

Neural Optimal Stopping Boundary

A. Max Reppen* H. Mete Soner† Valentin Tissot-Daguette‡

April 29, 2022

Abstract

A method based on deep artificial neural networks and empirical risk minimization is developed to calculate the boundary separating the stopping and continuation regions in optimal stopping. The algorithm parameterizes the stopping boundary as the graph of a function and introduces relaxed stopping rules based on fuzzy boundaries to facilitate efficient optimization. Several financial instruments, some in high dimensions, are analyzed through this method, demonstrating its effectiveness. The existence of the stopping boundary is also proved under natural structural assumptions.

Key words: Stopping boundary problems, American derivatives, Bermudan options, optimal stopping, deep learning, fuzzy boundary.

Mathematics Subject Classification: 91G20, 91G60, 68T07, 35R35.

1 Introduction

The classical decision problem of optimal stopping has found amazing range of applications from finance, statistics, marketing, phase transitions to engineering. While over the past decades many efficient methods have been developed for its numerical resolution, until recently essentially all computational approaches in high-dimensions first approximate the maximal value. The quantity that is equally important is the associated stopping rule, and the state is divided into two regions depending on whether at a point it is optimal to stop or to continue. Although approximate optimal actions can either be calculated directly, or are always available from the calculated value function in a greedy policy, a characterization and computation of these regions through interpretable criterion would provide an immediate link between the policy and the state.

In many important applications, continuation and stopping regions are separated by graphs of functions that can be used to construct interpretable optimal stopping decisions, and our goal is to directly compute these *stopping boundaries* whose performances are close to the maximal ones. In lower dimensions, they can be accurately approximated by methods

*Questrom School of Business, Boston University, Boston, MA, 02215, USA, email: amreppen@bu.edu.

†Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ, 08540, USA, email: soner@princeton.edu. Research of Soner was partially supported by the National Science Foundation grant DMS 2106462.

‡Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ, 08540, USA, email: v.tissot-daguette@princeton.edu

based on nonlinear partial differential equations or dynamic programming. However, in models with many degrees of freedom, this numerical problem is still challenging.

Our approach is based on a method proposed by E, Jentzen & Han [30, 31], which we call *deep empirical risk minimization* (deep ERM). In this framework, the controller is replaced by a deep artificial neural network and the random system dynamics is simulated via Monte-Carlo. Recent advances in optimization tools and computational capabilities enable us to optimize this system efficiently, thus providing an accurate tool potentially also applicable to other problems with stopping boundaries, including models with regime change or other singularities in financial economics, phase transitions in continuum mechanics, and Stefan type problems in solidification.

Deep ERM applies essentially to all problems that can be formulated as dynamic stochastic optimization, maybe after problem specific modifications. This and similar algorithms [2, 6, 8, 9, 15, 16, 29, 33, 45, 52] have been applied to many classical problems from quantitative finance, financial economics, and to nonlinear partial differential equations that have stochastic control representations. Also, the application to optimal stopping is carried out by Becker, Jentzen & Cheridito [6, 8] to compute an approximation of the optimal stopping rule and hence, the stopping region as well as the maximal value. They accurately price many American options of practical importance in very high dimensions with confidence intervals, showcasing the power and the flexibility of deep ERM in optimal stopping as well. We refer the readers to our accompanying papers [42, 43] for an introduction of deep ERM, and to the excellent surveys [23, 34, 46] and the references therein for more information.

The stopping regions are exactly the sets of all points at which the value and the pay-off functions agree, and they can be derived from the calculated value or computed directly. However, qualitatively these approximations may not always have accurate geometric properties, and lack interpretability. Thus a direct computation of these regions with their known structures is desirable, and it is the main goal of our approach. Indeed, in many financial applications they are characterized as the epi or hypo-graph of certain functions in a natural but problem specific coordinate system, described in Assumption 2.3. We first outline the algorithm and its properties under this assumption, and then verify it in Theorem 4.1 for all examples that we consider. In contrast to methods mapping each state point to stopping decisions, our representation of stopping boundaries as graphs provides topological guarantees, even when approximated.

In our method, the graph separating the continuation and stopping sets is approximated by a deep artificial neural network which then provides a stopping rule used to calculate an empirical reward function. However, it is not possible to directly use any gradient method for optimization over these “stop-or-go” decisions. Therefore, although at the high-level our approach is deep ERM, it requires one important modification to overcome this difficulty. We replace these sharp rules based on the hitting times, by stopping probabilities. In this relaxed formulation, when the state is not close to the boundary, we still stop or continue the process as in the non-relaxed case. But when it is close to the boundary, in a region called the *fuzzy (stopping) boundary*, we stop with a probability proportional to the distance to the boundary. This is analogous to mushy regions in solidification or phase fields models in continuum mechanics [3, 11, 25, 41, 48, 50]. Similarly [6, 8] uses stopping probabilities as relaxed control variables but without any connection to the geometric structures.

We numerically study several American and Bermudan options with this methodology. The numerical results for the American put options in Black & Scholes, and in Heston models verify the effectiveness of the method. The Bermudan max-call option, studied in high dimensions, shows once again the power of deep ERM. The numerical results for the look-back options is an example of the flexibility of this approach in handling path-dependent options. When possible all our computations is benchmarked to previous studies, and shown to be comparable. Also, confidence intervals and upper bounds computed in [6, 7, 8],

provide computable guarantees. Additionally, the stopping regions, Figures 6, 7, 8, for the two-dimensional max-call options are qualitatively similar to those obtained in [13].

Direct computation of the stopping boundary has been the object of several other studies as well. In low dimensions, techniques from nonlinear partial differential equations modeling obstacle problems and phase transitions allow for accurate calculations. Also, Garcia [26] proposes a method similar to ours based on a parsimonious parametrization of the stopping region. A recent study by Ciocan & Mišić [18] proposes a tree-based method to compute this partition. Alternatively, deep Galerkin method for the differential equations is used by [47] to compute the value function of basket options at every point and the stopping boundary is derived from this function. [51] studies the classical Stephan problem of melting ice in low dimensions, by the same approach. We also refer the readers to [4] for a review of numerical methods based on boundary parametrization as well as other approaches to optimal stopping.

For American option pricing with many degrees of freedom, the main alternative computational tool to deep ERM are the Monte-Carlo based regression methods described in Glasserman [28], Longstaff & Schwartz [39] and Tsitsiklis & van Roy [49]. The classical book of Detemple [22] and the recent article of Ludkovski [40] provide extensive information on simulation based computational techniques for optimal stopping. The key difference between these approaches, and the value computation through deep ERM is the choice of the set of basis functions that grows rapidly with the underlying dimension. Deep artificial neural networks do not require an *a priori* specification of the basis. Therefore, they are likely to be more effective for problems with many states, as supported by the experiments. Indeed, [7, 37] exploit this property and replace the preset hypothesis class by an artificial neural network in their regression.

The power of these new approaches are more apparent in high dimensional settings, and an interesting example is the American options with rough volatility models. These are infinite-dimensional models, and their numerical analysis is given in Bayer *et al.* [4, 5] and in Chevalier *at. al.* [17] by alternative methods. In a different class of problems with many states, [27] exploits the symmetry of the underlying problem to build an appropriate architecture of the neural networks. This approach is particularly relevant for mean-field games which models many agents that are identical. Additionally, [6, 7, 8] consider several models with many variables including the fractional Brownian motion.

The paper is organized as follows. The problem and the value function is defined in the next section. The algorithm is introduced in Section 3, and the financial structure is outlined in Section 4. Section 5 details the network architecture and parameters used in our experiments. One dimensional put options in the Black & Scholes model is studied in Section 6, in the Heston model in Section 7, Bermudan max-call options are the topic of Section 8, and the look-back options are the foci of Section 9. Appendix A proves Theorem 4.1, Appendix B proves the convergence of the reward function \mathcal{R}_ϵ as the width of the mushy region tends to zero. Appendix C provides the details of the algorithm.

2 Optimal Stopping

Let $T > 0$ be the finite time horizon, and (Ω, \mathbb{Q}) be the probability space with the filtration $\mathbb{F} = (\mathcal{F}_t)_{t \in [0, T]}$. The state process $X = (X_t)_{t \in [0, T]}$ is an \mathbb{F} -adapted, continuous, Markov process taking values in a Euclidean space \mathcal{X} . For a fixed subset $\mathcal{T} \subset [0, T]$, $\vartheta = \vartheta(\mathcal{T})$ is the set of \mathbb{F} -stopping times taking values in \mathcal{T} . Then, the *optimal stopping* problem, corresponding to a given *reward function* φ , is

$$\text{to maximize } v(\tau) := \mathbb{E}[\varphi(\tau, X_\tau)], \quad \text{over all } \tau \in \vartheta, \quad (2.1)$$

where for some $a > 0$, $\varphi(t, \cdot) \in \mathcal{L}_a$ defined as

$$\mathcal{L}_a := \{\phi : \mathcal{X} \rightarrow \mathbb{R} : \text{continuous and } |\phi(x)| \leq C[1 + |x|^a] \text{ for some } C > 0\}. \quad (2.2)$$

In financial examples, a is almost always equal to one. Also, $\mathcal{T} = [0, T]$ corresponds to an American option, while \mathcal{T} is a finite set for Bermudan ones. However, since for numerical calculations one has to discretize the time variable, from the onset we assume that $\mathcal{T} = \{t_0, t_1, \dots, t_{n-1}, T\}$, with n being large for the American ones.

We next define the value function which is a central tool in Markov optimal control [24]. For $t \in \mathcal{T}$ and $x \in \mathcal{X}$, let ϑ_t be the set of all $[t, T] \cap \mathcal{T}$ -valued stopping times, and set

$$v(t, x) := \sup_{\tau \in \vartheta_t} v(t, x, \tau) \quad \text{where} \quad v(t, x, \tau) := \mathbb{E}[\varphi(\tau, X_\tau) \mid X_t = x].$$

Finally, we introduce our notation $\mathbb{R}_+ = (0, \infty)$, and

$$\mathcal{T}^\circ := \mathcal{T} \setminus \{T\} = \{t_0, t_1, \dots, t_{n-1}\}.$$

2.1 Stopping region

The following subset of the state space is the *stopping region* at time $t \in \mathcal{T}$:

$$\mathcal{S}_t := \{x \in \mathcal{X} : \varphi(t, x) = v(t, x)\}.$$

As $v(T, x) = \varphi(T, x)$ for every $x \in \mathcal{X}$, we have $\mathcal{S}_T = \mathcal{X}$.

