_B(X_n)\big]$$

define finite Borel measures on $\mathbb{R}^d$. Since every finite Borel measure on $\mathbb{R}^d$ is tight (see, e.g., Aliprantis and Border, 2006), there exists a compact (possibly empty) subset $K \subseteq A$ such that

$$\begin{aligned} &\mathbb{E}\big[g(n, X_n) 1_K(X_n) + g(\tau_{n+1}, X_{\tau_{n+1}})(1 - 1_K(X_n))\big] \\ &\geq \mathbb{E}\Big[g(n, X_n) \tilde{f}(X_n) + g(\tau_{n+1}, X_{\tau_{n+1}})(1 - \tilde{f}(X_n))\Big] - \varepsilon/4. \end{aligned} \tag{10}$$

Let $\rho_K \colon \mathbb{R}^d \to [0, \infty]$ be the distance function given by $\rho_K(x) = \inf_{y \in K} \|x - y\|_2$. Then

$$k_j(x) = \max\{1 - j\rho_K(x), -1\}, \quad j \in \mathbb{N},$$

defines a sequence of continuous functions $k_j \colon \mathbb{R}^d \to [-1, 1]$ that converge pointwise to $1_K - 1_{K^c}$. So it follows from Lebesgue's dominated convergence theorem that there exists a $j \in \mathbb{N}$ such that

$$\begin{aligned} &\mathbb{E}\Big[g(n, X_n)\, 1_{\{k_j(X_n) \geq 0\}} + g(\tau_{n+1}, X_{\tau_{n+1}})(1 - 1_{\{k_j(X_n) \geq 0\}})\Big] \\ &\geq \mathbb{E}\big[g(n, X_n) 1_K(X_n) + g(\tau_{n+1}, X_{\tau_{n+1}})(1 - 1_K(X_n))\big] - \varepsilon/4. \end{aligned} \tag{11}$$

By Theorem 1 of Leshno et al. (1993), k_j can be approximated uniformly on compacts by functions of the form

$$\sum_{i=1}^{r} (v_i^T x + c_i)^+ - \sum_{i=1}^{s} (w_i^T x + d_i)^+ \tag{12}$$

for $r, s \in \mathbb{N}$, $v_1, \dots, v_r, w_1, \dots, w_s \in \mathbb{R}^d$ and $c_1, \dots, c_r, d_1, \dots, d_s \in \mathbb{R}$. So there exists a function $h \colon \mathbb{R}^d \to \mathbb{R}$ expressible as in (12) such that

$$\begin{aligned} &\mathbb{E}\big[g(n, X_n)\, 1_{\{h(X_n) \geq 0\}} + g(\tau_{n+1}, X_{\tau_{n+1}})(1 - 1_{\{h(X_n) \geq 0\}})\big] \\ &\geq \mathbb{E}\Big[g(n, X_n)\, 1_{\{k_j(X_n) \geq 0\}} + g(\tau_{n+1}, X_{\tau_{n+1}})(1 - 1_{\{k_j(X_n) \geq 0\}})\Big] - \varepsilon/4. \end{aligned} \tag{13}$$

Now note that for any integer $I \geq 2$, the composite mapping $1_{[0,\infty)} \circ h$ can be written as a neural net f^θ of the form (8) with depth I for suitable integers $q_1, \dots, q_{I-1}$ and parameter value $\theta \in \mathbb{R}^q$. Hence, one obtains from (9), (10), (11) and (13) that

$$\begin{aligned} &\mathbb{E}\Big[g(n, X_n)\, f^\theta(X_n) + g(\tau_{n+1}, X_{\tau_{n+1}})(1 - f^\theta(X_n))\Big] \\ &\geq \sup_{f \in \mathcal{D}} \mathbb{E}\big[g(n, X_n) f(X_n) + g(\tau_{n+1}, X_{\tau_{n+1}})(1 - f(X_n))\big] - \varepsilon, \end{aligned}$$

and the proof is complete. ■

We always choose $\theta_N \in \mathbb{R}^q$ such that[2] $f^{\theta_N} \equiv 1$. Then our candidate optimal stopping time

$$\tau^\Theta = \sum_{n=1}^{N} n f^{\theta_n}(X_n) \prod_{j=0}^{n-1} (1 - f^{\theta_j}(X_j)) \tag{14}$$

2. It is easy to see that this is possible.

is specified by the vector $\Theta = (\theta_0, \theta_1, \dots, \theta_{N-1}) \in \mathbb{R}^{Nq}$. The following is an immediate consequence of Theorem 1 and Proposition 4:

Corollary 5 *For a given optimal stopping problem of the form* (1)*, a depth $I \geq 2$ and a constant $\varepsilon > 0$, there exist positive integers $q_1, \dots, q_{I-1}$ and a vector $\Theta \in \mathbb{R}^{Nq}$ such that the corresponding stopping time* (14) *satisfies $\mathbb{E}\, g(\tau^\Theta, X_{\tau^\Theta}) \geq \sup_{\tau \in \mathcal{T}} \mathbb{E}\, g(\tau, X_\tau) - \varepsilon$.*

2.3. Parameter Optimization

We train neural networks of the form (8) with fixed depth $I \geq 2$ and given numbers $q_1, \dots, q_{I-1}$ of nodes in the hidden layers[3]. To numerically find parameters $\theta_n \in \mathbb{R}^q$ yielding good stopping decisions f^{θ_n} for all times $n \in \{0, 1, \dots, N-1\}$, we approximate expected values with averages of Monte Carlo samples calculated from simulated paths of the process $(X_n)_{n=0}^N$.

Let $(x_n^k)_{n=0}^N$, $k = 1, 2, \dots$ be independent realizations of such paths. We choose $\theta_N \in \mathbb{R}^q$ such that $f^{\theta_N} \equiv 1$ and determine determine $\theta_n \in \mathbb{R}^q$ for $n \leq N-1$ recursively. So, suppose that for a given $n \in \{0, 1, \dots, N-1\}$, parameters $\theta_{n+1}, \dots, \theta_N \in \mathbb{R}^q$, have been found so that the stopping decisions $f^{\theta_{n+1}}, \dots, f^{\theta_N}$ generate a stopping time

$$\tau_{n+1} = \sum_{m=n+1}^{N} m f^{\theta_m}(X_m) \prod_{j=n+1}^{m-1} (1 - f^{\theta_j}(X_j))$$

with corresponding expectation $\mathbb{E}\, g(\tau_{n+1}, X_{\tau_{n+1}})$ close to the optimal value V_{n+1}. If $n = N-1$, one has $\tau_{n+1} \equiv N$, and if $n \leq N-2$, τ_{n+1} can be written as

$$\tau_{n+1} = l_{n+1}(X_{n+1}, \dots, X_{N-1})$$

for a measurable function $l_{n+1} \colon \mathbb{R}^{d(N-n-1)} \to \{n+1, n+2, \dots, N\}$. Accordingly, denote

$$l_{n+1}^k = \begin{cases} N & \text{if } n = N-1 \\ l_{n+1}(x_{n+1}^k, \dots, x_{N-1}^k) & \text{if } n \leq N-2 \end{cases}.$$

If at time n, one applies the soft stopping decision F^θ and afterward behaves according to $f^{\theta_{n+1}}, \dots, f^{\theta_N}$, the realized reward along the k-th simulated path of X is

$$r_n^k(\theta) = g(n, x_n^k) F^\theta(x_n^k) + g(l_{n+1}^k, x_{l_{n+1}^k}^k)(1 - F^\theta(x_n^k)).$$

For large $K \in \mathbb{N}$,

$$\frac{1}{K} \sum_{k=1}^{K} r_n^k(\theta) \tag{15}$$

approximates the expected value

$$\mathbb{E}\Big[g(n, X_n) F^\theta(X_n) + g(\tau_{n+1}, X_{\tau_{n+1}})(1 - F^\theta(X_n)) \Big].$$

3. For a given application, one can try out different choices of I and $q_1, \dots, q_{I-1}$ to find a suitable trade-off between accuracy and efficiency. Alternatively, the determination of I and $q_1, \dots, q_{I-1}$ could be built into the training algorithm.

