Optimal Stopping in High-dimensions

H. Mete Soner, ORFE, Princeton

---

joint work with
Max Reppen, Boston University
Valentin Tissot-Daguette, Princeton

Stochastic Control and Quantitative Finance
Hebrew University
September 14 , 2022

Recent years have seen the development of generic and efficient algorithms for solving stochastic optimal control and stopping problems. Common structures are :

- Formulated as empirical risk minimization problems (distinct type of modeling from typical RL).
- Convergence is understood to hold in idealized settings.
- Implementations typically postulate a market model from which the training data is simulated.
- The methodology allows for completely data driven approaches, provided sufficient data availability.
- The black box nature of neural networks can make the results hard to interpret.

In this study : We impose structure (e.g., parametrized stopping boundary). So that we improve *interpretability*, while simultaneously improving training properties.

Outline

Problem

- Fix a time horizon $T > 0$, a probability space $(\Omega, \mathbb{Q})$, and a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$.
- $\mathcal{T} \subset [0, T]$ is a *finite* set of time point at which stopping is allowed.
- Let $\vartheta = \vartheta(\mathcal{T})$ is the set of $\mathbb{F}-$stopping times taking values in $\mathcal{T}$.
- The state process solves

$$\boxed{\mathrm{d}X_t = \mu(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t,}$$

  where W is the Brownian motion.
- For a given reward function φ, the problem is

$$\boxed{\text{to maximize} \quad v(x, t; \tau) := \mathbb{E}_{\mathbb{Q}}[\varphi(\tau, X_\tau) \mid X_t = x], \qquad \text{overall } \tau \in \vartheta.}$$

Define the stopping region as $\mathcal{S}_t := \{x \in \mathcal{X} \;:\; \varphi(t,x) = \sup_\tau v(t,x;\tau)\}$, and set

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

Assumption (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 $\phi \in \mathcal{L}_a$, $t < \bar{t} \in \mathcal{T}$,

$$\mathbb{E}[\phi(X_{\bar{t}}) \mid X_t = x] \in \mathcal{L}_a, \qquad x \in \mathcal{X}.$$

Proposition (Regularity)

Under the growth and continuity assumption, for each $t \in \mathcal{T}$, the value function $v(t,\cdot) := \sup_\tau v(t,\cdot,\tau) \in \mathcal{L}_a$, and the stopping region $\mathcal{S}_t$ is a relatively closed subset of $\mathcal{X}$.

Set

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

Assumption (Existence of a stopping boundary)

There exist measurable functions

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

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, and*

$$\boxed{\mathcal{S}_t = \left\{x \in \mathcal{X} \;:\; \eta\left(f^*(t, \Xi(x)) - \alpha(x)\right) \leq 0\right\}, \qquad \forall\, t \in \mathcal{T}^\circ.}$$

We prove that it holds for general types of option problems.

$$\mathcal{S}_t = \big\{x \in \mathcal{X} \;:\; \eta\big(f^*(t, \Xi(x)) - \alpha(x)\big) \leq 0\big\}, \qquad \forall\ t \in \mathcal{T}^\circ.$$

Example (American put option)

With Markovian dynamics in one dimension, $\mathcal{X} = \mathbb{R}_+$, $\alpha(x) = x$, $\Xi \equiv 0$, and $\eta = -1$. Thus, the stopping region is the hypo-graph of some stopping boundary f^*.

Example (American max-call option)

With two stocks and strike K : $\varphi(t, x) = (\max\{x_1, x_2\} - K)^+$ for $x = (x_1, x_2) \in \mathcal{X}$.
With Markovian dynamics, the state space is $\mathcal{X} = \mathbb{R}_+^2$.
There exists a stopping boundary f^* such that the triplet (α, Ξ, f^*) with

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

satisfies the assumption.