To obtain some regularity of the value function and the stopping regions, we make the following mild assumption satisfied in all applications. Recall that \mathcal{L}_a is defined in (2.2).

Assumption 2.1. (Growth and continuity) *There exists $a > 0$, such that for every $t \in \mathcal{T}$, $\varphi(t, \cdot) \in \mathcal{L}_a$. Moreover, for every $\phi \in \mathcal{L}_a$, $t < \bar{t} \in \mathcal{T}$, $v_\phi(t, \cdot, \bar{t}) \in \mathcal{L}_a$, where*

$$v_\phi(t, x, \bar{t}) := \mathbb{E}[\phi(X_{\bar{t}}) \mid X_t = x], \quad x \in \mathcal{X}.$$

The above assumed continuity essentially follows from the continuous dependence of the distribution of the state on its initial value, and it is always satisfied in all our applications.

The following are direct consequences.

Proposition 2.2. *Under the growth and continuity Assumption 2.1, for each $t \in \mathcal{T}$, the value function $v(t, \cdot) \in \mathcal{L}_a$, and the stopping region \mathcal{S}_t is a relatively closed subset of \mathcal{X} .*

Proof. To prove the first statement, we use backward induction in time $t_0 < t_1 < \dots < t_n = T$. First observe that by Assumption 2.1, $v(T, \cdot) = \varphi(T, \cdot) \in \mathcal{L}_a$. For the induction hypothesis, suppose that $v(t_{k+1}, \cdot) \in \mathcal{L}_a$ for some $t_{k+1} \in \mathcal{T}$. By dynamic programming,

$$v(t_k, x) = \max \{\varphi(t_k, x), c(t_k, x)\}, \quad \text{where} \quad c(t_k, x) := \mathbb{E}[v(t_{k+1}, X_{t_{k+1}}) \mid X_t = x].$$

As $v(t_{k+1}, \cdot)$ is assumed to be in \mathcal{L}_a , Assumption 2.1 implies that $c(t_k, x)$ is also in \mathcal{L}_a . Since the space \mathcal{L}_a is closed under maximization, the above equation implies that $v(t_k, \cdot) \in \mathcal{L}_a$. Hence, $v(t, \cdot) \in \mathcal{L}_a$ for every $t \in \mathcal{T}$, and in particular, it is continuous.

The stopping region \mathcal{S}_t is the zero level set of $v(t, \cdot) - \varphi(t, \cdot)$ which is shown to be continuous. Hence, it is relatively closed in \mathcal{X} . \square

Because the stopping regions are closed (hence Borel) and $\mathcal{S}_T = \mathcal{X}$, the hitting time

$$\tau^* := \min \{u \in \mathcal{T} : X_u \in \mathcal{S}_u\} \quad (2.3)$$

is a well-defined stopping time and belongs to ϑ . Moreover, τ^* is optimal, or equivalently, with $v(\tau)$ as in (2.1),

$$v(\tau^*) = \sup_{\tau \in \vartheta} v(\tau).$$

2.2 Stopping boundary

In all problems that we consider, the boundary of stopping regions are given as the graphs of certain functions, but possibly in a different coordinate system. To outline the algorithm in a general manner, we first assume the existence of this geometry. We then verify it for the financial examples, in Theorem 4.1 below, under mild natural structural conditions.

Although the description of the following assumption is technical, its interpretation is quite natural. In words, we assume that there exists a stopping boundary that can be characterized as the graph of some function f^* in the appropriate coordinates (α, Ξ) . In all applications, the choice of these functions is immediate as discussed in Example 2.1 below, and in Sections 6, 7, 8, 9. In particular, the positive real-valued α is related to the payoff of the option, and f^* in (2.4) describes the stopping boundary. The function Ξ is more abstract and is a transformation into coordinates that are meaningful to the specific problem.

Assumption 2.3. (Existence of a stopping boundary) *There exist measurable functions*

$$\alpha : \mathcal{X} \rightarrow \mathbb{R}_+, \quad \Xi : \mathcal{X} \rightarrow \Xi(\mathcal{X}) \subset \mathcal{X}, \quad f^* : \mathcal{T}^\circ \times \Xi(\mathcal{X}) \rightarrow [0, \infty],$$

and $\eta \in \{-1, +1\}$, so that the map $x \in \mathcal{X} \mapsto (\alpha(x), \Xi(x)) \in \mathbb{R}_+ \times \Xi(\mathcal{X})$ is a homeomorphism (i.e., one-to-one, onto, continuous with a continuous inverse) and for every $t \in \mathcal{T}^\circ$,

$$\mathcal{S}_t = \{x \in \mathcal{X} : \eta(f^*(t, \Xi(x)) - \alpha(x)) \leq 0\}. \quad (2.4)$$

The pair $(\alpha(x), \Xi(x))$ is a representation of x that is adapted to the problem. The space $\Xi(\mathcal{X})$ is a *latent space* with (Hausdorff) dimension strictly less than that of \mathcal{X} . The parameter η determines whether the stopping region is an epigraph or a hypograph of f^* . In the financial applications, typically $\eta = 1$ corresponds to a call type option and the stopping region is an epigraph. While $\eta = -1$ is related to put options.

The surjectivity of α ensures that this characterization is non-trivial, and it additionally implies that for every $c > 0$, the point corresponding to the pair $(f^*(t, \Xi(x)) + \eta c, \Xi(x))$ is in the stopping region \mathcal{S}_t , and respectively, to the pair $(f^*(t, \Xi(x)) - \eta c, \Xi(x))$ is not in \mathcal{S}_t . Therefore, the graph of f^* in the coordinates $(\alpha(x), \Xi(x))$ given by

$$\{x \in \mathcal{X} : \alpha(x) = f^*(t, \Xi(x))\},$$

is included in the boundary of \mathcal{S}_t . In fact, in all examples it separates the stopping and continuation regions. Thus, we call f^* the optimal *stopping boundary*.

Example 2.1. In the financial problem of an American max-call option with two stocks and a given strike K , we have $\varphi(t, x) = e^{-rt}(\max\{x_1, x_2\} - K)^+$ for $x = (x_1, x_2) \in \mathcal{X}$. With Markovian dynamics, the state space is $\mathcal{X} = \mathbb{R}_+^2$. In Theorem 4.1 below, we show that there exists a stopping boundary f^* such that the triplet (α, Ξ, f^*) with

$$\alpha(x) = \max\{x_1, x_2\}, \quad \Xi(x) = \frac{x}{\alpha(x)}, \quad \eta = 1,$$

satisfies the above assumption. Figures 6, 7, 8 below provide examples of these regions. In our numerical approach, the latent variable $\Xi(x)$ is the input to the network approximating the stopping boundary. As $\Xi(x) = \Xi(\gamma x)$ for all $x \in \mathbb{R}_+^2$, $\gamma > 0$, every ray emanating from the origin is mapped to a single point by the map Ξ , and the threshold $f^*(t, \Xi(x))$ separates the stopping and continuation regions on this ray.

In the simpler example of an American put option with one Markovian stock, one simply takes $\mathcal{X} = \mathbb{R}_+$, $\alpha(x) = x$, $\Xi \equiv 0$, and $\eta = -1$. So that the stopping region is the hypograph of some stopping boundary f^* . Several other examples are discussed in the sections below. \square

As an immediate consequence of the above assumption, we can restrict the maximization in (2.1) to stopping times given by stopping boundaries. For future reference, we record this fact. Let \mathfrak{F} be the set of all measurable functions $f : \mathcal{T}^\circ \times \Xi(\mathcal{X}) \rightarrow [0, \infty]$. For $f \in \mathfrak{F}$, the corresponding stopping time is given by,

$$\tau_f := \min \{u \in \mathcal{T} : \eta(f(u, \Xi(X_u)) - \alpha(X_u)) \leq 0 \text{ or } u = T\} \in \vartheta.$$

Lemma 2.4. *Under the growth and continuity Assumption 2.1, and the existence of a stopping boundary Assumption 2.3,*

$$\sup_{\tau \in \vartheta} v(\tau) = \sup_{f \in \mathfrak{F}} v(\tau_f).$$

Proof. As $\tau_f \in \vartheta$, we have $\sup_{\tau \in \vartheta} v(\tau) \geq v(\tau_f)$ for all $f \in \mathfrak{F}$. Let f^* be as in Assumption 2.3. Then, for $u \in \mathcal{T}$, we have $X_u \in \mathcal{S}_u$ if and only if, either $\eta(f^*(u, \Xi(X_u)) - \alpha(X_u)) \leq 0$, or $u = T$. Therefore, τ_{f^*} is equal to the optimal hitting time τ^* defined in (2.3). These imply that $\sup_{\tau \in \vartheta} v(\tau) \geq \sup_{f \in \mathfrak{F}} v(\tau_f) \geq v(\tau_{f^*}) = v(\tau^*) = \sup_{\tau \in \vartheta} v(\tau)$. \square

3 The Algorithm

Our goal is to construct an efficient algorithm for the calculation of the interpretable stopping decisions based on boundaries approximated by artificial neural networks. We train the networks by the *deep empirical risk minimization* algorithm proposed by Weinan E, Jiequn Han, and Arnulf Jentzen [30, 31].

This method approximates the expected value of a reward function by Monte-Carlo simulations and optimizes it by stochastic gradient ascent. For the problem of optimal stopping, the natural choice for the reward function is $\varphi(\tau_\theta, X_{\tau_\theta})$. However, the map $\theta \mapsto \tau_\theta$ is piece-wise constant, making optimization by gradient ascent impossible. Therefore, we replace the hitting times by stopping probabilities based on fuzzy boundaries that are also used in some problems of solidification [11, 25, 48, 50].