Since $r_n^k(\theta)$ is almost everywhere differentiable in θ, a stochastic gradient ascent method can be applied to find an approximate optimizer $\theta_n \in \mathbb{R}^q$ of (15). The same simulations $(x_n^k)_{n=0}^N$, $k = 1, 2, \dots$ can be used to train the stopping decisions f^{θ_n} at all times $n \in \{0, 1, \dots, N-1\}$. In the numerical examples in Section 4 below, we employed mini-batch gradient ascent with Xavier initialization (Glorot and Bengio, 2010), batch normalization (Ioffe and Szegedy, 2015) and Adam updating (Kingma and Ba, 2015).

Remark 6 *If the Markov process X starts from a deterministic initial value $x_0 \in \mathbb{R}^d$, the initial stopping decision is given by a constant $f_0 \in \{0,1\}$. To learn f_0 from simulated paths of X, it is enough to compare the initial reward $g(0, x_0)$ to a Monte Carlo estimate $\hat{C}$ of $\mathbb{E}\, g(\tau_1, X_{\tau_1})$, where $\tau_1 \in \mathcal{T}_1$ is of the form*

$$\tau_1 = \sum_{n=1}^{N} n f^{\theta_n}(X_n) \prod_{j=1}^{n-1} (1 - f^{\theta_j}(X_j))$$

for $f^{\theta_N} \equiv 1$ and trained parameters $\theta_1, \dots, \theta_{N-1} \in \mathbb{R}^q$. Then one sets $f_0 = 1$ (that is, stop immediately) if $g(0, x_0) \geq \hat{C}$ and $f_0 = 0$ (continue) otherwise. The resulting stopping time is of the form

$$\tau^\Theta = \begin{cases} 0 & \text{if } f_0 = 1 \\ \tau_1 & \text{if } f_0 = 0. \end{cases}$$

3. Bounds, Point Estimates and Confidence Intervals

In this section we derive lower and upper bounds as well as point estimates and confidence intervals for the optimal value $V_0 = \sup_{\tau \in \mathcal{T}} \mathbb{E}\, g(\tau, X_\tau)$.

3.1. Lower Bound

Once the stopping decisions f^{θ_n} have been trained, the stopping time τ^Θ given by (14) yields a lower bound $L = \mathbb{E}\, g(\tau^\Theta, X_{\tau^\Theta})$ for the optimal value $V_0 = \sup_{\tau \in \mathcal{T}} \mathbb{E}\, g(\tau, X_\tau)$. To estimate it, we simulate a new set[4] of independent realizations $(y_n^k)_{n=0}^N$, $k = 1, 2, \dots, K_L$, of $(X_n)_{n=0}^N$. τ^Θ is of the form $\tau^\Theta = l(X_0, \dots, X_{N-1})$ for a measurable function $l \colon \mathbb{R}^{dN} \to \{0, 1, \dots, N\}$. Denote $l^k = l(y_0^k, \dots, y_{N-1}^k)$. The Monte Carlo approximation

$$\hat{L} = \frac{1}{K_L} \sum_{k=1}^{K_L} g(l^k, y_{l^k}^k)$$

gives an unbiased estimate of the lower bound L, and by the law of large numbers, $\hat{L}$ converges to L for $K_L \to \infty$.

4. In particular, we assume that the samples $(y_n^k)_{n=0}^N$, $k = 1, \dots, K_L$, are drawn independently from the realizations $(x_n^k)_{n=0}^N$, $k = 1, \dots, K$, used in the training of the stopping decisions.

3.2. Upper Bound

The Snell envelope of the reward process $(g(n, X_n))_{n=0}^N$ is the smallest[5] supermartingale with respect to $(\mathcal{F}_n)_{n=0}^N$ that dominates $(g(n, X_n))_{n=0}^N$. It is given[6] by

$$H_n = \operatorname{ess\,sup}_{\tau \in \mathcal{T}_n} \mathbb{E}[g(\tau) \mid \mathcal{F}_n], \quad n = 0, 1, \dots, N;$$

see, e.g., Peskir and Shiryaev (2006) or Lamberton and Lapeyre (2008). Its Doob–Meyer decomposition is

$$H_n = H_0 + M_n^H - A_n^H,$$

where M^H is the $(\mathcal{F}_n)$-martingale given[6] by

$$M_0^H = 0 \quad \text{and} \quad M_n^H - M_{n-1}^H = H_n - \mathbb{E}[H_n \mid \mathcal{F}_{n-1}], \quad n = 1, \dots, N,$$

and A^H is the nondecreasing $(\mathcal{F}_n)$-predictable process given[6] by

$$A_0^H = 0 \quad \text{and} \quad A_n^H - A_{n-1}^H = H_{n-1} - \mathbb{E}[H_n \mid \mathcal{F}_{n-1}], \quad n = 1, \dots, N.$$

Our estimate of an upper bound for the optimal value V_0 is based on the following variant[7] of the dual formulation of optimal stopping problems introduced by Rogers (2002) and Haugh and Kogan (2004).

Proposition 7 *Let $(\varepsilon_n)_{n=0}^N$ be a sequence of integrable random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. Then*

$$V_0 \ge \mathbb{E}\left[\max_{0 \le n \le N} \left(g(n, X_n) - M_n^H - \varepsilon_n\right)\right] + \mathbb{E}\left[\min_{0 \le n \le N} \left(A_n^H + \varepsilon_n\right)\right]. \tag{16}$$

Moreover, if $\mathbb{E}[\varepsilon_n \mid \mathcal{F}_n] = 0$ for all $n \in \{0, 1, \dots, N\}$, one has

$$V_0 \le \mathbb{E}\left[\max_{0 \le n \le N} \left(g(n, X_n) - M_n - \varepsilon_n\right)\right] \tag{17}$$

for every $(\mathcal{F}_n)$-martingale $(M_n)_{n=0}^N$ starting from 0.

Proof First, note that

$$\begin{aligned}
&\mathbb{E}\left[\max_{0 \le n \le N} \left(g(n, X_n) - M_n^H - \varepsilon_n\right)\right] \le \mathbb{E}\left[\max_{0 \le n \le N} \left(H_n - M_n^H - \varepsilon_n\right)\right] \\
&= \mathbb{E}\left[\max_{0 \le n \le N} \left(H_0 - A_n^H - \varepsilon_n\right)\right] = V_0 - \mathbb{E}\left[\min_{0 \le n \le N} \left(A_n^H + \varepsilon_n\right)\right],
\end{aligned}$$

which shows (16).

Now, assume that $\mathbb{E}[\varepsilon_n \mid \mathcal{F}_n] = 0$ for all $n \in \{0, 1, \dots, N\}$, and let τ be an X-stopping time. Then

$$\mathbb{E}\, \varepsilon_\tau = \mathbb{E}\left[\sum_{n=0}^N 1_{\{\tau = n\}} \varepsilon_n\right] = \mathbb{E}\left[\sum_{n=0}^N 1_{\{\tau = n\}} \mathbb{E}[\varepsilon_n \mid \mathcal{F}_n]\right] = 0.$$

5. in the $\mathbb{P}$-almost sure order
6. up to $\mathbb{P}$-almost sure equality
7. See also the discussion on noisy estimates in Andersen and Broadie (2004).