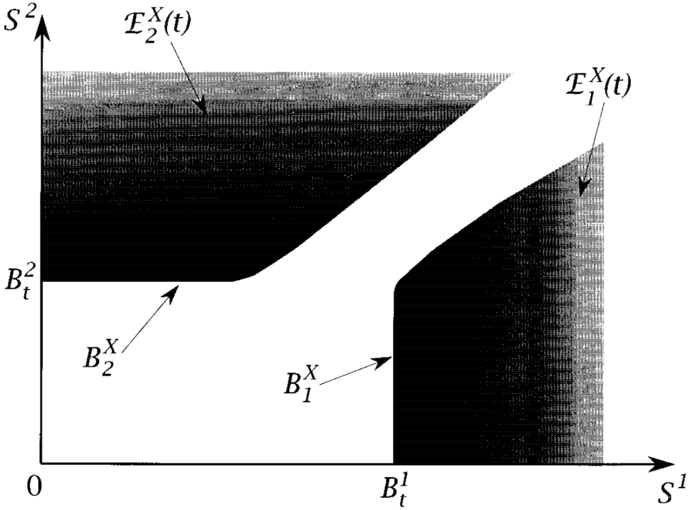


Image from *The Valuation of American Options on Multiple Assets*, by Broadie, M, and Detemple, J, published in *Mathematical Finance*, 7/3,241–286, 1997.

Instead of parameterizing the stopping rule directly, we parameterize the stopping boundary. Why ?

- *Interpretability :* The stopping boundary gives us an actual rule about which we can reason. A $\{0, 1\}$ output for an evaluated point is not as informative.
- *Topological guarantee :* With other methods there is no guarantee that the stopping region is connected even in 1-D. The stopping boundary approach can make sure that the resulting stopping rule has the desired topological properties.
- *Computational :* The boundary *structure* constrains the solution relative to learning stopping times, which makes it "easier" by reducing the hypothesis space.
- *Extensions :* to other free boundary problems is possible but we believe these option problems are good benchmark problems.

Deep Empirical Minimization

Outline

All control problems can be formulated as

$$\boxed{\text{to maximize} \quad v(\mathrm{u}) := \mathbb{E}\left[\ell(\mathrm{u}, \xi)\right],}$$

where ξ is the increments of a stochastic process and $\mathrm{u} \in \mathcal{C}$ is the feedback control or action.

Weinan E, Jiequn Han, and Arnulf Jentzen proposed an algorithm based on *empirical risk minimization*

Deep learning approximation for stochastic control problems, Han & E, W., *NIPS*, 2016.

Solving high-dimensional partial differential equations using deep learning, Han, Jentzen & E, W., *Proceedings of the National Academy of Sciences*, 115/34, 8505–8510, 2018.

For hedging problems, similar algorithm developed in

Deep hedging, by Bühler, H., Gonon, L., Teichmann, J., and Wood, B., *Quantitative Finance*, 2019.

For a batch B of sample paths $\{X^1, \ldots, X^B\}$ (simulated or obtained from data), let $\hat{\mathbb{E}}^B$ be the empirical expectation, and set

$$v(\tau; B) = \hat{\mathbb{E}}^B\big[\, e^{-r\tau}\varphi(\tau, X_\tau) \big| X_t = x \big],$$

- *Parameterize* the stopping decision by θ as a function of the state and construct the corresponding stopping time $\tau(\theta)$;
- *Optimize by stochastic gradient ascent :* for $n = 0, 1, \ldots$:
  - *Generate* a batch B,
  - *Compute* the derivative $d := \nabla_\theta\, v(\tau(\theta); B)$;
  - *Update* $\theta \leftarrow \theta + \kappa d$.
- *Stop* if n is sufficiently large and the improvement of the value is ‘small’.

Difficulty : As such, the stopping rule is locally constant (0 or 1), so the gradient is zero.

We resolve this difficulty by using randomized stopping and fuzzy boundaries.