3.1 Fuzzy Stopping Boundary

Let $\eta \in \{-1, 1\}$ be as in Assumption 2.3. For $f \in \mathfrak{F}$, set

$$d(t, x; f) := \eta(f(t, \Xi(x)) - \alpha(x)), \quad t \in \mathcal{T}^\circ, x \in \mathcal{X}.$$

In the original problem, once a boundary f is chosen, we stop the process only when $d(t, X_t; f) \leq 0$. As training is not possible with this sharp “stop-or-continue” rule, we introduce the fuzzy boundaries. Namely, we fix a *tuning parameter* $\epsilon > 0$, and at time $t \in \mathcal{T}^\circ$, define

$$\mathfrak{B}_t^\epsilon(f) := \{x \in \mathcal{X} : |d(t, x; f)| < \epsilon\}$$

be the *fuzzy region*.

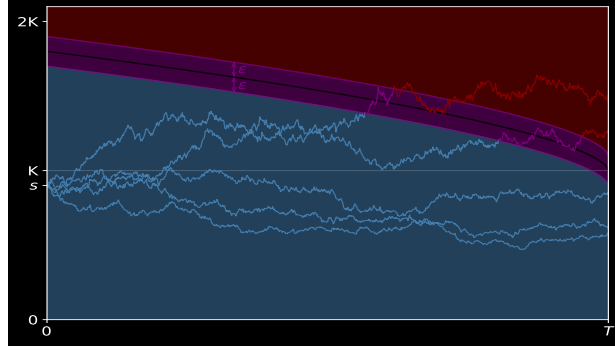


Figure 1: Fuzzy region of an American call (in purple) connecting the stopping region (in red) to the continuation region (in blue).

An illustration of the fuzzy region is given in Figure 1 for an American call option. If the state is not in $\mathfrak{B}_t^\epsilon(f)$, we continue or stop, as before, with probability one. But if we are in the fuzzy region, we stop with probability $p_t(X_t, f)$, where

$$p_t(x, f) := \left(\frac{\epsilon - d(t, x, f)}{2\epsilon} \right)^+ \wedge 1, \quad t \in \mathcal{T}^\circ, f \in \mathfrak{F}, x \in \mathcal{X}.$$

At maturity, we stop at all states, or equivalently, we set $p_T(\cdot, f) \equiv 1$. Precisely, $p_t(X_t, f)$ is the probability of stopping at time $t \in \mathcal{T}$, conditioned on the event that the process is not stopped prior to t . The relaxed control problem is then defined using the reward function given by

$$\mathcal{R}_\epsilon(X, f) := \sum_{t \in \mathcal{T}} p_t(X_t, f) b_t(X, f) \varphi(t, X_t), \quad (3.1)$$

where $b_t(X, f)$ is the probability of not stopping strictly before time t and it is obtained as the solution of the following difference equations,

$$b_{t+1}(X, f) = b_t(X, f)(1 - p_t(X_t, f)), \quad t \in \mathcal{T}^\circ,$$

with $b_0(X, f) = 1$. In other words, $b_t(X, f)$ is the unused *stopping-budget* remaining. The process $b_t(X, f)$ is non-increasing in time and takes values in $[0, 1]$. Connection between $\mathbb{E}[\mathcal{R}_\epsilon(X, f)]$ and $v(\tau_f)$ is discussed in Appendix B.

We refer to [43] for a different exposition of this relaxation.

3.2 Deep ERM

In this method, we approximate the stopping boundary by a set of deep neural networks that we abstractly parameterize by θ taking values in a finite-dimensional space Θ as follows:

$$\mathfrak{N} = \{ g(\cdot; \theta) : \mathcal{T}^\circ \times \Xi(\mathcal{X}) \rightarrow \mathbb{R}_+ : \theta \in \Theta \} \subset \mathfrak{F}.$$

Let $\tau_\theta = \tau_{g(\cdot; \theta)}$ be the hitting time corresponding to the stopping boundary $g(\cdot; \theta)$. The following is the pseudocode of the deep empirical risk minimization based on the fuzzy stopping boundary.

1. *Initialize* $\theta \in \Theta$.
2. *Simulate* independent state trajectories, $\{X^{(1)}, \dots, X^{(B)}\}$, where B is the *batch size*.

3. Compute the *empirical reward* function:

$$R_\epsilon(\theta) := \frac{1}{B} \sum_{i=1}^B \mathcal{R}_\epsilon(X^{(i)}, g(\cdot; \theta)),$$

where $\mathcal{R}_\epsilon(X, f)$ is as in (3.1). A modification $\mathcal{R}_{\epsilon, \lambda}$, given in (4.4), allows for *important sampling* discussed in Section 4.1 below.

4. *Optimize and update by stochastic gradient ascent* with the learning rate process ζ :

$$\theta \leftarrow \theta + \zeta \nabla_\theta R_\epsilon(\theta).$$

5. *Stop* after M number of iterations.

6. *Compute the initial price* given by the training network, using J many Monte-Carlo simulations with a sharp boundary.

A formal discussion of the algorithm is given in the next subsection. In Appendix C, we also provide more details of the above algorithm for the benefit of readers interested in the implementation. [26] proposes a similar approach by assuming that the stopping region has a finite dimensional representation without a formal justification. Also, Chapter 8.2 of [28] discusses a general approximation for the value calculation by parametrizing the stopping times. In our approach we approximate the stopping boundary directly.

3.3 Discussion

The general theory of stochastic gradient ascent [44] implies that the above algorithm constructs an approximation of

$$\theta_\epsilon^* \in \arg \max_{\theta \in \Theta} \mathbb{E}[\mathcal{R}_\epsilon(X, g(\cdot; \theta))].$$

In the context of optimal stopping, [10] also provides rates of convergence. Moreover, in Lemma B.1 below, we show that for each $\theta \in \Theta$,

$$\lim_{\epsilon \downarrow 0} \mathbb{E}[\mathcal{R}_\epsilon(X, g(\cdot; \theta))] = \mathbb{E}[\varphi(\tau_\theta, X_{\tau_\theta})] = v(\tau_\theta).$$

Therefore,

$$\mathbb{E}[\mathcal{R}_\epsilon(X, g(\cdot; \theta_\epsilon^*))] = \sup_{\theta \in \Theta} \mathbb{E}[\mathcal{R}_\epsilon(X, g(\cdot; \theta))] \approx \sup_{\theta \in \Theta} v(\tau_\theta).$$

Also, by the celebrated universal approximation theorem [20, 32], continuous stopping boundaries are well approximated by the hypothesis class \mathfrak{N} , and in all our applications the optimal stopping boundaries are continuous. Thus, in view of this approximation capability of \mathfrak{N} and Lemma 2.4, we formally have

$$\mathbb{E}[\mathcal{R}_\epsilon(X, g(\cdot; \theta_\epsilon^*))] = \sup_{\theta \in \Theta} \mathbb{E}[\mathcal{R}_\epsilon(X, g(\cdot; \theta))] \approx \sup_{\theta \in \Theta} v(\tau_\theta) \approx \sup_{f \in \mathfrak{F}} v(\tau_f) = \sup_{\tau \in \mathcal{D}} v(\tau).$$

Hence, the above algorithm constructs an artificial neural network with an asymptotic performance formally close to the optimal value, justifying the algorithm. However, there are several technical issues in front a rigorous statement and we postpone this convergence analysis to a future study.

4 Financial Examples

We numerically study several American or Bermudan options to illustrate and also assess the proposed algorithm. In these applications, \mathbb{Q} is a risk-neutral measure and the Markov state process $X = (S, Y, Z)$ has three components: S is the prices of the underlying stocks possibly including the past values to make it Markov, a factor process Y used in models like Heston, and a functional Z of the stock process, making the pay-off of a path-dependent option a function of the current value of X . Typical examples of Z are the running maximum or minimum in look-back options, and the average stock price for the Asian options. Depending on the model and the option, Y , Z , or neither may be present.

In our examples, the pay-off function always has the following form:

$$\varphi(t, x) = e^{-rt} (\eta(\alpha(s, z) - \beta(s, z) - K))^+, \quad t \in \mathcal{T}, \quad x = (s, y, z) \in \mathcal{X}, \quad (4.1)$$

where $\alpha > 0, \beta \geq 0$ are *positively homogenous of degree one*, and the parameter $\eta \in \{-1, +1\}$ determines the option type: $\eta = 1$ is a call and $\eta = -1$ is a put. The choice of α and β may not be unique, and we would choose $\beta \equiv 0$ when possible. It is clear that φ grows linearly. Additionally, in all our examples, the stock price distribution depends smoothly on the initial data. Hence, the growth and continuity Assumption 2.1 holds with $a = 1$.

The following stopping boundary result is proven in the Appendix. For max-call options a similar result is proved in Proposition A.5 in [13], and for look-back options in [21].

Theorem 4.1. *Suppose that in (4.1), $\alpha : \mathcal{X} \rightarrow \mathbb{R}_+, \beta : \mathcal{X} \rightarrow [0, \infty)$ are positively homogenous of degree one, and that X scales linearly in the initial values $S_t = s, Z_t = z$, i.e., for $x = (s, y, z)$, $\gamma > 0$, and a bounded, continuous, integrable function ϕ , the following holds:*

$$\mathbb{E}[\phi(u, X_u) \mid X_t = (\gamma s, y, \gamma z)] = \mathbb{E}[\phi(u, \gamma X_u) \mid X_t = (s, y, z)].$$

Further assume that the growth and continuity Assumption 2.1 holds, and the price of the European option is strictly positive for all $t < T$. Then, the existence of a stopping boundary Assumption 2.3 is satisfied with α as in (4.1),

$$\Xi(s, y, z) = \left(\frac{s}{\alpha(s, z)}, y, \frac{z}{\alpha(s, z)} \right), \quad (4.2)$$

and

$$f^*(t, \xi) := \begin{cases} \sup \{ \alpha(s, z) : x \in \mathcal{S}_t \text{ s.t. } \Xi(x) = \xi \}, & \text{if } \eta = -1, \\ \inf \{ \alpha(s, z) : x \in \mathcal{S}_t \text{ s.t. } \Xi(x) = \xi \}, & \text{if } \eta = 1, \end{cases} \quad (4.3)$$

where we use the convention that the sup over the empty set is zero and inf is infinity.