So one obtains from the optional stopping theorem (see, e.g., Grimmett and Stirzaker, 2001),

$$\mathbb{E}\, g(\tau, X_\tau) = \mathbb{E}[g(\tau, X_\tau) - M_\tau - \varepsilon_\tau] \le \mathbb{E}\left[\max_{0 \le n \le N} (g(n, X_n) - M_n - \varepsilon_n)\right]$$

for every $(\mathcal{F}_n)$-martingale $(M_n)_{n=0}^N$ starting from 0. Since $V_0 = \sup_{\tau \in \mathcal{T}} \mathbb{E}\, g(\tau, X_\tau)$, this implies (17). ■

For every $(\mathcal{F}_n)$-martingale $(M_n)_{n=0}^N$ starting from 0 and each sequence of integrable error terms $(\varepsilon_n)_{n=0}^N$ satisfying $\mathbb{E}[\varepsilon_n \mid \mathcal{F}_n] = 0$ for all n, the right side of (17) provides an upper bound[8] for V_0, and by (16), this upper bound is tight if $M = M^H$ and $\varepsilon \equiv 0$. So we try to use our candidate optimal stopping time τ^Θ to construct a martingale close to M^H. The closer τ^Θ is to an optimal stopping time, the better the value process[9]

$$H_n^\Theta = \mathbb{E}\left[g(\tau_n^\Theta, X_{\tau_n^\Theta}) \mid \mathcal{F}_n\right], \quad n = 0, 1, \ldots, N,$$

corresponding to

$$\tau_n^\Theta = \sum_{m=n}^N m f^{\theta_m}(X_m) \prod_{j=n}^{m-1} (1 - f^{\theta_j}(X_j)), \quad n = 0, 1, \ldots, N,$$

approximates the Snell envelope $(H_n)_{n=0}^N$. The martingale part of $(H_n^\Theta)_{n=0}^N$ is given by $M_0^\Theta = 0$ and

$$M_n^\Theta - M_{n-1}^\Theta = H_n^\Theta - \mathbb{E}\big[H_n^\Theta \mid \mathcal{F}_{n-1}\big] = f^{\theta_n}(X_n) g(n, X_n) + (1 - f^{\theta_n}(X_n)) C_n^\Theta - C_{n-1}^\Theta, \; n \ge 1, \tag{18}$$

for the continuation values[10]

$$C_n^\Theta = \mathbb{E}[g(\tau_{n+1}^\Theta, X_{\tau_{n+1}^\Theta}) \mid \mathcal{F}_n] = \mathbb{E}[g(\tau_{n+1}^\Theta, X_{\tau_{n+1}^\Theta}) \mid X_n], \quad n = 0, 1, \ldots, N-1.$$

Note that C_N^Θ does not have to be specified. It formally appears in (18) for $n = N$. But $(1 - f^{\theta_N}(X_N))$ is always 0. To estimate M^Θ, we generate a third set[11] of independent realizations $(z_n^k)_{n=0}^N$, $k = 1, 2, \ldots, K_U$, of $(X_n)_{n=0}^N$. In addition, for every z_n^k, we simulate J continuation paths $\tilde{z}_{n+1}^{k,j}, \ldots, \tilde{z}_N^{k,j}$, $j = 1, \ldots, J$, that are conditionally independent[12] of each

8. Note that for the right side of (17) to be a valid upper bound, it is sufficient that $\mathbb{E}[\varepsilon_n \mid \mathcal{F}_n] = 0$ for all n. In particular, $\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_N$ can have any arbitrary dependence structure.
9. Again, since H_n^Θ, M_n^Θ and C_n^Θ are given by conditional expectations, they are only specified up to $\mathbb{P}$-almost sure equality.
10. The two conditional expectations are equal since $(X_n)_{n=0}^N$ is Markov and τ_{n+1}^Θ only depends on $(X_{n+1}, \ldots, X_{N-1})$.
11. The realizations $(z_n^k)_{n=0}^N$, $k = 1, \ldots, K_U$, must be drawn independently of $(x_n^k)_{n=0}^N$, $k = 1, \ldots, K$, so that our estimate of the upper bound does not depend on the samples used to train the stopping decisions. But theoretically, they can depend on $(y_n^k)_{n=0}^N$, $k = 1, \ldots, K_L$, without affecting the unbiasedness of the estimate $\hat{U}$ or the validity of the confidence interval derived in Subsection 3.3 below.
12. More precisely, the tuples $(\tilde{z}_{n+1}^{k,j}, \ldots, \tilde{z}_N^{k,j})$, $j = 1, \ldots, J$, are simulated according to $p_n(z_n^k, \cdot)$, where p_n is a transition kernel from $\mathbb{R}^d$ to $\mathbb{R}^{(N-n)d}$ such that $p_n(X_n, B) = \mathbb{P}[(X_{n+1}, \ldots, X_N) \in B \mid X_n]$ $\mathbb{P}$-almost surely for all Borel sets $B \subseteq \mathbb{R}^{(N-n)d}$. We generate them independently of each other across j and k. On the other hand, the continuation paths starting from z_n^k do not have to be drawn independently of those starting from $z_{n'}^k$ for $n \neq n'$.

other and of $z^k_{n+1}, \dots, z^k_N$. Let us denote by $\tau^{k,j}_{n+1}$ the value of τ^Θ_{n+1} along $\tilde{z}^{k,j}_{n+1}, \dots, \tilde{z}^{k,j}_N$. Estimating the continuation values as

$$C^k_n = \frac{1}{J} \sum_{j=1}^J g\left(\tau^{k,j}_{n+1}, \tilde{z}^{k,j}_{\tau^{k,j}_{n+1}}\right), \quad n = 0, 1, \dots, N-1,$$

yields the noisy estimates

$$\Delta M^k_n = f^{\theta_n}(z^k_n) g(n, z^k_n) + (1 - f^{\theta_n}(z^k_n)) C^k_n - C^k_{n-1}$$

of the increments $M^\Theta_n - M^\Theta_{n-1}$ along the k-th simulated path $z^k_0, \dots, z^k_N$. So

$$M^k_n = \begin{cases} 0 & \text{if } n = 0 \\ \sum_{m=1}^n \Delta M^k_m & \text{if } n \geq 1 \end{cases}$$

can be viewed as realizations of $M^\Theta_n + \varepsilon_n$ for estimation errors ε_n with standard deviations proportional to $1/\sqrt{J}$ such that $\mathbb{E}[\varepsilon_n \mid \mathcal{F}_n] = 0$ for all n. Accordingly,

$$\hat{U} = \frac{1}{K_U} \sum_{k=1}^{K_U} \max_{0 \leq n \leq N} \left(g\left(n, z^k_n\right) - M^k_n\right),$$

is an unbiased estimate of the upper bound

$$U = \mathbb{E}\left[\max_{0 \leq n \leq N} \left(g(n, X_n) - M^\Theta_n - \varepsilon_n\right)\right],$$

which, by the law of large numbers, converges to U for $K_U \to \infty$.

3.3. Point Estimate and Confidence Intervals

Our point estimate of V_0 is the average

$$\frac{\hat{L} + \hat{U}}{2}.$$

To derive confidence intervals, we assume that $g(n, X_n)$ is square-integrable[13] for all n. Then

$$g(\tau^\theta, X_{\tau^\Theta}) \quad \text{and} \quad \max_{0 \leq n \leq N} \left(g(n, X_n) - M^\Theta_n - \varepsilon_n\right)$$

are square-integrable too. Hence, one obtains from the central limit theorem that for large K_L, $\hat{L}$ is approximately normally distributed with mean L and variance $\hat{\sigma}^2_L / K_L$ for

$$\hat{\sigma}^2_L = \frac{1}{K_L - 1} \sum_{k=1}^{K_L} \left(g(l^k, y^k_{l^k}) - \hat{L}\right)^2.$$