For a batch B of sample paths $\{X^1, \ldots, X^B\}$ (simulated or obtained from data), let $\hat{\mathbb{E}}^B$ be the empirical expectation, and set

$$v(\tau; B) = \hat{\mathbb{E}}^B \big[e^{-r\tau} \varphi(\tau, X_\tau) \big| X_t = x \big],$$

- *Parameterize* the stopping decision by θ as a function of the state and construct the corresponding stopping time $\tau(\theta)$;
- *Optimize by stochastic gradient ascent :* for $n = 0, 1, \ldots$:
  - *Generate* a batch B,
  - *Compute* the derivative $d := \nabla_\theta\, v(\tau(\theta); B)$;
  - *Update* $\theta \leftarrow \theta + \kappa d$.
- *Stop* if n is sufficiently large and the improvement of the value is 'small'.

Difficulty : As such, the stopping rule is locally constant (0 or 1), so the gradient is zero.

We resolve this difficulty by using randomized stopping and fuzzy boundaries.

We consider two benchmark problems, based on

Deep optimal stopping, by Becker, S., Cheridito, P. and Jentzen, A., published in *Journal of Machine Learning Research*, 19/8,1271–1291, 2019.

Put : $\varphi(\tau, \{X_u\}_{u\in[0,\tau]}) = (K - X_\tau)^+$, where r, K is given ;

Max-Call : For $X_u = (X_u^1, \ldots, X_u^d) \in \mathbb{R}^d$, $\varphi(\tau, \{X_u\}_{u\in[0,\tau]}) = (\max_i X_u^i - K)^+$, (with dividends).

- For simplicity, consider $d = 1$. The stopping time is based on a *stopping boundary* :

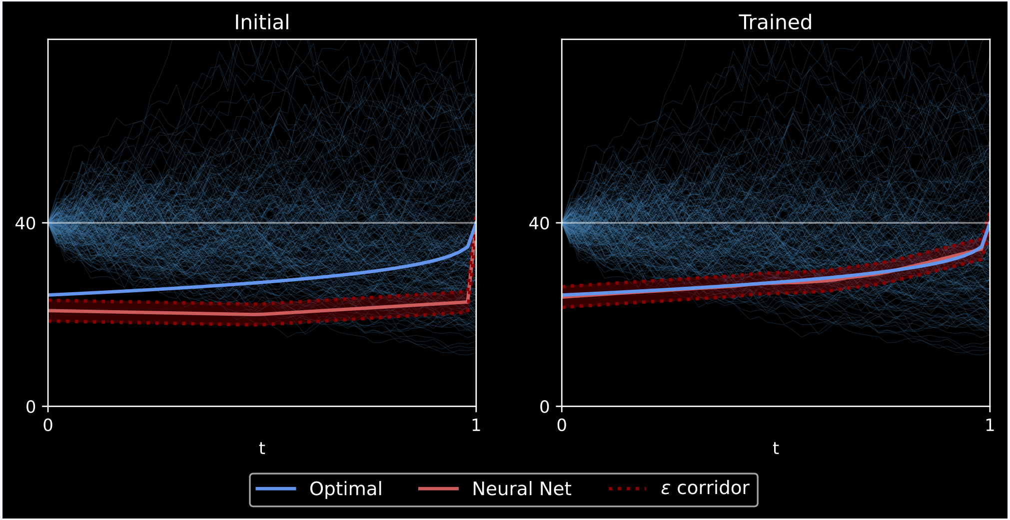
$$f : [0, T] \to \mathbb{R}_+ \qquad \mapsto \qquad \tau = \inf \{ u \in [0, T] \, : \, X_u \leq f(u) \} .$$

- We *train for the stopping boundary* by maximizing the composition :