The positivity of the European price ensures that on stopping regions $\varphi > 0$.

4.1 Importance sampling

To effectively train the network $g(\cdot; \theta)$ by deep ERM, it is crucial that the simulated paths X_t hit the fuzzy boundary $\mathfrak{B}_t^i(g(\cdot; \theta))$ frequently enough for each exercise date. We achieve this by importance sampling that we now outline.

In all our examples, S evolves according to a stochastic differential equation,

$$dS_t = S_t (\mu_t dt + \sigma_t dW_t), \quad s_0 \in \mathbb{R}^d,$$

where the processes $\mu \in \mathbb{R}^d, \sigma \in \mathbb{R}^{d \times d}$ may depend on t, S_t and other exogenous factors. For $\lambda = (\lambda_1, \dots, \lambda_d) \in \mathbb{R}^d$, let \mathbb{Q}_λ be the measure obtained by the Girsanov transformation so that $W_t^\lambda := W_t + \lambda t$ is a \mathbb{Q}_λ Brownian motion. Hence, under \mathbb{Q}_λ ,

$$dS_t = S_t ([\mu_t - \sigma_t \lambda] dt + \sigma_t dW_t^\lambda).$$

We then adjust the hitting frequency of the simulated trajectories, by choosing λ appropriately, thus allowing for more efficient training. Although in all our experiments we have employed constant λ , one could also use time dependent ones as well.

However, as the probability measure is modified, we need to account for it in the reward function. This is achieved by the Radon-Nikodym derivative given by,

$$\mathcal{Z}_t^\lambda(S) := \frac{d\mathbb{Q}}{d\mathbb{Q}_\lambda} \Big|_{\mathcal{F}_t} = \exp(\lambda \cdot W_t^\lambda - \frac{1}{2}|\lambda|^2 t).$$

Since W^λ can be expressed in terms of S , the process \mathcal{Z}^λ can be viewed as a function of the state process S . We then modify the reward function given in (3.1) by,

$$\mathcal{R}_{\epsilon,\lambda}(X, f) := \sum_{t \in \mathcal{T}} \mathcal{Z}_t^\lambda(S) p_t(X_t, f) b_t(X, f) \varphi(t, X_t). \quad (4.4)$$

5 Network Architecture and Parameters

All our numerical experiments have been carried out with Tensorflow 2.7 on a 2021 Macbook pro with 64GB unified memory and Apple M1 Max chip. The code is implemented in Python and run on CPU (10-core) only. Throughout the examples, we use the same neural network architecture and training parameters, given below:

1. We use only one deep neural network that takes both the time and the state vector as input. Then, the trained boundary is a function of the continuous time variable, delivering smoother dependence on time. This is in contrast to majority of previous studies that use one deep neural network for each time in the discrete set \mathcal{T} .

2. The single feedforward neural network that we employ consists of 2 hidden layers with leaky rectified linear unit (Leaky ReLU) activation function. Each layer has $20 + \bar{d}$ many nodes where \bar{d} is the dimension of the latent space $\Xi(\mathcal{X})$. The output layer uses the standard rectified linear unit (ReLU) activation. An important parameter is the *bias* $\theta_0 \in \mathbb{R}_+$ in the *output layer* that sets an initial boundary level. Namely, the output function g is given by,

$$g(t, \xi; \theta_1, \theta_0, \theta_{in}) = (\theta_1 \cdot g_{in}(t, \xi; \theta_{in}) + \theta_0)^+,$$

where $g_{in}(t, \xi; \theta_{in}) \in \mathbb{R}^{20+\bar{d}}$ is the value of the neural network before entering the output layer, with parameters θ_{in} . In our experiments, we have observed that the algorithm is not sensitive to the initial bias θ_0 , as long as it is set to a reasonable value. For example, for options with a given strike K , we choose $\theta_0 = \frac{3K}{2}$, $\theta_1 = \frac{K}{2}$ for calls and puts, respectively.

3. We set the number of stochastic gradient iterations to $M = 3,000$ with a fixed simulation batch size of $B = 512$.

4. The learning rate process ζ is taken from the Adam optimizer [35].

5. The fuzzy region width ϵ is chosen to be the strike K times the standard deviation of the increments of $\alpha(X)$. While the strike K adjusts to the scale of the payoff, the second term reflects the typical variation of the process $\alpha(X)$ between exercise dates.

5. After the training is completed, the initial price is computed using the sharp boundary and $J = 2^{22} = 4,194,304$ many Monte Carlo simulations.

6 One-dimensional Put Option

As an initial study, we consider an at the money Bermudan put option in the Black Scholes model. Hence, $\varphi(t, s) = e^{-rt}(K - s)^+$, $\eta = -1$, and $X_t = S_t$ with $\mathcal{X} = \mathbb{R}_+$.

We take the same model and parameters as in Section 4.3.1.2 of [8]. Namely, S_t follows the Black-Scholes model with parameters

$$s_0 = K = 40, \quad r = 6\%, \quad \sigma = 40\%, \quad T = 1, \quad n = 50, \quad (6.1)$$

so that $\mathcal{T} = \{0, \frac{1}{50}, \dots, 1\}$. For efficient training, we want the the stock price to reach low values to cross the boundary. Towards this goal, we use importance sampling with $\lambda = 0.275$, so that S becomes a super-martingale under \mathbb{Q}_λ with 5% negative drift. The fuzzy region width is approximately equal to $\epsilon = K\sigma\sqrt{T/n}$.

It is classical that the existence of a stopping boundary Assumption 2.3 holds with $\alpha(s) = s$ and $\Xi(s) \equiv 0$. Table 1 summarizes the performance values and compares it to the values computed by Becker *et.al.* [6]. The first column in that table is the average price of ten experiments, and the second column the highest price. In practice, one would of course retain the trained boundary yielding the highest price.

Bermudan with parameters as in (6.1)

Average Price	Highest Price	Runtime	Price in [8]
5.308 (0.003)	5.313 (0.003)	57.9	5.311

Table 1: Bermudan Put Option with $n = 50$ in Black-Scholes model. First column is the average of ten experiments with its standard deviation in brackets. The second column is the highest price among the realizations and its standard deviation in brackets. Third column is the average runtime (in seconds) per experiment for the training phase.

Figure 2 displays the initial boundary and the trained one. The optimal boundary, shown in blue, is computed using a finite-difference scheme. Figure 3 is the approximation of the optimal stopping boundary for the American option with same parameters. As the American option does not restrict the trading dates, we use $n = 250$. Also, to capture the known singular behavior of the optimal boundary, we use a non-uniform discretization that is finer near maturity.

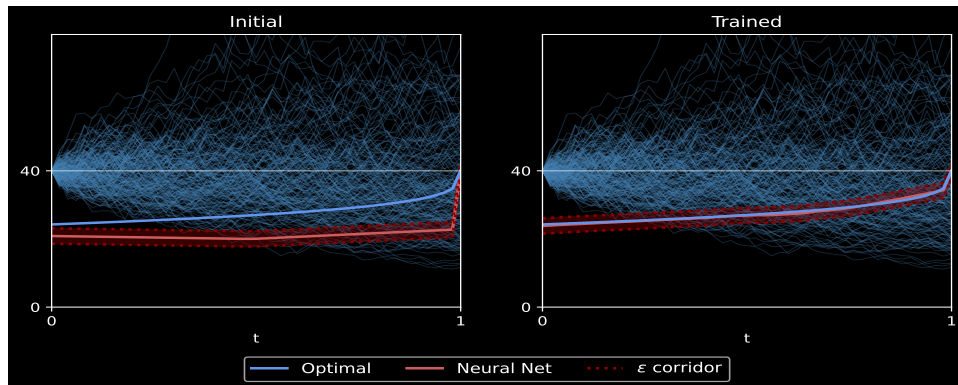


Figure 2: Black-Scholes model with parameters as in (6.1). Blue trajectories are simulated paths.

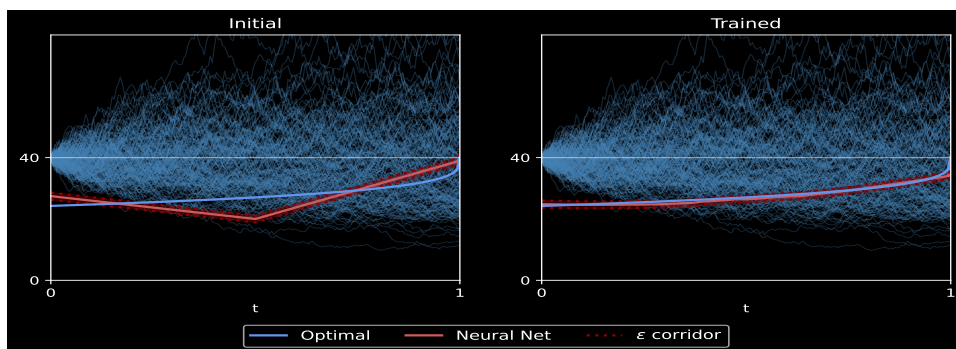


Figure 3: Almost American option in Black-Scholes model with parameters as in (6.1) except $n = 250$.