13. See condition (3).

So, for every $\alpha \in (0,1]$,

$$\left[\hat{L} - z_{\alpha/2} \frac{\hat{\sigma}_L}{\sqrt{K_L}} \, , \infty\right)$$

is an asymptotically valid $1-\alpha/2$ confidence interval for L, where $z_{\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution. Similarly,

$$\left(-\infty \, , \hat{U} + z_{\alpha/2} \frac{\hat{\sigma}_U}{\sqrt{K_U}}\right] \quad \text{with} \quad \hat{\sigma}_U^2 = \frac{1}{K_U - 1} \sum_{k=1}^{K_U} \left(\max_{0 \le n \le N} \left(g\left(n, z_n^k\right) - M_n^k \right) - \hat{U} \right)^2,$$

is an asymptotically valid $1-\alpha/2$ confidence interval for U. It follows that for every constant $\varepsilon > 0$, one has

$$\begin{aligned} &\mathbb{P}\left[V_0 < \hat{L} - z_{\alpha/2} \frac{\hat{\sigma}_L}{\sqrt{K_L}} \quad \text{or} \quad V_0 > \hat{U} + z_{\alpha/2} \frac{\hat{\sigma}_U}{\sqrt{K_U}}\right] \\ &\le \mathbb{P}\left[L < \hat{L} - z_{\alpha/2} \frac{\hat{\sigma}_L}{\sqrt{K_L}}\right] + \mathbb{P}\left[U > \hat{U} + z_{\alpha/2} \frac{\hat{\sigma}_U}{\sqrt{K_U}}\right] \le \alpha + \varepsilon \end{aligned}$$

as soon as K_L and K_U are large enough. In particular,

$$\left[\hat{L} - z_{\alpha/2} \frac{\hat{\sigma}_L}{\sqrt{K_L}} \, , \hat{U} + z_{\alpha/2} \frac{\hat{\sigma}_U}{\sqrt{K_U}}\right] \tag{19}$$

is an asymptotically valid $1-\alpha$ confidence interval for V_0.

4. Examples

In this section we test[14] our method on three examples: the pricing of a Bermudan max-call option, the pricing of a callable multi barrier reverse convertible and the problem of optimally stopping a fractional Brownian motion.

4.1. Bermudan Max-Call Options

Bermudan max-call options are one of the most studied examples in the numerics literature on optimal stopping problems (see, e.g., Longstaff and Schwartz, 2001; Rogers, 2002; García, 2003; Boyle et al., 2003; Haugh and Kogan, 2004; Broadie and Glasserman, 2004; Andersen and Broadie, 2004; Broadie and Cao, 2008; Berridge and Schumacher, 2008; Belomestny, 2011, 2013; Jain and Oosterlee, 2015; Lelong, 2016). Their payoff depends on the maximum of d underlying assets.

Assume the risk-neutral dynamics of the assets are given by a multi-dimensional Black–Scholes model[15]

$$S_t^i = s_0^i \exp\left([r - \delta_i - \sigma_i^2/2]t + \sigma_i W_t^i\right), \quad i = 1, 2, \ldots, d, \tag{20}$$

14. All computations were performed in single precision (float32) on a NVIDIA GeForce GTX 1080 GPU with 1974 MHz core clock and 8 GB GDDR5X memory with 1809.5 MHz clock rate. The underlying system consisted of an Intel Core i7-6800K 3.4 GHz CPU with 64 GB DDR4-2133 memory running Tensorflow 1.11 on Ubuntu 16.04.
15. We make this assumption so that we can compare our results to those obtained with different methods in the literature. But our approach works for any asset dynamics as long as it can efficiently be simulated.

for initial values $s_0^i \in (0,\infty)$, a risk-free interest rate $r \in \mathbb{R}$, dividend yields $\delta_i \in [0,\infty)$, volatilities $\sigma_i \in (0,\infty)$ and a d-dimensional Brownian motion W with constant instantaneous correlations[16] $\rho_{ij} \in \mathbb{R}$ between different components W^i and W^j. A Bermudan max-call option on $S^1, S^2, \dots, S^d$ has payoff $\left(\max_{1\le i\le d} S_t^i - K\right)^+$ and can be exercised at any point of a time grid $0 = t_0 < t_1 < \dots < t_N$. Its price is given by

$$\sup_\tau \mathbb{E}\left[e^{-r\tau}\left(\max_{1\le i\le d} S_\tau^i - K\right)^+\right],$$

where the supremum is over all S-stopping times taking values in $\{t_0, t_1, \dots, t_N\}$ (see, e.g., Schweizer, 2002). Denote $X_n^i = S_{t_n}^i$, $n = 0,1,\dots,N$, and let $\mathcal{T}$ be the set of X-stopping times. Then the price can be written as $\sup_{\tau\in\mathcal{T}} \mathbb{E}\, g(\tau, X_\tau)$ for

$$g(n,x) = e^{-rt_n}\left(\max_{1\le i\le d} x^i - K\right)^+,$$

and it is straight-forward to simulate $(X_n)_{n=0}^N$.

In the following we assume the time grid to be of the form $t_n = nT/N$, $n = 0,1,\dots,N$, for a maturity $T > 0$ and $N+1$ equidistant exercise dates. Even though $g(n, X_n)$ does not carry any information that is not already contained in X_n, our method worked more efficiently when we trained the optimal stopping decisions on Monte Carlo simulations of the $d+1$-dimensional Markov process $(Y_n)_{n=0}^N = (X_n, g(n,X_n))_{n=0}^N$ instead of $(X_n)_{n=0}^N$. Since Y_0 is deterministic, we first trained stopping times $\tau_1 \in \mathcal{T}_1$ of the form

$$\tau_1 = \sum_{n=1}^N n f^{\theta_n}(Y_n) \prod_{j=1}^{n-1}(1 - f^{\theta_j}(Y_k))$$

for $f^{\theta_N} \equiv 1$ and $f^{\theta_1}, \dots, f^{\theta_{N-1}} \colon \mathbb{R}^{d+1} \to \{0,1\}$ given by (8) with $I = 2$ and $q_1 = q_2 = d+40$. Then we determined our candidate optimal stopping times as

$$\tau^\Theta = \begin{cases} 0 & \text{if } f_0 = 1 \\ \tau_1 & \text{if } f_0 = 0 \end{cases}$$

for a constant $f_0 \in \{0,1\}$ depending[17] on whether it was optimal to stop immediately at time 0 or not (see Remark 6 above).

It is straight-forward to simulate from model (20). We conducted 3,000+d training steps, in each of which we generated a batch of 8,192 paths of $(X_n)_{n=0}^N$. To estimate the lower bound L we simulated $K_L =$ 4,096,000 trial paths. For our estimate of the upper bound U, we produced $K_U =$ 1,024 paths $(z_n^k)_{n=0}^N$, $k = 1,\dots,K_U$, of $(X_n)_{n=0}^N$ and $K_U \times J$ realizations $(v_n^{k,j})_{n=1}^N$, $k = 1,\dots,K_U$, $j = 1,\dots,J$, of $(W_{t_n} - W_{t_{n-1}})_{n=1}^N$ with $J =$ 16,384. Then for all n and k, we generated the i-th component of the j-th continuation path departing from z_n^k according to

$$\tilde{z}_m^{i,k,j} = z_n^{i,k} \exp\left([r - \delta_i - \sigma_i^2/2](m-n)\Delta t + \sigma_i[v_{n+1}^{i,k,j} + \dots + v_m^{i,k,j}]\right), \quad m = n+1,\dots,N.$$

16. That is, $\mathbb{E}[(W_t^i - W_s^i)(W_t^j - W_s^j)] = \rho_{ij}(t-s)$ for all $i \neq j$ and $s < t$.
17. In fact, in none of the examples in this paper it is optimal to stop at time 0. So $\tau^\Theta = \tau_1$ in all these cases.

Symmetric case

We first considered the special case, where $s_0^i = s_0$, $\delta_i = \delta$, $\sigma_i = \sigma$ for all $i = 1, \dots, d$, and $\rho_{ij} = \rho$ for all $i \neq j$. Our results are reported in Table 1.