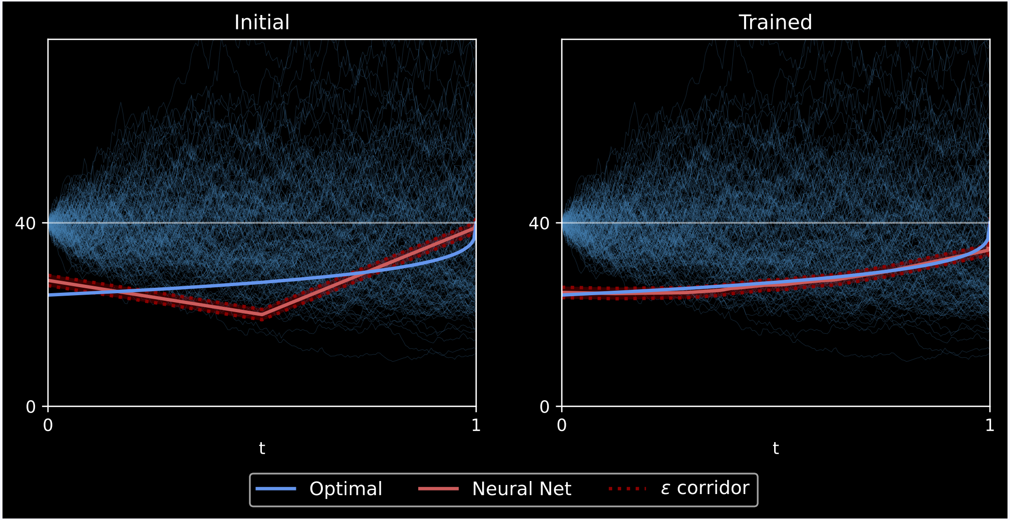
$$\boxed{\theta \quad \mapsto \quad f(\cdot;\theta) \quad \text{(boundary)} \quad \mapsto \quad \tau \quad \text{(stopping time)} \quad \mapsto \quad v(\tau) \quad \text{(reward).}}$$

- Difficulties :
  1. The map from the stopping boundary to the reward is very flat and difficult to train for.
  2. Again, $\{0,1\}$-problem : the derivative $\nabla_\theta v(\tau(\theta); B)$ is zero.
- Our solutions :
  1. We use importance sampling to make the boundary relevant for more data points.
  2. For the $\{0,1\}$-problem, we relax the problem using a fuzzy boundary : during training we stop with some probability depending on the distance from the boundary.

Training progress



Training progress



Outline

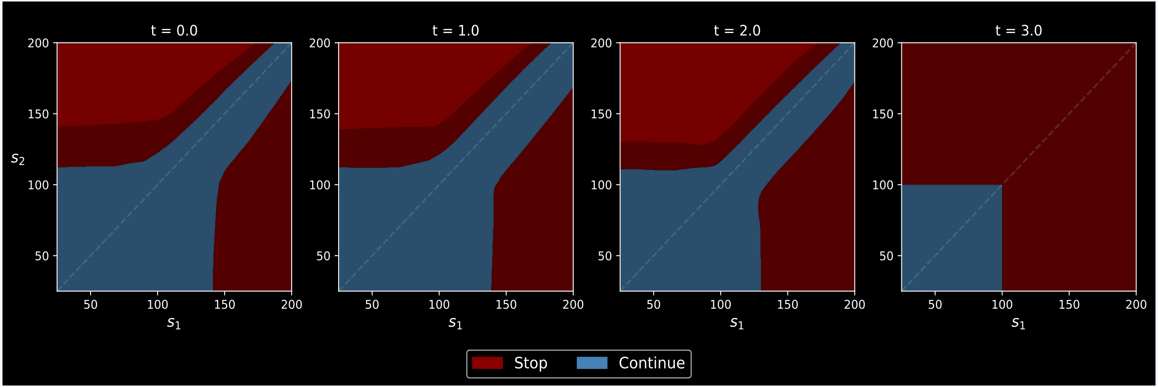
Another benchmarking example studied by the same group is :

- Let $X_t \in \mathbb{R}^d$, with $dX_t^{(i)} = X_t^{(i)}[r\mathrm{d}t + \sigma \mathrm{d}W_t^{(i)}]$, where $W^{(i)}$ are independent Brownian motions.
- One may stop at N different time points to receive

$$\varphi(t, X_t) = \left(\max_i X_t^{(i)} - K\right)^+ .$$

- Following Deep Optimal Stopping, $N = 9$.
- Nunmerically, we have obtained similar precision as in the Deep Optimal Stopping, but with lower computational resources, and topological guarantees.

Max-call option



The stocks are asymmetric. The top left light red region is the bottom right region mirrored in the diagonal.