7 Put Option in the Heston Model

Consider the Bermudan put option of the Section 6 (i.e., $s_0 = K = 40$, $T = 1$, $n = 50$) but now in the Heston model. Then, under a risk-neutral measure \mathbb{Q} , the stock and factor dynamics are given by,

$$\begin{aligned} \frac{dS_t}{S_t} &= rdt + \sqrt{Y_t} dW_t, \\ dY_t &= (\kappa(\bar{y} - Y_t) - \gamma_{\mathbb{Q}} Y_t)dt + \sigma_Y \sqrt{Y_t} d\tilde{W}_t, \end{aligned}$$

where W, \tilde{W} are Brownian motions with constant correlations ρ , and $\gamma_{\mathbb{Q}}$ is the volatility risk premium related to the risk-neutral measure \mathbb{Q} . We use the parameters

$$r = 6\%, \quad \kappa = 1, \quad \sqrt{Y_0} = \sqrt{\bar{y}} = 40\%, \quad \gamma_{\mathbb{Q}} = 0, \quad \sigma_Y = 0.1, \quad \rho = -0.5.$$

In particular, the Feller condition $2\kappa\bar{y} \geq \sigma_Y^2$ is satisfied. To allow for a comparison with results of Section 6, we set the initial value and the long term mean $\sqrt{\bar{y}}$ of the stochastic volatility \sqrt{Y} equal to the Black-Scholes volatility σ from that section. Moreover, we use importance sampling with the same Girsanov parameter, $\lambda = 0.275$. The variance process is simulated using the Milstein scheme [36] with $n' = 4n = 200$ time steps.

In this example, $X = (S, Y)$, $\mathcal{X} = \mathbb{R}_+^2$, $\alpha(x) = s$, $\beta \equiv 0$, and $\Xi(x) = (1, y)$. We simplify and set Ξ to y . Figure 4 shows that the stopping boundary is a function of time t and the spot variance $Y_t = y$. As φ is convex and bounded, the value function $v(t, s, y)$ is non-decreasing in y for all $t \in [0, T]$, cf. [38]. Consequently, the map $y \mapsto f(t, y)$ is non-increasing and the trained neural network $g(\cdot, \cdot; \theta^M)$ has the same property as seen in Figure 4. Notice also that $y \mapsto f(t, y)$ becomes less steep as t approaches maturity. This is line with the fact that the vega of the option decreases over time until the option is exercised. Since $s \mapsto v(t, s, y)$ is non-increasing for fixed (t, y) , we therefore expect that the rectangle $[0, s] \times [0, y]$ is contained in \mathcal{S}_t whenever $(s, y) \in \mathcal{S}_t$. This is also confirmed in Figure 4.

Table 2 compares the prices of the claims when the stopping decision exploits the spot variance (i.e., $f = f(t, y)$) and when it does not (i.e., $f = f(t)$). We observe that the price obtained with the “blind” strategy is only slightly inferior. Interestingly, the neural network coincides with the optimal boundary in the Black-Scholes model. Also note that the longer runtime compared to Table 1 is due to the thin partition needed to simulate Y .

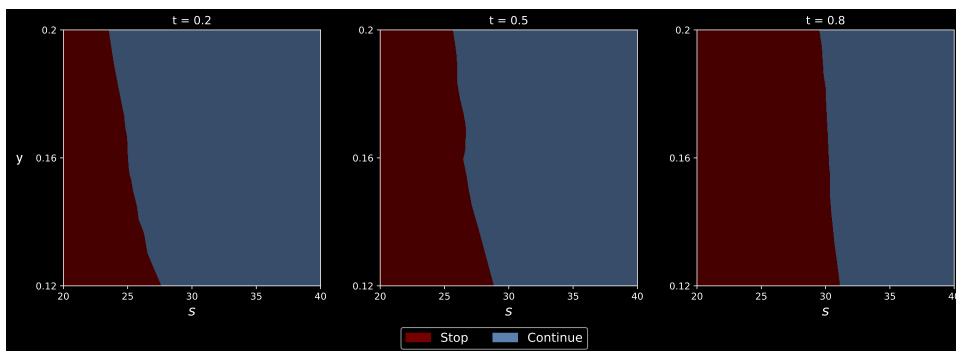


Figure 4: Stopping and continuation regions for a Bermudan put in the Heston model.

Bermudan Put Option ($n = 50$), Heston model

Threshold	Average Price	Highest Price	Runtime
$f(t, y)$	5.290 (0.004)	5.294 (0.003)	194.3
$f(t)$	5.288 (0.004)	5.293 (0.003)	193.9

Table 2: Second column is the average of ten experiments with its standard deviation in brackets. Third column is the highest price among the realizations and its standard deviation in brackets. Fourth column is the average run-time (in seconds) per experiment for the training phase.

Finally, Figure 5 displays a trajectory of (X, Y) together with the trained stopping boundary $g(\cdot; \theta^M)$. As can be seen, the threshold is typically below the Black-Scholes boundary when Y_t is above its mean $\bar{y} = 0.16$ and vice versa.

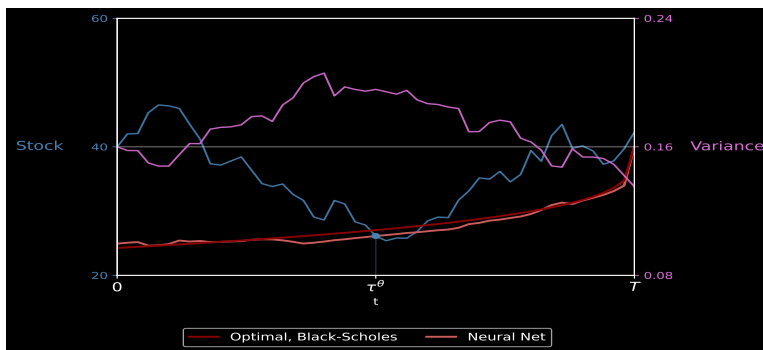


Figure 5: Stock price, variance and threshold process for a Bermudan put in the Heston model.

8 Bermudan Max-Call

We now apply the deep empirical risk minimization to the classical example of the max-call option on dividend paying stocks as studied in the seminal paper by Broadie & Detemple [13], and also in the book by Detemple [22].

We assume that the stock price process $S_t = (S_t^{(1)}, \dots, S_t^{(d)}) \in \mathbb{R}_+^d$ is the state process and solves

$$dS_t^{(i)} = S_t^{(i)} \left((r - \delta_i)dt + \sigma_i dW_t^{(i)} \right),$$

where δ_i is the dividend rate. In this example, the reward function is given by,

$$\varphi(t, s) = e^{-rt}(\alpha(s) - K)^+, \quad \text{where } \alpha(s) = \max\{s_1, \dots, s_d\}, \quad s \in \mathbb{R}_+^d.$$

Hence, $X = S$, $\mathcal{X} = \mathbb{R}_+^d$, $\eta = 1$, and $\beta \equiv 0$. In view of Theorem 4.1, the existence of a stopping boundary Assumption 2.3 holds with α and $\Xi(s) = s/\alpha(s)$. In two dimensions, the stopping regions can be visualized effectively as seen in the Figures 6, 7 reported from [43]. These are stopping regions in two space dimensions obtained with initial data $s_0 = 90$ and $s_0 = 100$ with parameters given in (8.1). Clearly the stopping boundary is independent of the initial condition and the below numerical results verify it. Also they are very similar to those obtained in [13].

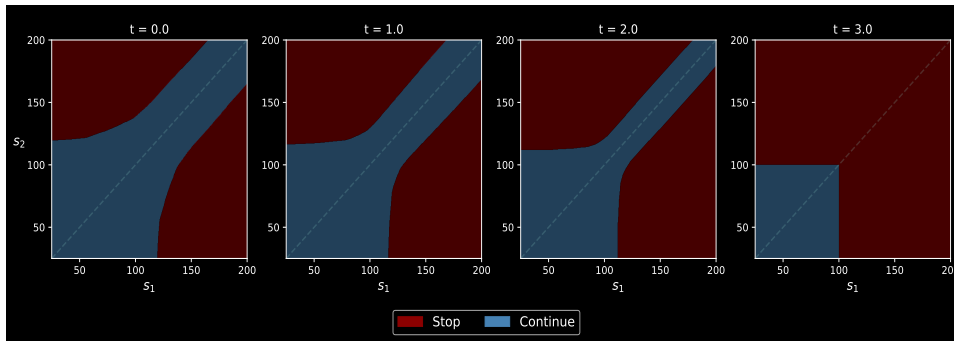


Figure 6: Stopping Boundaries with $s_0 = 90$, $K = 100$, $\sigma_i = 0.2$, $r = 0.05$, $\delta_i = 0.1$, reported from [43].

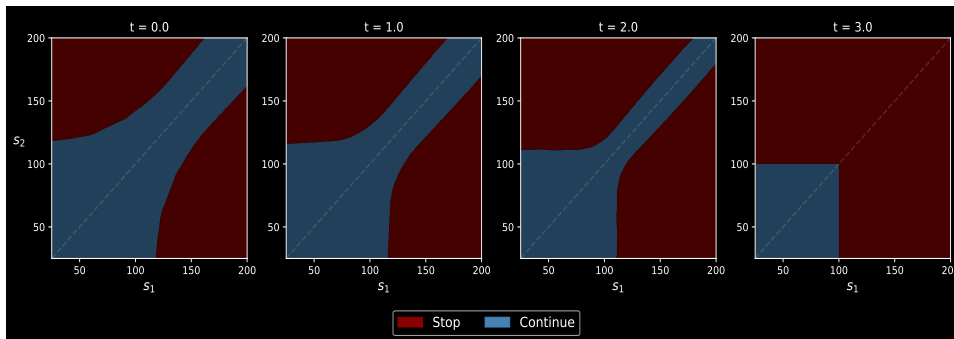


Figure 7: Stopping Boundaries with $s_0 = 100$, $K = 100$, $\sigma_i = 0.2$, $r = 0.05$, $\delta_i = 0.1$, reported from [43].

8.1 Numerical results: Symmetric case

As in [6, 8], we use the following parameter set:

$$K = 100, s_0 = 90, 100, 110, \sigma_i = 0.2, r = 0.05, \delta_i = 0.1. \quad (8.1)$$

We take the maturity to be 3 years and $n = 9$. Thus, each time interval corresponds to four months. Numerical experiments for these parameters are reported in the accompanying paper [43]. They compare well to the results obtained in [6, 8].