Asymmetric case

As a second example, we studied model (20) with $s_0^i = s_0$, $\delta_i = \delta$ for all $i = 1, 2, \dots, d$, and $\rho_{ij} = \rho$ for all $i \neq j$, but different volatilities $\sigma_1 < \sigma_2 < \dots < \sigma_d$. For $d \leq 5$, we chose the specification $\sigma_i = 0.08 + 0.32 \times (i-1)/(d-1)$, $i = 1, 2, \dots, d$. For $d > 5$, we set $\sigma_i = 0.1 + i/(2d)$, $i = 1, 2, \dots, d$. The results are given in Table 2.

4.2. Callable Multi Barrier Reverse Convertibles

A MBRC is a coupon paying security that converts into shares of the worst-performing of d underlying assets if a prespecified trigger event occurs. Let us assume that the price of the i-th underlying asset in percent of its starting value follows the risk-neutral dynamics

$$S_t^i = \begin{cases} 100 \exp\big([r - \sigma_i^2/2]t + \sigma_i W_t^i\big) & \text{for } t \in [0, T_i) \\ 100(1-\delta_i) \exp\big([r - \sigma_i^2/2]t + \sigma_i W_t^i\big) & \text{for } t \in [T_i, T] \end{cases} \tag{21}$$

for a risk-free interest rate $r \in \mathbb{R}$, volatility $\sigma_i \in (0, \infty)$, maturity $T \in (0, \infty)$, dividend payment time $T_i \in (0, T)$, dividend rate $\delta_i \in [0, \infty)$ and a d-dimensional Brownian motion W with constant instantaneous correlations $\rho_{ij} \in \mathbb{R}$ between different components W^i and W^j.

Let us consider a MBRC that pays a coupon c at each of N time points $t_n = nT/N$, $n = 1, 2, \dots, N$, and makes a time-T payment of

$$G = \begin{cases} F & \text{if } \min_{1 \leq i \leq d} \min_{1 \leq m \leq M} S_{u_m}^i > B \text{ or } \min_{1 \leq i \leq d} S_T^i > K \\ \min_{1 \leq i \leq d} S_T^i & \text{if } \min_{1 \leq i \leq d} \min_{1 \leq m \leq M} S_{u_m}^i \leq B \text{ and } \min_{1 \leq i \leq d} S_T^i \leq K, \end{cases}$$

where $F \in [0, \infty)$ is the nominal amount, $B \in [0, \infty)$ a barrier, $K \in [0, \infty)$ a strike price and u_m the end of the m-th trading day. Its value is

$$\sum_{n=1}^N e^{-rt_n} c + e^{-rT} \mathbb{E}\, G \tag{22}$$

and can easily be estimated with a standard Monte Carlo approximation.

A callable MBRC can be redeemed by the issuer at any of the times $t_1, t_2, \dots, t_{N-1}$ by paying back the notional. To minimize costs, the issuer will try to find a $\{t_1, t_2, \dots, T\}$-valued stopping time such that

$$\mathbb{E}\left[\sum_{n=1}^{\tau} e^{-rt_n} c + 1_{\{\tau < T\}} e^{-r\tau} F + 1_{\{\tau = T\}} e^{-rT} G\right]$$

is minimal.

Let $(X_n)_{n=1}^N$ be the $d+1$-dimensional Markov process given by $X_n^i = S_{t_n}^i$ for $i = 1, \dots, d$, and

$$X_n^{d+1} := \begin{cases} 1 & \text{if the barrier has been breached before or at time } t_n \\ 0 & \text{else.} \end{cases}$$

d	s_0	$\hat{L}$	t_L	$\hat{U}$	t_U	Point est.	95% CI	Literature
2	90	8.072	28.7	8.075	25.4	8.074	$[8.060, 8.081]$	8.075
2	100	13.895	28.7	13.903	25.3	13.899	$[13.880, 13.910]$	13.902
2	110	21.353	28.4	21.346	25.3	21.349	$[21.336, 21.354]$	21.345
3	90	11.290	28.8	11.283	26.3	11.287	$[11.276, 11.290]$	11.29
3	100	18.690	28.9	18.691	26.4	18.690	$[18.673, 18.699]$	18.69
3	110	27.564	27.6	27.581	26.3	27.573	$[27.545, 27.591]$	27.58
5	90	16.648	27.6	16.640	28.4	16.644	$[16.633, 16.648]$	$[16.620, 16.653]$
5	100	26.156	28.1	26.162	28.3	26.159	$[26.138, 26.174]$	$[26.115, 26.164]$
5	110	36.766	27.7	36.777	28.4	36.772	$[36.745, 36.789]$	$[36.710, 36.798]$
10	90	26.208	30.4	26.272	33.9	26.240	$[26.189, 26.289]$	
10	100	38.321	30.5	38.353	34.0	38.337	$[38.300, 38.367]$	
10	110	50.857	30.8	50.914	34.0	50.886	$[50.834, 50.937]$	
20	90	37.701	37.2	37.903	44.5	37.802	$[37.681, 37.942]$	
20	100	51.571	37.5	51.765	44.3	51.668	$[51.549, 51.803]$	
20	110	65.494	37.3	65.762	44.4	65.628	$[65.470, 65.812]$	
30	90	44.797	45.1	45.110	56.2	44.953	$[44.777, 45.161]$	
30	100	59.498	45.5	59.820	56.3	59.659	$[59.476, 59.872]$	
30	110	74.221	45.3	74.515	56.2	74.368	$[74.196, 74.566]$	
50	90	53.903	58.7	54.211	79.3	54.057	$[53.883, 54.266]$	
50	100	69.582	59.1	69.889	79.3	69.736	$[69.560, 69.945]$	
50	110	85.229	59.0	85.697	79.3	85.463	$[85.204, 85.763]$	
100	90	66.342	95.5	66.771	147.7	66.556	$[66.321, 66.842]$	
100	100	83.380	95.9	83.787	147.7	83.584	$[83.357, 83.862]$	
100	110	100.420	95.4	100.906	147.7	100.663	$[100.394, 100.989]$	
200	90	78.993	170.9	79.355	274.6	79.174	$[78.971, 79.416]$	
200	100	97.405	170.1	97.819	274.3	97.612	$[97.381, 97.889]$	
200	110	115.800	170.6	116.377	274.5	116.088	$[115.774, 116.472]$	
500	90	95.956	493.4	96.337	761.2	96.147	$[95.934, 96.407]$	
500	100	116.235	493.5	116.616	761.7	116.425	$[116.210, 116.685]$	
500	110	136.547	493.7	136.983	761.4	136.765	$[136.521, 137.064]$	

Table 1: Summary results for max-call options on d symmetric assets for parameter values of $r = 5\%$, $\delta = 10\%$, $\sigma = 20\%$, $\rho = 0$, $K = 100$, $T = 3$, $N = 9$. t_L is the number of seconds it took to train τ^{Θ} and compute $\hat{L}$. t_U is the computation time for $\hat{U}$ in seconds. 95% CI is the 95% confidence interval (19). The last column lists values calculated with a binomial lattice method by Andersen and Broadie (2004) for $d = 2$–3 and the 95% confidence intervals of Broadie and Cao (2008) for $d = 5$.