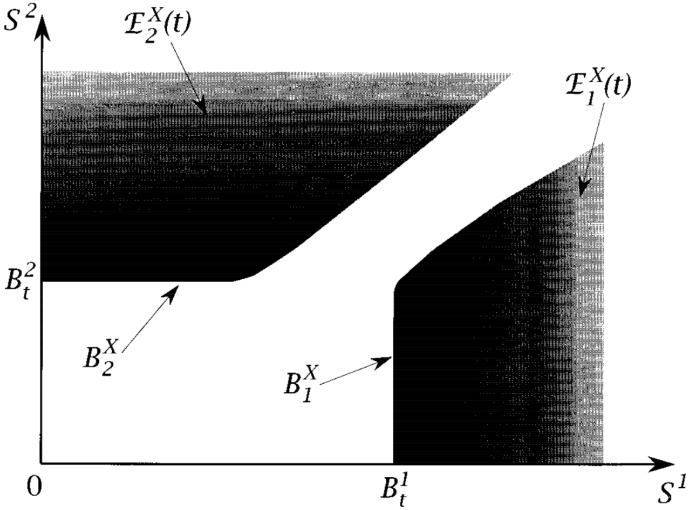


Image from *The Valuation of American Options on Multiple Assets*, by Broadie, M, and Detemple, J, published in *Mathematical Finance*, 7/3,241–286, 1997.

- The problem cannot be solved by directly parameterizing the 'control' τ.
- For any stopping decision χ_θ taking values in $\{0,1\} \equiv \{\text{continue}, \text{stop}\}$, $\nabla_\theta \chi_\theta = 0$.
  As a consequence, for $\tau(\theta)$ obtained from χ_θ, $\nabla_\theta \tau = 0$

  $$\nabla_\theta v(\tau(\theta); B) \sim v'(\tau(\theta); B) \nabla_\theta \tau(\theta) = 0.$$

  Gradient ascent does not make progress.
- Consider the *relaxed* version :
  Let $\mathrm{u} : \mathcal{T} \to [0,1]$ be a non-decreasing function representing the cumulative *probability of stopping* :
  e.g., $\mathrm{u}_s = 0.6$ for $s \in [0, T]$ means that we have stopped with 60% probability by time s.
- The stopping event is independent of the stock price and we set

  $$v(\mathrm{u}) := \mathbb{E}\left[\sum_{k=1}^{N} \varphi(t_k, X_{t_k})\,(u(t_k) - u(t_{k-1}))\right].$$

- Then, the gradient ascent works, and a maximum value is obtained with precision.

The algorithm

Outline

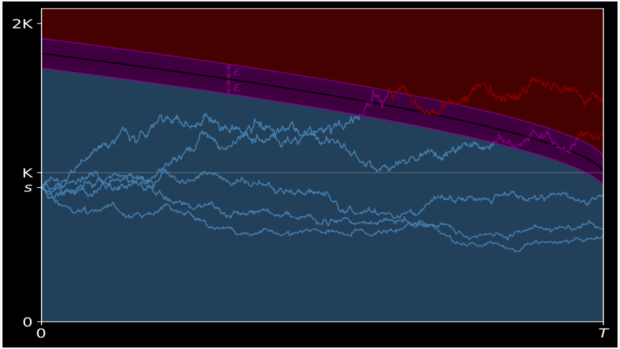
- Let $\mathfrak{F}$ be the set of all measurable functions $f : \mathcal{T}^\circ \times \ \Xi(\mathcal{X}) \to [0, \infty]$.
- For $f \in \mathfrak{F}$, define the distance to the boundary by,

$$d(t, X_t; f) := \eta\left(f(u, \Xi(X_u)) - \alpha(X_u)\right).$$

- 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* $\varepsilon > 0$, and at time $t \in \mathcal{T}^\circ$, define

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

be the *fuzzy region*.