We also compute the initial price of this max option on $d \in \{2, 5, 10, 20, 50\}$ assets. Since we observe that the drift of the maximum stock price process is of order $\ln(d)$, we use important sampling with $\lambda_i = (r - \delta_i - \mu(d))/\sigma$ and $\mu(d) = -0.01 \ln(d)$ to ensure that the maximum stock value does not cross the stopping boundary too soon. Utilizing the fact that the stocks are exchangeable in this example, we order the stock prices and input the top five assets to the network. Table 3 summarizes the results. Despite a low cutoff, the obtained prices are close or above the benchmark. Moreover, the runtime remains moderate as d increases. The case $d = 5$ is a classical example first introduced by Broadie and Glasserman [14] and well-studied later in the literature. Table 3 also contains the tight confidence intervals, obtained in [1], [12], and [6] using a primal-dual method.

Max-call option with parameters in (8.1) and $s_0 = 100$

d	Average Price	Highest Price	Runtime	Price in [8]	Confidence Intervals
2	13.883 (0.009)	13.898 (0.008)	29.1	13.901	[13.892, 13.934] ([1])
5	26.130 (0.010)	26.151 (0.009)	31.8	26.147	[26.115, 26.164] ([12])
10	38.336 (0.015)	38.355 (0.011)	33.1	38.272	[38.300, 38.367] ([6])
20	51.728 (0.018)	51.753 (0.011)	36.0	51.572	[51.549, 51.803] ([6])
50	69.860 (0.012)	69.881 (0.011)	43.5	69.572	[69.560, 69.945] ([6])

Table 3: Max Option on $d \in \{5, 10, 20, 50\}$ symmetric assets. The second column is the average of ten experiments with its standard deviation in brackets. The third column is the highest price among the realizations and its standard deviation in brackets. The fourth column is the average runtime (in seconds) per experiment for the training phase.

8.2 Numerical results: Asymmetric case

We next investigate the Bermudan max option with asymmetric assets. We consider two stocks with same volatility σ but different dividend rates. We use the same parameters for K, r, σ as in the previous subsection, and $s_0 = 100, \delta = (5\%, 15\%)$.

We choose the parameter λ in the Girsanov theorem, so as to make the assets symmetric under an equivalent measure. More precisely, we choose $\lambda = [(r + \mu(2))(1, 1) - \delta]/\sigma \in \mathbb{R}^2$ so that both assets have the same drift $\mu(2) \in \mathbb{R}$ under \mathbb{Q}_λ , and $\mu(2) = -0.01 \ln(2)$ is as in the previous subsection. The asymmetry is then captured by the Radon-Nikodym derivative appearing in the reward function.

The average and highest price after 10 realizations are given in Table 4. The last column is a benchmark price obtained from the Least Square Monte Carlo (LSMC) algorithm [39] with 2^{22} simulations. As can be seen, our method in average computes a price comparable to the benchmark, but may construct better stopping strategies reported in the second column. Notice also that having different dividends gives a premium over the symmetric case as shown in the first row of Table 3.

Max-call option with asymmetric assets

Average Price	Highest Price	Runtime	LSMC Price
15.551 (0.014)	15.575 (0.011)	30.9	15.558 (0.009)

Table 4: Max Option on $d = 2$ asymmetric assets

Figure 8 displays the stopping and continuation region over time. In particular, we clearly see the asymmetry of the problem and the stopping region reflected in the figure. For $i = 1, 2$, let the connected components of \mathcal{S}_t be given by, $\mathcal{S}_t^{(i)} := \{s \in \mathcal{S}_t : \alpha(s) = s_i\}$. In Figure 8, we observe that the light red region $\mathcal{I}(\mathcal{S}_t^{(1)})$ with $\mathcal{I}(s_1, s_2) := (s_2, s_1)$, is contained in $\mathcal{S}_t^{(2)}$. That is, if it is optimal to stop at $S_t = (s_1, s_2)$, then the same is true at $\mathcal{I}(S_t) = (s_2, s_1)$. The structure of the asymmetry is therefore consistent with our expectations.

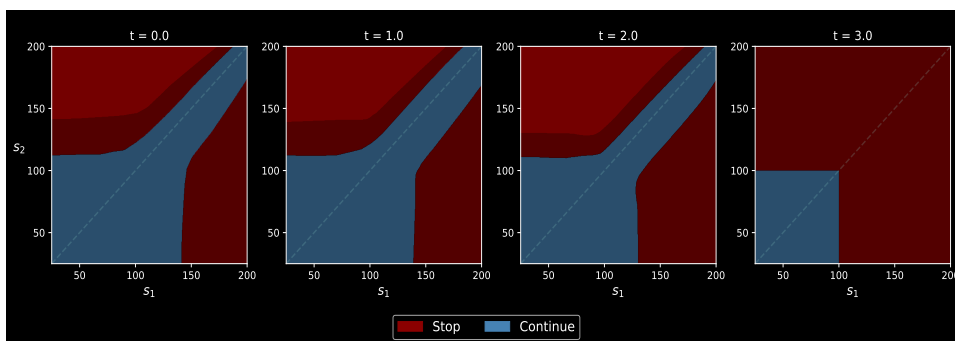


Figure 8: Stopping and continuation regions for a max option with $d = 2$ assets and $T = 3$. The light red region is the reflection of the lower stopping region $\{s \in \mathcal{S}_t : s_1 \geq s_2\}$ through the diagonal.

Remark 8.1. For max-call options on two assets, [26] also use a two-dimensional parametrization of the stopping region given by

$$\mathcal{S}_t \approx \left\{ (s_1, s_2) \in \mathbb{R}_+^2 : \alpha(s) = \max\{s_1, s_2\} \geq \theta_t^1, \text{ and } |s_1 - s_2| \geq \theta_t^2 \right\}.$$

While the above hypothesis class of regions provides satisfactory numerical results, they are restricted to two dimensions, and only give a simple polygonal approximation of the actual stopping regions that are more complex as can be seen in the Figures 6, 7, 8.

9 Look-back Options

Look-back options provides exposure to the minimum or maximum values attained during the tenure of the option. There are several types of look-back options that are commonly traded as summarized in the Table 5. In these examples, when the stock price $S_t \in \mathbb{R}_+$ is modeled as a Markov process, then we would have $X = (S, Z)$, where Z_t is either the *running maximum* \bar{S}_t or *running minimum* \underline{S}_t given by,

$$\bar{S}_t := \max_{u \in [0, t]} S_u, \quad \underline{S}_t := \min_{u \in [0, t]} S_u.$$

Option	$e^{rt}\varphi(t, X_t)$	η	$\alpha(s, z)$	$\beta(s, z)$
Fixed Strike Call	$(\bar{S}_t - K)^+$	1	z	0
Fixed Strike Put	$(K - \underline{S}_t)^+$	-1	z	0
Floating Strike Call	$(S_t - \gamma \underline{S}_t)^+$	1	s	γz
Floating Strike Put	$(\bar{S}_t - S_t)^+$	-1	s	γz

Table 5: Different Lookback Options.

Clearly, $X = (S, Z)$ is a Markov process with $\mathcal{X} = \{x = (s, z) \in \mathbb{R}_+^2 \mid 0 < s \leq z\}$, when $Z_t = \bar{S}_t$, or $\mathcal{X} = \{x = (s, z) \in \mathbb{R}_+^2 \mid 0 < z \leq s\}$ when $Z_t = \underline{S}_t$. The scaling factor γ appearing in the payoff of floating strike options is introduced to reduce the price of these otherwise expensive contracts. We therefore choose $\gamma \in [1, \infty)$ and $\gamma \in (0, 1]$ for call and put options, respectively. When $\gamma = 1$, the floating strike look-back put (call) option delivers precisely the drawdown (drawup) of the stock. We refer the reader to Dai and Kwok [21] for an introduction and discussion of American look-back claims.

9.1 Fixed strike call

Although our setting covers all look-backs, as an example we only consider the fixed strike call. Then, $Z_t = \bar{S}_t$, $\mathcal{X} = \{x = (s, z) \in \mathbb{R}_+^2 \mid 0 < s \leq z\}$, $\eta = 1$, $\alpha(s, z) = z$. By Theorem 4.1, a stopping boundary exist with $\Xi(x) = (s/z, 1)$. Notice that $1 - (s/z) = (z - s)/z$ is the relative drawdown of the stock.

As in [21], we use

$$T = 1/2, \quad r = 2\%, \quad \delta = 4\%, \quad \sigma = 30\%. \quad (9.1)$$

To approximate the American option, we consider a regular partition with $n = 200$ many exercise dates. Moreover, we employ an even thinner partition with $n' = 800$ points for our simulations. This allows us to better approximate the running maximum. Importance sampling is not needed in this example. We visualize the the stopping region in the (s, y) plane, similar to Fig. 3 in [21], where the authors use a finite difference scheme to compute the initial price.

Average Price	Highest Price	Runtime	European Price	Upper Bound
16.827 (0.008)	16.844 (0.007)	216.7	16.808	16.979

Table 6: American look-back call with parameters in (9.1) and $K = 100$.

Table 6 summarizes the result for $s_0 = K = 100$. First column in that table is the average of ten experiments with the standard deviation in brackets. The second column is the highest price among the realizations and its standard deviation in brackets. The third column is the average runtime (in seconds) per experiment for the training phase. The last column is the upper bound given by the forward value of the European price. The European price, denoted by v_e gives a lower bound on the American price. The last column is the forward value of European price, $e^{rT}v_e$, is an upper bound for the American price. Since here $e^{rT} \approx 1.01$,

the American option only offers little premium over its European counterparts. This is nevertheless fortunate from a performance perspective as it gives a tight interval in which the American price must lie (see, e.g., [19]). We indeed obtain prices that are within the bounds.

We display the stopping and continuation region in Figure 9. Obtaining an overall accurate boundary turns out to be challenging, as visiting all the pairs $(s, y) \in \mathcal{X}$ is difficult, especially when $s \ll y$. We effectively resolve this issue by randomizing $s_0 \in \mathcal{X}$ and setting the initial running maximum to $y_0 = K \vee s_0$. As in Fig. 3 of [21], we observe a flat boundary when $s \ll K$. That is, the neural network correctly set $g(t, s/y; \theta)$ approximately equal to the strike when the ratio s/y is small, or equivalently, when the relative drawdown is large.