d	s_0	$\hat{L}$	t_L	$\hat{U}$	t_U	Point est.	95% CI	Literature
2	90	14.325	26.8	14.352	25.4	14.339	$[14.299, 14.367]$	
2	100	19.802	27.0	19.813	25.5	19.808	$[19.772, 19.829]$	
2	110	27.170	26.5	27.147	25.4	27.158	$[27.138, 27.163]$	
3	90	19.093	26.8	19.089	26.5	19.091	$[19.065, 19.104]$	
3	100	26.680	27.5	26.684	26.4	26.682	$[26.648, 26.701]$	
3	110	35.842	26.5	35.817	26.5	35.829	$[35.806, 35.835]$	
5	90	27.662	28.0	27.662	28.6	27.662	$[27.630, 27.680]$	$[27.468, 27.686]$
5	100	37.976	27.5	37.995	28.6	37.985	$[37.940, 38.014]$	$[37.730, 38.020]$
5	110	49.485	28.2	49.513	28.5	49.499	$[49.445, 49.533]$	$[49.155, 49.531]$
10	90	85.937	31.8	86.037	34.4	85.987	$[85.857, 86.087]$	
10	100	104.692	30.9	104.791	34.2	104.741	$[104.603, 104.864]$	
10	110	123.668	31.0	123.823	34.4	123.745	$[123.570, 123.904]$	
20	90	125.916	38.4	126.275	45.6	126.095	$[125.819, 126.383]$	
20	100	149.587	38.2	149.970	45.2	149.779	$[149.480, 150.053]$	
20	110	173.262	38.4	173.809	45.3	173.536	$[173.144, 173.937]$	
30	90	154.486	46.5	154.913	57.5	154.699	$[154.378, 155.039]$	
30	100	181.275	46.4	181.898	57.5	181.586	$[181.155, 182.033]$	
30	110	208.223	46.4	208.891	57.4	208.557	$[208.091, 209.086]$	
50	90	195.918	60.7	196.724	81.1	196.321	$[195.793, 196.963]$	
50	100	227.386	60.7	228.386	81.0	227.886	$[227.247, 228.605]$	
50	110	258.813	60.7	259.830	81.1	259.321	$[258.661, 260.092]$	
100	90	263.193	98.5	264.164	151.2	263.679	$[263.043, 264.425]$	
100	100	302.090	98.2	303.441	151.2	302.765	$[301.924, 303.843]$	
100	110	340.763	97.8	342.387	151.1	341.575	$[340.580, 342.781]$	
200	90	344.575	175.4	345.717	281.0	345.146	$[344.397, 346.134]$	
200	100	392.193	175.1	393.723	280.7	392.958	$[391.996, 394.052]$	
200	110	440.037	175.1	441.594	280.8	440.815	$[439.819, 441.990]$	
500	90	476.293	504.5	477.911	760.7	477.102	$[476.069, 478.481]$	
500	100	538.748	504.6	540.407	761.6	539.577	$[538.499, 540.817]$	
500	110	601.261	504.9	603.243	760.8	602.252	$[600.988, 603.707]$	

Table 2: Summary results for max-call options on d asymmetric assets for parameter values of $r = 5\%$, $\delta = 10\%$, $\rho = 0$, $K = 100$, $T = 3$, $N = 9$. t_L is the number of seconds it took to train τ^Θ and compute $\hat{L}$. t_U is the computation time for $\hat{U}$ in seconds. 95% CI is the 95% confidence interval (19). The last column reports the 95% confidence intervals of Broadie and Cao (2008).

Then the issuer's minimization problem can be written as

$$\inf_{\tau \in \mathcal{T}} \mathbb{E}\, g(\tau, X_\tau), \tag{23}$$

where $\mathcal{T}$ is the set of all X-stopping times and

$$g(n,x) = \begin{cases} \sum_{m=1}^{n} e^{-rt_m} c + e^{-rt_n} F & \text{if } 1 \le n \le N-1 \text{ or } x^{d+1} = 0 \\ \sum_{m=1}^{N} e^{-rt_m} c + e^{-rt_N} h(x) & \text{if } n = N \text{ and } x^{d+1} = 1, \end{cases}$$

where

$$h(x) = \begin{cases} F & \text{if } \min_{1 \le i \le d} x^i > K \\ \min_{1 \le i \le d} x^i & \text{if } \min_{1 \le i \le d} x^i \le K. \end{cases}$$

Since the issuer cannot redeem at time 0, we trained stopping times of the form

$$\tau^\Theta = \sum_{n=1}^{N} n f^{\theta_n}(Y_n) \prod_{j=1}^{n-1} (1 - f^{\theta_j}(Y_k)) \in \mathcal{T}_1$$

for $f^{\theta_N} \equiv 1$ and $f^{\theta_1}, \dots, f^{\theta_{N-1}} \colon \mathbb{R}^{d+1} \to \{0,1\}$ given by (8) with $I = 2$ and $q_1 = q_2 = d+40$. Since (23) is a minimization problem, τ^Θ yields an upper bound and the dual method a lower bound.

We simulated the model (21) like (20) in Subsection 4.1 with the same number of trials except that here we used the lower number $J = 1{,}024$ to estimate the dual bound. Numerical results are reported in Table 3.

4.3. Optimally Stopping a Fractional Brownian Motion

A fractional Brownian motion with Hurst parameter $H \in (0,1]$ is a continuous centered Gaussian process $(W^H_t)_{t \ge 0}$ with covariance structure

$$\mathbb{E}[W^H_t W^H_s] = \frac{1}{2}\left(t^{2H} + s^{2H} - |t-s|^{2H}\right);$$

see, e.g., Mandelbrot and Van Ness (1968) or Samoradnitsky and Taqqu (1994). For $H = 1/2$, W^H is a standard Brownian motion. So, by the optional stopping theorem, one has $\mathbb{E}\, W^{1/2}_\tau = 0$ for every $W^{1/2}$-stopping time τ bounded above by a constant; see, e.g., Grimmett and Stirzaker (2001). However, for $H \neq 1/2$, the increments of W^H are correlated – positively for $H \in (1/2, 1]$ and negatively for $H \in (0, 1/2)$. In both cases, W^H is neither a martingale nor a Markov process, and there exist bounded W^H-stopping times τ such that $\mathbb{E}\, W^H_\tau > 0$; see, e.g., Kulikov and Gusyatnikov (2016) for two classes of simple stopping rules $0 \le \tau \le 1$ and estimates of the corresponding expected values $\mathbb{E}\, W^H_\tau$.

To approximate the supremum

$$\sup_{0 \le \tau \le 1} \mathbb{E}\, W^H_\tau \tag{24}$$

d	ρ	$\hat{L}$	t_L	$\hat{U}$	t_U	Point est.	95% CI	Non-callable
2	0.6	98.235	24.9	98.252	204.1	98.243	[98.213, 98.263]	106.285
2	0.1	97.634	24.9	97.634	198.8	97.634	[97.609, 97.646]	106.112
3	0.6	96.930	26.0	96.936	212.9	96.933	[96.906, 96.948]	105.994
3	0.1	95.244	26.2	95.244	211.4	95.244	[95.216, 95.258]	105.553
5	0.6	94.865	41.0	94.880	239.2	94.872	[94.837, 94.894]	105.530
5	0.1	90.807	41.1	90.812	238.4	90.810	[90.775, 90.828]	104.496
10	0.6	91.568	71.3	91.629	300.9	91.599	[91.536, 91.645]	104.772
10	0.1	83.110	71.7	83.137	301.8	83.123	[83.078, 83.153]	102.495
15	0.6	89.558	94.9	89.653	359.8	89.606	[89.521, 89.670]	104.279
15	0.1	78.495	94.7	78.557	360.5	78.526	[78.459, 78.571]	101.209
30	0.6	86.089	158.5	86.163	534.1	86.126	[86.041, 86.180]	103.385
30	0.1	72.037	159.3	72.749	535.6	72.393	[71.830, 72.760]	99.279

Table 3: Summary results for callable MBRCs with d underlying assets for $F = K = 100$, $B = 70$, $T = 1$ year (= 252 trading days), $N = 12$, $c = 7/12$, $\delta_i = 5\%$, $T_i = 1/2$, $r = 0$, $\sigma_i = 0.2$ and $\rho_{ij} = \rho$ for $i \neq j$. t_U is the number of seconds it took to train τ^Θ and compute $\hat{U}$. t_L is the number of seconds it took to compute $\hat{L}$. The last column lists fair values of the same MBRCs without the callable feature. We estimated them by averaging 4,096,000 Monte Carlo samples of the payoff. This took between 5 (for $d = 2$) and 44 (for $d = 30$) seconds.

over all W^H-stopping times $0 \leq \tau \leq 1$, we denote $t_n = n/100$, $n = 0, 1, 2, \ldots, 100$, and introduce the 100-dimensional Markov process $(X_n)_{n=0}^{100}$ given by

$$\begin{aligned}
X_0 &= (0, 0, \ldots, 0)\\
X_1 &= (W^H_{t_1}, 0, \ldots, 0)\\
X_2 &= (W^H_{t_2}, W^H_{t_1}, 0, \ldots, 0)\\
&\vdots\\
X_{100} &= (W^H_{t_{100}}, W^H_{t_{99}}, \ldots, W^H_{t_1}).
\end{aligned}$$

The discretized stopping problem

$$\sup_{\tau \in \mathcal{T}} \mathbb{E}\, g(X_\tau), \tag{25}$$

where $\mathcal{T}$ is the set of all X-stopping times and $g \colon \mathbb{R}^{100} \to \mathbb{R}$ the projection $(x^1, \ldots, x^{100}) \mapsto x^1$, approximates (24) from below.