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

- If the state is not in $\mathfrak{B}_t^{\varepsilon}(f)$, we continue or stop, as before, with probability one.
- If we are in the fuzzy region, we stop with probability $p_t(X_t, f)$, where

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

- 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 to maximize $\mathbb{E}[\mathcal{R}_\varepsilon(X,f)]$ over all $f \in \mathfrak{F}$, where

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

  and $b_t(X,f)$ is the probability of not stopping strictly before time t.
- $b_t(X,f)$ is obtained as the solution of,

$$b_{t+1}(X,f) = b_t(X,f)(1 - p_t(X_t,f)), \qquad t \in \mathcal{T}^\circ, \quad \text{with} \quad b_0(X,f) = 1.$$

- As $\lim_{\varepsilon \downarrow 0} \mathcal{R}_\varepsilon(X,f) = \varphi(\tau_f, X_{\tau_f})$, this problem approximates the original one.

Outline

Pseudo Code

1. *Initialize* $\theta \in \Theta$.
2. *Simulate* independent state trajectories, $\{X^{(1)}, \ldots, X^{(B)}\}$, where B is the *batch size*.
3. Compute the *empirical reward* function :

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

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

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

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.

Network Architecture and Parameters

1. 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.
2. We use only one deep neural network that takes both the time and the state vector as input.
3. 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 $20 + d$ many nodes
4. We set the number of stochastic gradient iterations to $M = 3,000$ with a fixed simulation batch size of $B = 512$.
5. The learning rate process ζ is taken from the Adam optimizer.
6. 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.
7. 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.

We do not use one network per time point but one network that takes time as input as well..

This comes with some benefits :

- It appears to avoid the issue of training degradation with more time points.
- Excessive overfitting at time t is less likely, because part of that overfitting would affect other time points $s \neq t$, which would be unfavorable.
- We can change the time discretization during training.
  This potentially lets us get closer to American than Bermudan.

Put : $\varphi(\tau, \{S_u\}_{u\in[0,\tau]}) = (K - S_\tau)^+$, where r, K is given

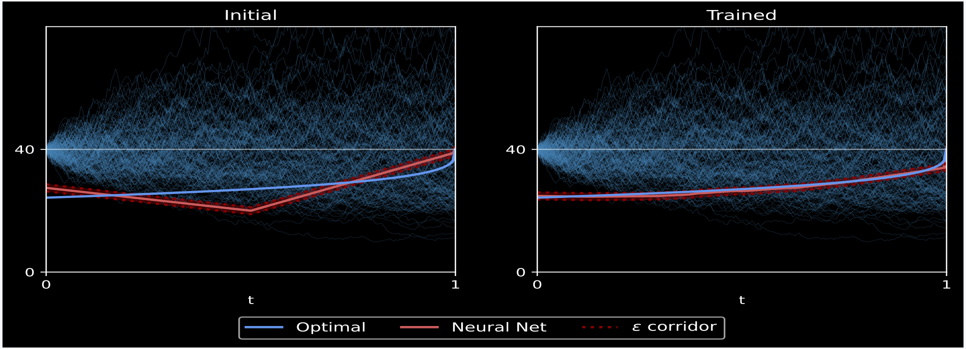


Figure 1: Almost American option in Black-Scholes model with $n = 250$.

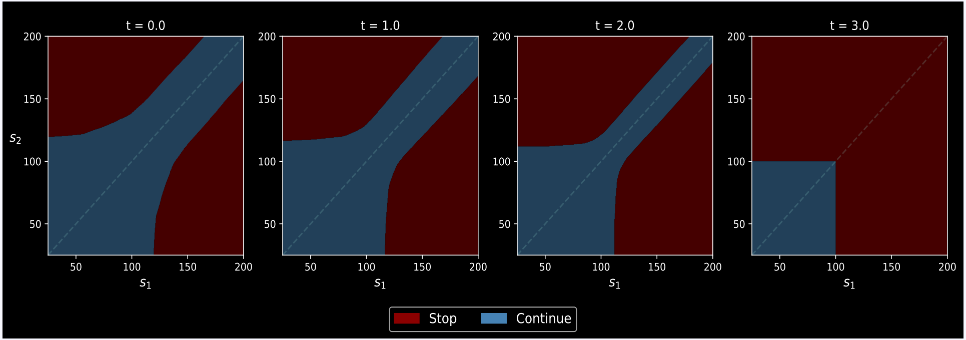


Figure 2: Stopping Boundaries with $s_0 = 90$, $K = 100$, $\sigma_i = 0.2$, $r = 0.05$, $\delta_i = 0.1$.