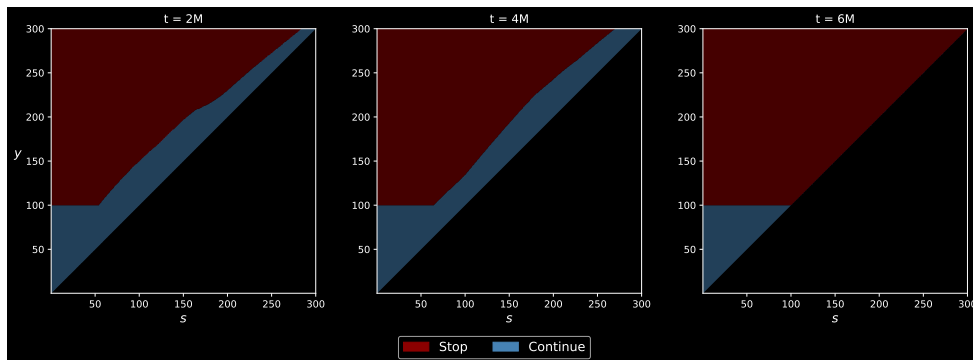


Figure 9: Stopping boundaries for a look-back call with parameters in (9.1) and $K = 100$.

10 Conclusion

We have developed an algorithm for the computation of the “graph-like” stopping boundaries separating the continuation and stopping regions in optimal stopping. While the method at the high-level is empirical risk minimization, a relaxation based on fuzzy boundaries motivated from phase-field models for liquid-to-solid phase transitions is used. Through numerical experiments, the method is shown to be effective in high-dimensions. The method has the potential to incorporate market details like price impact, transaction costs, and market restrictions. It is also possible to apply the technique to other models from financial economics and other obstacle type problems as long as a control representation is available.

A Proof of Theorem 4.1

We prove the theorem when the parameter η in (4.1) is equal to -1 . The other case of $\eta = 1$ is proved *mutatis mutandis*. We fix $t \in \mathcal{T}^\circ$.

Firstly, it is clear that for Ξ as in (4.2), the map $x = (s, y, z) \mapsto (\alpha(s, z), \Xi(x))$ is a homeomorphism and α is onto.

Recall that ϑ is the set of all \mathcal{T} -valued stopping times, $\vartheta_t = \vartheta \cap [t, T]$, the pay-off $\varphi(t, x) = \varphi(t, s, z)$ is given by (4.1), and the European price, v_e , is strictly positive. As the European price is equal to price of the stopping time $\tau \equiv T$, $v(t, x) \geq v_e > 0$. Therefore, the stopping region is characterized by

$$x \in \mathcal{S}_t \quad \Leftrightarrow \quad v(t, x) = \varphi(t, s, z) > 0.$$

We start by showing that \mathcal{S}_t is star-shaped around the origin.

Lemma A.1. *If $x = (s, y, z) \in \mathcal{S}_t$, then $(\gamma s, y, \gamma z) \in \mathcal{S}_t$ for every $\gamma \in (0, 1]$.*

Proof. Fix $x = (s, y, z) \in \mathcal{S}_t$ and $\gamma \in (0, 1]$. Then, by the definition of the stopping region \mathcal{S}_t , $v(t, x) = \varphi(t, s, z) > 0$. In particular, as $\eta = -1$, $\alpha(s, z) - \beta(s, z) < K$ and therefore,

$$\alpha(\gamma s, \gamma z) - \beta(\gamma s, \gamma z) = \gamma(\alpha(s, z) - \beta(s, z)) < \gamma K \leq K.$$

Hence, $\varphi(t, \gamma s, \gamma z) = -(\alpha(\gamma s, \gamma z) - \beta(\gamma s, \gamma z) - K) > 0$. Moreover,

$$\mathbb{E}[\varphi(\tau, S_\tau, Z_\tau) \mid X_t = (\gamma s, y, \gamma z)] = \mathbb{E}[\varphi(\tau, \gamma S_\tau, \gamma Z_\tau) \mid X_t = (s, y, z)].$$

for all $t \in \mathcal{T}^\circ$, $x = (s, y, z)$, $\tau \in \vartheta_t$, and $\gamma > 0$. By (4.1) and the homogeneity of α and β ,

$$\begin{aligned} \varphi(\tau, \gamma S_\tau, \gamma Z_\tau) &= e^{-r\tau} (-(\alpha(\gamma S_\tau, \gamma Z_\tau) - \beta(\gamma S_\tau, \gamma Z_\tau) - K))^+ \\ &= e^{-r\tau} (-(\gamma\alpha(S_\tau, Z_\tau) - \gamma\beta(S_\tau, Z_\tau) - K))^+ \\ &\leq e^{-r\tau} \gamma (-(\alpha(S_\tau, Z_\tau) - \beta(S_\tau, Z_\tau) - K))^+ + e^{-r\tau} (1 - \gamma)K \\ &= \gamma\varphi(\tau, S_\tau, Z_\tau) + e^{-r\tau} (1 - \gamma)K. \end{aligned}$$

We combine the above inequalities to arrive at the following:

$$\begin{aligned} \mathbb{E}[\varphi(\tau, S_\tau, Z_\tau) \mid X_t = (\gamma s, y, \gamma z)] &\leq \gamma \mathbb{E}[\varphi(\tau, S_\tau, Z_\tau) \mid X_t = (s, y, z)] + e^{-rt}(\gamma - 1)K \\ &\leq \gamma v(t, x) + e^{-rt}(1 - \gamma)K \\ &= \gamma\varphi(t, x) + e^{-rt}(1 - \gamma)K. \end{aligned}$$

As $v(t, x) = \varphi(t, s, z) > 0$, again by the homogeneity of α and β ,

$$\begin{aligned} \varphi(t, \gamma s, \gamma z) &= v(t, x) \\ &= \sup_{\tau \in \vartheta_t} \mathbb{E}[\varphi(\tau, S_\tau, Z_\tau) \mid X_t = (\gamma s, y, \gamma z)] \\ &\leq \gamma\varphi(t, s, z) + e^{-rt}(1 - \gamma)K \\ &= e^{-rt} (-(\gamma\alpha(s, z) - \gamma\beta(s, z) - \gamma K)) + e^{-rt}(1 - \gamma)K \\ &= e^{-rt} (-(\alpha(\gamma s, \gamma z) - \beta(\gamma s, \gamma z) - K)) \\ &= \varphi(t, \gamma s, \gamma z). \end{aligned}$$

Hence, $(\gamma s, y, \gamma z) \in \mathcal{S}_t$. □

Let f^* be as in (4.3). It is clear that if $\alpha(\bar{x}) > f^*(t, \Xi(\bar{x}))$, then $\bar{x} \notin \mathcal{S}_t$. Now, suppose $\alpha(\bar{x}) < f^*(t, \Xi(\bar{x}))$ for some $\bar{x} = (\bar{s}, \bar{y}, \bar{z})$. Then, by the above definition, there exists $x = (s, y, z) \in \mathcal{S}_t$ such that $\alpha(x) \geq \alpha(\bar{x})$ and $\Xi(x) = \Xi(\bar{x})$. By the definition of Ξ , we conclude that $y = \bar{y}$ and $(\bar{s}, \bar{z}) = (\gamma s, \gamma z)$ with $\gamma := \alpha(\bar{s}, \bar{z})/\alpha(s, z)$. Therefore, $(\gamma s, y, \gamma z) = \bar{x}$, and as $\gamma \leq 1$, by the above lemma $\bar{x} \in \mathcal{S}_t$. Summarizing, we have shown that

$$\{\alpha < f^*\} \subset \mathcal{S}_t, \quad \text{and} \quad \{\alpha > f^*\} \subset \mathcal{X} \setminus \mathcal{S}_t.$$

Now suppose that $\alpha(x) = f^*(t, \Xi(x))$. Since α is always strictly positive, $f^*(t, \Xi(x)) > 0$. Then, there is a sequence $x_n \in \mathcal{S}_t$ with $\Xi(x_n) = \Xi(x)$ and $\alpha(x_n) \uparrow f^*(t, \Xi(x)) = \alpha(x)$. As (α, Ξ) is a homeomorphism, we conclude that x_n converges to x . Moreover, \mathcal{S}_t is relatively closed in \mathcal{X} , implying that $x \in \mathcal{S}_t$. Hence, $\{\alpha = f^*\} \subset \mathcal{S}_t$. Consequently, the triplet (α, Ξ, f^*) satisfies (2.4), and the existence of a stopping boundary Assumption 2.3 holds.

B Convergence of \mathcal{R}_ϵ

The following limit result justifies our choice of the reward function \mathcal{R}_ϵ .

Lemma B.1. *Suppose that X_t has no atoms for all $t \in (0, T]$ and $x_0 = X_0$ is not on the boundary f , i.e., $\alpha(x_0) \neq f(0, \Xi(x_0))$. Then,*

$$\lim_{\epsilon \downarrow 0} \mathbb{E}[\mathcal{R}_\epsilon(X, f)] = \mathbb{E}[\varphi(\tau_f, X_{\tau_f})] = v(\tau_f). \quad (\text{B.1})$$

Proof. Under above assumptions, for every $t \in \mathcal{T}^\circ$, $d(t, X_t; f) \neq 0$ with probability one. Then as ϵ tends to zero, $p_t(X_t, f)$ converges to one if $d(t, X_t; f) > 0$ and respectively, to zero if $d(t, X_t; f) < 0$. Consequently, $b_t(X, f)$ converges to one for all $t \leq \tau_f$ and respectively, to zero for all $t > \tau_f$. These limit statements directly imply that

$$\lim_{\epsilon \downarrow 0} \mathcal{R}_\epsilon(X, f) = \varphi(\tau_f, X_{\tau_f}), \quad \text{a.s.}$$

Moreover, as $b_t, p_t \in [0, 1]$, Assumption 2.1 implies that

$$|\mathcal{R}_\epsilon(X, f)| \leq \sum_{t \in \mathcal{T}} \varphi(t, X_t) \leq C \sum_{t \in \mathcal{T}} [1 + |X_t|^a] =: \Phi.$$

Since \mathcal{T} is finite, by Assumption 2.1, Φ is integrable. Then, the claimed limit (B.1) follows from the dominated convergence theorem. \square

C Algorithm Details

The following is a brief outline of the training part of our algorithm.