We computed estimates of (25) for $H \in \{0.01, 0.05, 0.1, 0.15, \ldots, 1\}$ by training networks of the form (8) with depth $I = 2$, $d = 100$ and $q_1 = q_2 = 140$. To simulate the vector $Y = (W^H_{t_n})_{n=0}^{100}$, we used the representation $Y = BZ$, where BB^T is the Cholesky decomposition of the covariance matrix of Y and Z a 100-dimensional random vector with independent standard normal components. We carried out 6,000 training steps with a batch size of 2,048. To estimate the lower bound L we generated $K_L = 4{,}096{,}000$ simulations of Z. For our estimate of the upper bound U, we first simulated $K_U = 1{,}024$ realizations v^k,

$k = 1, \dots, K_U$ of Z and set $w^k = Bv^k$. Then we produced another $K_U \times J$ simulations $\tilde{v}^{k,j}$, $k = 1, \dots, K_U$, $j = 1, \dots, J$, of Z, and generated for all n and k, continuation paths starting from

$$z_n^k = (w_n^k, \dots, w_1^k, 0, \dots, 0)$$

according to

$$\tilde{z}_m^{k,j} = (\tilde{w}_m^{k,j}, \dots, \tilde{w}_{n+1}^{k,j}, w_n^k, \dots, w_1^k, 0 \dots, 0), \quad m = n+1, \dots, 100,$$

with

$$\tilde{w}_l^{k,j} = \sum_{i=1}^{n} B_{li} v_i^k + \sum_{i=n+1}^{l} B_{li} \tilde{v}_i^{k,j}, \quad l = n+1, \dots, m.$$

For $H \in \{0.01, ..., 0.4\} \cup \{0.6, ..., 1.0\}$, we chose $J = 16{,}384$, and for $H \in \{0.45, 0.5, 0.55\}$, $J = 32{,}768$. The results are listed in Table 4 and depicted in graphical form in Figure 1. Note that for $H = 1/2$ and $H = 1$, our 95% confidence intervals contain the true values, which in these two cases, can be calculated exactly. As mentioned above, $W^{1/2}$ is a Brownian motion, and therefore, $\mathbb{E}\, W_\tau^{1/2} = 0$ for every $(W_{t_n}^{1/2})_{n=0}^{100}$-stopping time τ. On the other hand, one has[18] $W_t^1 = tW_1^1$, $t \geq 0$. So, in this case, the optimal stopping time is given[18] by

$$\tau = \begin{cases} 1 & \text{if } W_{t_1}^1 > 0 \\ t_1 & \text{if } W_{t_1}^1 \leq 0, \end{cases}$$

and the corresponding expectation by

$$\mathbb{E}\, W_\tau^1 = \mathbb{E}\left[W_1^1 1_{\left\{W_{t_1}^1 > 0\right\}} - W_{t_1}^1 1_{\left\{W_{t_1}^1 \leq 0\right\}}\right] = 0.99\, \mathbb{E}\left[W_1^1 1_{\{W_1^1 > 0\}}\right] = 0.99/\sqrt{2\pi} = 0.39495...$$

Moreover, it can be seen that for $H \in (1/2, 1)$, our estimates are up to three times higher than the expected payoffs generated by the heuristic stopping rules of Kulikov and Gusyatnikov (2016). For $H \in (0, 1/2)$, they are up to five times higher.

Acknowledgments

We thank Philippe Ehlers, Ariel Neufeld, Martin Stefanik, the action editor and the referees for fruitful discussions and helpful comments.

References

Charalambos D. Aliprantis and Kim C. Border. *Infinite Dimensional Analysis*. Springer, Berlin, 3rd Edition, 2006.

Leif Andersen. A simple approach to the pricing of Bermudan swaptions in the multifactor LIBOR market model. *The Journal of Computational Finance*, 3(2):5–32, 2000.

18. up to $\mathbb{P}$-almost sure equality

H	$\hat{L}$	$\hat{U}$	Point est.	95% CI
0.01	1.518	1.519	1.519	$[1.517, 1.520]$
0.05	1.293	1.293	1.293	$[1.292, 1.294]$
0.10	1.048	1.049	1.049	$[1.048, 1.050]$
0.15	0.838	0.839	0.839	$[0.838, 0.840]$
0.20	0.658	0.659	0.658	$[0.657, 0.659]$
0.25	0.501	0.504	0.503	$[0.501, 0.505]$
0.30	0.369	0.370	0.370	$[0.368, 0.371]$
0.35	0.255	0.256	0.255	$[0.254, 0.257]$
0.40	0.155	0.158	0.156	$[0.154, 0.158]$
0.45	0.067	0.075	0.071	$[0.066, 0.075]$
0.50	0.000	0.005	0.002	$[0.000, 0.005]$
0.55	0.057	0.065	0.061	$[0.057, 0.065]$
0.60	0.115	0.118	0.117	$[0.115, 0.119]$
0.65	0.163	0.165	0.164	$[0.163, 0.166]$
0.70	0.206	0.207	0.207	$[0.205, 0.208]$
0.75	0.242	0.245	0.244	$[0.242, 0.245]$
0.80	0.276	0.278	0.277	$[0.276, 0.279]$
0.85	0.308	0.309	0.308	$[0.307, 0.310]$
0.90	0.336	0.339	0.337	$[0.335, 0.339]$
0.95	0.365	0.367	0.366	$[0.365, 0.367]$
1.00	0.395	0.395	0.395	$[0.394, 0.395]$

Table 4: Estimates of $\sup_{\tau \in \{0, t_1, \dots, 1\}} \mathbb{E}\, W^H_\tau$. For all $H \in \{0.01, 0.05, \dots, 1\}$, it took about 430 seconds to train τ^Θ and compute $\hat{L}$. The computation of $\hat{U}$ took about 17,000 seconds for $H \in \{0.01, \dots, 0.4\} \cup \{0.6, \dots, 1\}$ and about 34,000 seconds for $H \in \{0.45, 0.5, 0.55\}$.

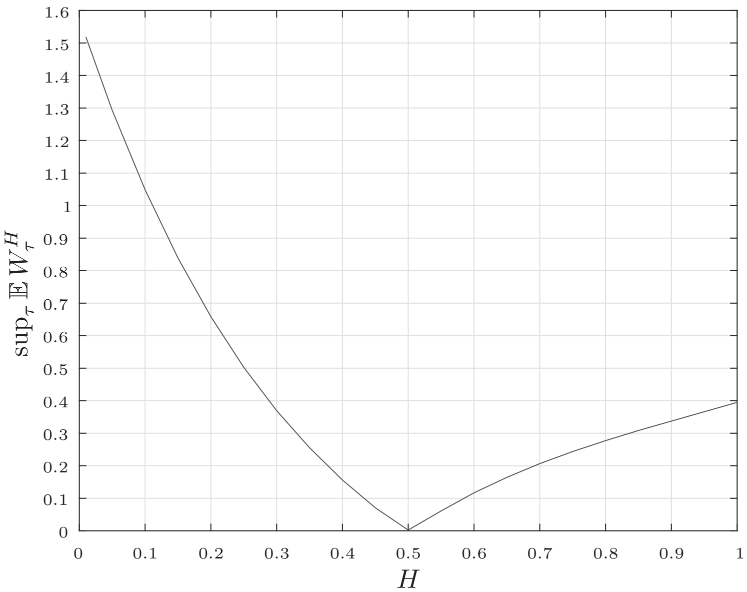

Figure 1: Estimates of $\sup_{\tau \in \{0, t_1, \ldots, 1\}} \mathbb{E}\, W^H_\tau$ for different values of H.