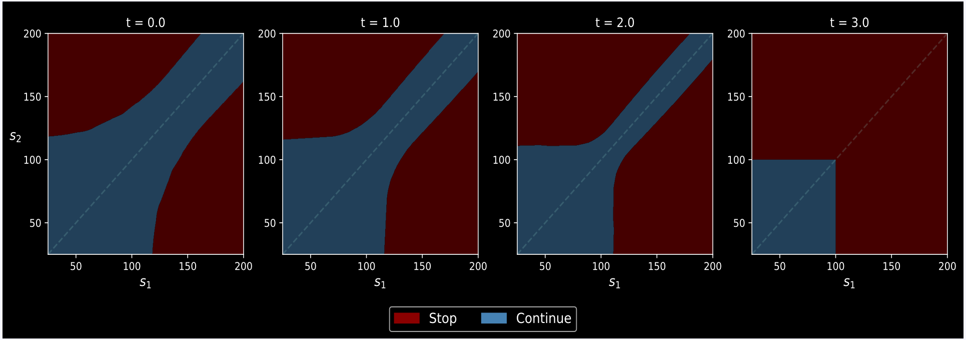


Figure 3: Stopping Boundaries with $s_0 = 100$, $K = 100$, $\sigma_i = 0.2$, $r = 0.05$, $\delta_i = 0.1$.

Max-call option with $s_0 = 100$

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

Table 1: 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.

- We ran experiments for the Max-Call options up to 500 dimensions.
- This is done on a personal computer (no GPU !)
- We could results as good or better than the ones obtained in Becker, Cheridito & Jentzen.
- Naive methods however, ran into problems and gave wrong results. For instance, a straightforward approach to a Put option in one-dimension, resulted in zero free boundary and a Call with non-trivial boundary.
- Large number of simulations were used. So a model for simulations is needed.

There are many recent papers using the recent computational techniques. Here is a partial list.

- *Statistical learning for probability-constrained stochastic optimal control*, Balata, Ludkovski, Maheshwari & Palczewski.
- *Asset Pricing with General Transaction Costs : Theory and Numerics*, Gonon, Muhle-Karbe & Shi.
- *Deep learning for discrete-time hedging in incomplete markets*, Fecamp, Mikael & Warin.
- *Machine learning for semi linear PDEs*, Chan-Wai-Nam, Mikael & Warin.
- *Learning a functional control for high-frequency finance*, Leal, Laurière & Lehalle.
- *Deep neural networks algorithms for stochastic control problems on finite horizon, part I : convergence analysis*, *part 2 : Numerical applications*, Huré, Pham, Bachouch & Langrené.
- A very related active area is data augmentation or data-driven models.
- Network architecture, random features is also actively studied.

- Deep empirical risk minimization is an effective and a flexible tool.
- It can handle details of the markets as well as complicated and general dynamics with ease.
- One has to be careful with the complexity of the networks and the size of the training data.
- Optimization step is the least understood part.

THANK YOU FOR YOUR ATTENTION.

Max Reppen of Boston University, and HMS

Deep empirical risk minimization in finance : looking into future

Max Reppen of Boston University, Valentin Tissot-Daguette of Princeton and HMS

Deep stochastic optimization in finance

Neural Optimal Stopping Boundary.

- Deep empirical risk minimization is an effective and a flexible tool.
- It can handle details of the markets as well as complicated and general dynamics with ease.
- One has to be careful with the complexity of the networks and the size of the training data.
- Optimization step is the least understood part.

THANK YOU FOR YOUR ATTENTION.

Max Reppen of Boston University, and HMS

Deep empirical risk minimization in finance : looking into future

Max Reppen of Boston University, Valentin Tissot-Daguette of Princeton and HMS

Deep stochastic optimization in finance

Neural Optimal Stopping Boundary.