Stopping Boundary Training:

1. Initialize $\theta_0 \in \Theta$
2. For $m = 0, \dots, M - 1$
 - Simulate trajectories $(X_{t_k}^j)_{k=0}^n$, $j = 1, \dots, B$.
 - For $k = 0, \dots, n - 1$, $j = 1, \dots$, compute the following quantities:
 - signed distances $d_{k,m,j} = \eta \left(g(t_k, \Xi(X_{t_k}^j); \theta^m) - \alpha(X_{t_k}^j) \right)$
 - stopping probabilities $p_{k,m,j} = \left(\frac{\epsilon - d_{k,m,j}}{2\epsilon} \right)^+ \wedge 1$ ($p_{n,m,j} = 1$)
 - stopping budgets $b_{k+1,m,j} = b_{k,m,j}(1 - p_{k,m,j})$ ($b_{0,m,j} = 1$)
 - reward function $R_\epsilon(\theta_m) = \frac{1}{B} \sum_{k=1}^B \sum_{k=0}^n p_{k,m,j} b_{k,m,j} \varphi(t_k, X_{t_k}^j)$
 - Update: $\theta_{m+1} = \theta_m + \zeta_m \nabla_{\theta} R_\epsilon(\theta_m)$
3. Return θ_M \square

In the above algorithm, we fix the width of the fuzzy region $\epsilon > 0$, and the learning rate process ζ_m is taken from the Adam optimizer [35]. Also in all financial examples, we use important sampling as discussed in the subsection 4.1. That is, chosen a parameter λ , we modify the dynamics of the state process X and use it in the simulations. Then, we adjust the reward function as in (4.4). The choice of the tuning parameters ϵ and λ are discussed in the specific examples.

References

- [1] L. Andersen and M. Broadie. Primal-dual simulation algorithm for pricing multidimensional American options. *Management Science*, 50(9):1222–1234, 2004.
- [2] A. Bachouch, C. Huré, N. Langrené, and H. Pham. Deep neural networks algorithms for stochastic control problems on finite horizon, part 2: Numerical applications. *arXiv:1812.05916*, 2018.
- [3] G. Barles, H. M. Soner, and P. E. Souganidis. Front propagation and phase field theory. *SIAM Journal on Control and Optimization*, 31(2):439–469, 1993.
- [4] C. Bayer, R. Tempone, and S. Wolfers. Pricing American options by exercise rate optimization. *Quantitative Finance*, 20(11):1749–1760, 2020.
- [5] C. Bayer, J. Qiu, and Y. Yao. Pricing options under rough volatility with backward SPDEs. *SIAM Journal on Financial Mathematics*, 13(1):179–212, 2022.
- [6] S. Becker, P. Cheridito, and A. Jentzen. Deep optimal stopping. *Journal of Machine Learning Research*, 20(74):1–25, 2019.
- [7] S. Becker, P. Cheridito, and A. Jentzen. Pricing and hedging American-style options with deep learning. *Journal of Risk and Financial Management*, 13(7):158, 2020.
- [8] S. Becker, P. Cheridito, A. Jentzen, and T. Welti. Solving high-dimensional optimal stopping problems using deep learning. *European Journal of Applied Mathematics*, 32(3):470–514, 2021.
- [9] S. Becker, A. Jentzen, M. S. Müller, and P. von Wurstemberger. Learning the random variables in Monte Carlo simulations with stochastic gradient descent: Machine learning for parametric PDEs and financial derivative pricing. *arXiv:2202.02717*, 2022.
- [10] D. 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.
- [11] W. J. Boettinger, J. A. Warren, C. Beckermann, and A. Karma. Phase-field simulation of solidification. *Annual Review of Materials Research*, 32(1):163–194, 2002.
- [12] M. Broadie and M. Cao. Improved lower and upper bound algorithms for pricing american options by simulation. *Quantitative Finance*, 8:845–861, 12 2008.
- [13] M. Broadie and J. Detemple. The valuation of American options on multiple assets. *Mathematical Finance*, 7:241–286, 1997.
- [14] M. Broadie and P. Glasserman. A stochastic mesh method for pricing high-dimensional american options. *Journal of Computational Finance*, 7:35–72, 2004.
- [15] H. Buehler, L. Gonon, J. Teichmann, and B. Wood. Deep hedging. *Quantitative Finance*, 19(8):1271–1291, 2019.
- [16] H. Buehler, L. Gonon, J. Teichmann, B. Wood, and B. Mohan. Deep hedging: hedging derivatives under generic market frictions using reinforcement learning. Technical report, Swiss Finance Institute, 2019.
- [17] E. Chevalier, S. Pulido, and E. Zúñiga. American options in the Volterra Heston model. *arXiv:2103.11734*, 2021.

- [18] D. Ciocan and V. Mišić. Interpretable optimal stopping. *Management Science*, 68(3):1616–1638, 2022.
- [19] A. Conze and Viswanathan. Path dependent options: The case of lookback options. *The Journal of Finance*, 46(5):1893–1907, 1991.
- [20] G. V. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2:303–314, 1989.
- [21] M. Dai and Y. K. Kwok. American options with lookback payoff. *SIAM Journal on Applied Mathematics*, 66(1):206–227, 2005. ISSN 00361399.
- [22] J. Detemple. *American-Style Derivatives: Valuation and Computation*. Chapman and Hall/CRC Financial Mathematics Series. CRC Press, 2005.
- [23] S. Fecamp, J. Mikael, and X. Warin. Deep learning for discrete-time hedging in incomplete markets. *Journal of computational Finance*, 2020.
- [24] W. H. Fleming and H. M. Soner. *Controlled Markov processes and viscosity solutions*, volume 25. Springer Science & Business Media, 2006.
- [25] E. Fried and M. E. Gurtin. Dynamic solid-solid transitions with phase characterized by an order parameter. *Physica D: Nonlinear Phenomena*, 72(4):287–308, 1994.
- [26] D. Garcia. 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.
- [27] M. Germain, M. Laurière, H. Pham, and X. Warin. Deepsets and their derivative networks for solving symmetric pdes. *arXiv preprint arXiv:2103.00838*, 2021.
- [28] P. Glasserman. *Monte Carlo methods in financial engineering*. Springer, 2004.
- [29] L. Gonon, J. Muhle-Karbe, and X. Shi. Asset pricing with general transaction costs: Theory and numerics. *arXiv:1905.05027*, 2019.
- [30] J. Han and W. E. Deep learning approximation for stochastic control problems. *NIPS*, 2016.
- [31] J. Han, A. Jentzen, and W. E. Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences*, 115(34):8505–8510, 2018.
- [32] K. Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257, 1991. ISSN 0893-6080.
- [33] C. Huré, H. Pham, A. Bachouch, and N. Langrené. Deep neural networks algorithms for stochastic control problems on finite horizon, part I: convergence analysis. *arXiv:1812.04300*, 2018.
- [34] C. Huré, H. Pham, and X. Warin. Some machine learning schemes for high-dimensional nonlinear pdes. *arXiv:1902.01599*, 33, 2019.
- [35] D. Kingma and J. A. Ba. Adam: a method for stochastic optimization. *Proceedings of the International Conference on Learning Representations (ICLR)*, 2015.

- [36] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, 1992.
- [37] M. Kohler, A. Krzyżak, and N. Todorovic. Pricing of high-dimensional american options by neural networks. *Mathematical Finance*, 20, 08 2008.
- [38] D. Lamberton and G. Terenzi. Properties of the American price function in the Heston-type models. *arXiv:1904.01653*, 2019.
- [39] F. A. Longstaff and E. S. Schwartz. Valuing American options by simulation: a simple least-squares approach. *The Review of Financial Studies*, 14(1):113–147, 2001.
- [40] M. Ludkovski. mlosp: Towards a unified implementation of regression Monte Carlo algorithms. *arXiv:2012.00729*, 2020.
- [41] S. Osher and R. Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces*. Springer, 2003.
- [42] A. M. Reppen and H. M. Soner. Deep empirical risk minimization in finance: an assessment. *arXiv:2011.09349*, 2021.
- [43] A. M. Reppen, H. M. Soner, and V. Tissot-Daguette. Deep stochastic optimization in finance. 2022. preprint.
- [44] H. Robbins and D. Siegmund. A convergence theorem for non negative almost supermartingales and some applications. In *Optimizing methods in statistics*, pages 233–257. Elsevier, 1971.
- [45] J. Ruf and W. Wang. Hedging with neural networks. *arXiv:2004.08891*, 2020.
- [46] J. Ruf and W. Wang. Neural networks for option pricing and hedging: a literature review. *Journal of Computational Finance*, 24(1), 2020.
- [47] J. Sirignano and K. Spiliopoulos. Dgm: A deep learning algorithm for solving partial differential equations. *Journal of Computational Physics*, 375:1339–1364, 2018.
- [48] H. M. Soner. Convergence of the phase-field equations to the Mullins-Sekerka problem with kinetic undercooling. *Archive for Rat. Mech. and Analysis*, 131(2):139–197, 1995.
- [49] J. N. Tsitsiklis and B. Van Roy. Regression methods for pricing complex American-style options. *IEEE Transactions on Neural Networks*, 12(4):694–703, 2001.
- [50] V. R. Voller and C. Prakash. A fixed grid numerical modelling methodology for convection-diffusion mushy region phase-change problems. *International journal of heat and mass transfer*, 30(8):1709–1719, 1987.
- [51] S. Wang and P. Perdikaris. Deep learning of free boundary and Stefan problems. *Journal of Computational Physics*, 428:109914, 2021.
- [52] X. Warin. Variance optimal hedging with application to electricity markets. *Journal of Computational Finance*, 23(3):33–59, 2019.