Leif Andersen and Mark Broadie. Primal-dual simulation algorithm for pricing multidimensional American options. *Management Science*, 50(9):1222–1234, 2004.

Vlad Bally, Gilles Pagès, and Jacques Printems. A quantization tree method for pricing and hedging multidimensional American options. *Mathematical Finance*, 15(1):119–168, 2005.

Jérôme Barraquand and Didier Martineau. Numerical valuation of high dimensional multivariate American securities. *The Journal of Financial and Quantitative Analysis*, 30(3):383–405, 1995.

Denis Belomestny. On the rates of convergence of simulation-based optimization algorithms for optimal stopping problems. *The Annals of Applied Probability*, 21(1):215–239, 2011.

Denis Belomestny. Solving optimal stopping problems via empirical dual optimization. *The Annals of Applied Probability*, 23(5):1988–2019, 2013.

Denis Belomestny, Christian Bender, and John Schoenmakers. True upper bounds for Bermudan products via non-nested Monte Carlo. *Mathematical Finance*, 19(1):53–71, 2009.

Denis Belomestny, Mark Joshi, and John Schoenmakers. Multilevel dual approach for pricing American style derivatives. *Finance and Stochastics*, 17(4):717–742, 2013.

Denis Belomestny, John Schoenmakers, Vladimir Spokoiny, and Yuri Tavyrikov. Optimal stopping via reinforced regression. *arXiv:1808.02341*, 2018.

Steffan J. Berridge and Johannes M. Schumacher. An irregular grid approach for pricing high-dimensional American options. *Journal of Computational and Applied Mathematics*, 222(1):94–111, 2008.

Phelim P. Boyle, Adam W. Kolkiewicz, and Ken Seng Tan. An improved simulation method for pricing high-dimensional American derivatives. *Mathematics and Computers in Simulation*, 62:315–322, 2003.

Mark Broadie and Menghui Cao. Improved lower and upper bound algorithms for pricing American options by simulation. *Quantitative Finance*, 8(8):845–861, 2008.

Mark Broadie and Paul Glasserman. A stochastic mesh method for pricing high-dimensional American options. *The Journal of Computational Finance*, 7(4):35–72, 2004.

Jacques F. Carriere. Valuation of the early-exercise price for options using simulations and nonparametric regression. *Insurance: Mathematics and Economics*, 19(1):19–30, 1996.

Nan Chen and Paul Glasserman. Additive and multiplicative duals for American option pricing. *Finance and Stochastics*, 11(2):153–179, 2007.

Mark H. A. Davis and Ioannis Karatzas. A deterministic approach to optimal stopping. In *Probability, Statistics and Optimisation: A Tribute to Peter Whittle (ed. Frank P. Kelly)*, pages 455–466. John Wiley & Sons, NewYork, 1994.

Vijay V. Desai, Vivek F. Farias, and Ciamac C. Moallemi. Pathwise optimization for optimal stopping problems. *Management Science*, 58(12):2292–2308, 2012.

Daniel Egloff, Michael Kohler, and Nebojsa Todorovic. A dynamic look-ahead Monte Carlo algorithm for pricing Bermudan options. *The Annals of Applied Probability*, 17(4):1138–1171, 2007.

Diego García. Convergence and biases of Monte Carlo estimates of American option prices using a parametric exercise rule. *Journal of Economic Dynamics and Control*, 27(10):1855–1879, 2003.

Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, *PMLR*, 9:249–256, 2010.

Geoffrey Grimmett and David Stirzaker. *Probability and Random Processes*. Oxford University Press, 3rd Edition, 2001.

Jiequn Han and Weinan E. Deep learning approximation for stochastic control problems. *Deep Reinforcement Learning Workshop, NIPS*, 2016.

Martin B. Haugh and Leonid Kogan. Pricing American options: a duality approach. *Operations Research*, 52(2):258–270, 2004.

Sergey Ioffe and Christian Szegedy. Batch normalization: accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on Machine Learning*, *PMLR*, 37:448–456, 2015.

Shashi Jain and Cornelis W. Oosterlee. The stochastic grid bundling method: efficient pricing of Bermudan options and their Greeks. *Applied Mathematics and Computation*, 269:412–431, 2015.

Farshid Jamshidian. The duality of optimal exercise and domineering claims: a Doob–Meyer decomposition approach to the Snell envelope. *Stochastics*, 79:27–60, 2007.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 2015.

Michael Kohler, Adam Krzyżak, and Nebojsa Todorovic. Pricing of high-dimensional American options by neural networks. *Mathematical Finance*, 20(3):383–410, 2010.

Anastasia Kolodko and John Schoenmakers. Iterative construction of the optimal Bermudan stopping time. *Finance and Stochastics*, 10(1):27–49, 2006.

Alexander V. Kulikov and Pavel P. Gusyatnikov. Stopping times for fractional Brownian motion. In *Computational Management Science*, Volume 682 of *Lecture Notes in Economics and Mathematical Systems*, pages 195–200. Springer International Publishing, 2016.

Damien Lamberton and Bernard Lapeyre. *Introduction to Stochastic Calculus Applied to Finance*. Chapman & Hall/CRC, 2nd Edition, 2008.

Jérôme Lelong. Pricing American options using martingale bases. *arXiv:1604.03317*, 2016.

Moshe Leshno, Vladimir Ya. Lin, Allan Pinkus, and Shimon Schocken. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, 6(6):861–867, 1993.

Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. *International Conference on Learning Representations*, 2016.

Francis A. Longstaff and Eduardo S. Schwartz. Valuing American options by simulation: a simple least-squares approach. *The Review of Financial Studies*, 14(1):113–147, 2001.

Benoit B. Mandelbrot and John W. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10(4):422–437, 1968.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu et al. Human-level control through deep reinforcement learning. *Nature*, 518:529–533, 2015.

Goran Peskir and Albert N. Shiryaev. *Optimal Stopping and Free-Boundary Problems*. Lectures in Mathematics. Birkhäuser Basel, 2006.

Chris Rogers. Monte Carlo valuation of American options. *Mathematical Finance*, 12(3):271–286, 2002.

Chris Rogers. Dual valuation and hedging of Bermudan options. *SIAM Journal on Financial Mathematics*, 1(1):604–608, 2010.

Gennady Samoradnitsky and Murad S. Taqqu. *Stable Non-Gaussian Random Processes*. Chapman and Hall/CRC, 1994.

John Schulman, Sergey Levine, Philipp Moritz, Michael Jordan, and Pieter Abbeel. Trust region policy optimization. In *Proceedings of the 32nd International Conference on Machine Learning*, *PMLR*, 37:1889–1897, 2015.

Martin Schweizer. On Bermudan options. In *Advances in Finance and Stochastics*, pages 257–270. Springer Berlin Heidelberg, 2002.

David Silver, Aja Huang, Chris J. Maddison, Arthur Guez et al. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529:484–489, 2016.

Justin Sirignano and Konstantinos Spiliopoulos. DGM: A deep learning algorithm for solving partial differential equations. *Journal of Computational Physics*, 375:1339–1364, 2018.

Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning*. The MIT Press, 1998.

James A. Tilley. Valuing American options in a path simulation model. *Transactions of the Society of Actuaries*, 45:83–104, 1993.

John N. Tsitsiklis and Benjamin Van Roy. Regression methods for pricing complex American-style options. *IEEE Transactions on Neural Networks*, 12(4):694–703, 2001.