Deep optimal stopping and switching

Approximating the indicator function with NN

Claudia Viaro

Contents

1 Introduction **2**

2 Optimal stopping problems **2**

2.1 Approximation methods with neural networks 4

3 Optimal switching **6**

4 Approximating the indicator function **8**

4.1 Stopping decisions $0 - 1$ 8

4.2 Neural network 9

5 An application of the optimal switching problem **10**

5.1 Training 10

5.2 Testing - Lower bound 10

5.3 Testing - Upper bound 10

1 Introduction

The present report is intended to

An updated record of the implementations produced can be found at the following links, distinguishing the two approaches: Optimal Switching.

2 Optimal stopping problems

The treatment of a sequential decision making setting can be looked at from a more established and traditional perspective, that of an optimal stopping problem. We are interested in finding an optimal control policy that determines when to perform an intervention, given some observed characteristics, so as to maximize the expected total discounted return. It is clear that the problem at hand belong once again to the general class of Markov Decision Processes (MDP).

Fields of application largely include financial strategies (asset selling, hedging) but also sequential hypothesis testing. This topic is also relevant for healthcare applications such as in Ajdari et al. (2019), in determining when to stop the treatment of patients receiving fractionated radiotherapy treatments. Some applications also feature handling of engagement platforms (Liyanage et al. 2019).

Definition 1: Optimal stopping

Let $(\Omega, \mathbb{F} = (\mathcal{F})_{n=0}^{N}, \mathbb{P})$ be a complete probability space, with $N \in \mathbb{N}$ be a natural number. Consider the time discretization with equidistant time grid $t_0, t_1, \dots, t_N \in [0, T]$ and $0 = t_0 < t_1 < \cdot < t_N = T$. The discretization allows for a few simplifications involving the use of simpler algorithms for dicrete-state MDP (as opposed to semi/continuous MDP) as well as simplifications regarding the computation of one-step conditional expectations.

Let a d-dimensional discrete-time Markov process describe the asset price under a Black-Scholes market $(X_{t_n})_{n=0}^{N}$ with $X_0 = x_0$ and be defined on the probability space given:

$$X_t^i = (r - \delta_i)dt + \sigma_i dW_t^i \tag{1}$$

where $r \in [0, \infty]$ is the market risk free rate, $\delta_i \in [0, \infty]$ is the dividend yield, $\sigma_i \in [0, \infty]$ is the volatility and $W : [0, T] \times \Omega \to \mathbb{R}^d$ be a standard Brownian motion on the same probability space, with $i = 1, \dots, d$ stock prices.

Call a random variable $\tau : \Omega \to \{t_0, t_1, \dots, t_N\}$ an X-stopping time if $\{\tau = n\} \in \mathcal{F}$. The notion of reward (hence the price) we seek to maximize is:

$$V_n = \sup_{\tau \in \mathcal{T}_n} \mathbb{E}[g(\tau, X_\tau)] \tag{2}$$

where $\mathcal{T}_n$ denotes the family of all stopping times and $g(n, X_{t_n})$ is the discounted payoff process.

The problem is then a Markov decision process with state space $\mathcal{S} = \{0, 1, \dots, N\} \times \mathbb{R}^d \times \{0, 1\}$, binary action space $\mathcal{A} = \{0, 1\}$, where $a = 0$ stands for holding on the option and $a = 1$ for stopping the process (e.g. exercising the option) and reward function

$$R((n, X_{t_n}), a) = \begin{cases} g(n, X_{t_n}) & \text{if a=1} \\ 0 & \text{if a=0} \end{cases} \tag{3}$$

for $n = 0, \dots, N$, transition kernel p induced by the dynamics of the $\mathbb{F}$-Markovian process $(X_{t_n})_{n=0}^{N}$. The state space includes time, the d-dimensional Markovian process and an absorbing state that captures the event of exercising or not the option, as we do not consider repeated controls.

If perfect knowledge of the MDP is assumed and its size is limited, optimal stopping problems with finitely many stopping opportunities can be solved exactly, such as with backward induction on binomial/trinomial trees or lattice. It is in fact known that the Snell envelope V^* provides an optimal solution to 2 (Peskir & Shiryaev 2006):

$$\begin{aligned} U_N &:= g(X_N), \\ U_n &:= \max(g(X_n), \mathbb{E}[\alpha U_{n+1}|X_n]), \end{aligned} \tag{4}$$

where α is a discount factor. Equivalently, U_n can be expressed as the optimal stopping problem of the form:

$$U_n = \sup_{\tau \in \mathcal{T}_n} \mathbf{E}[\alpha^{\tau - n} g(X_\tau|X_n)] \tag{5}$$

where $\mathcal{T}_n$ is the set of all stopping times $\tau \leq n$. The associated optimal stopping time τ^* is the first time the immediate reward dominates the continuation value:

$$\begin{aligned} \tau_N &:= N, \\ \tau_n &:= \begin{cases} n, & \text{if } g(X_n) \geq \mathbb{E}[\alpha U_{n+1}|X_n] \\ \tau_{n+1}, & \text{otherwise.} \end{cases} \end{aligned} \tag{6}$$

On the other hand, when the state space is large (e.g. when the payoff depends on several underlying assets or when the payoff depends on the history of underlying's prices), the classical algorithms used in the mathematical finance for exotic derivative pricing are no longer computationally tractable. Several approximations methods have been studied in the context of American/Bermudan option pricing, from approximation of the Snell envelope or continuation values (Longstaff & Schwartz 2001, Carriere 1996), to dual methods (Haugh & Kogan 2004). More recently, numerical approximation methods that are based on deep learning have been exploited.

2.1 Approximation methods with neural networks

A more recent line of research use backward recursion and stochastic gradient based methods to tackle high dimensional stopping problems. Casting these optimal pricing problems as MDPs allows to tackle them by means of Dynamic Programming or Reinforcement Learning algorithms, thus providing an interesting and valuable alternative to the traditional methods of derivative pricing. In particular, RL tend to be good at handling large state spaces by effectively leveraging sampling and function approximation methodologies in the context of solving the Bellman Optimality Equation, thus breaking the 'curse of dimensionality' problem (Rao & Jelvis n.d.).

In the literature there has been proposed two different approaches with respect to the object to be approximated by stochastic gradient based methods. We can distinguish approximation of the continuation value and the indicator function. These methods do not provide theoretical convergence guarantees due to the use of stochastic gradient methods with non-convex loss functions.

Continuation value

Consider m realizations from a d-dimensional Markovian stochastic process X under $\mathbb{P}$, where the i-th realization is $x_0, x_1^i, \dots, x_N^i$, with fixed spot price x_0. For each m, the price under stopping strategy 2, can be written as the following backward recursions:

- Tsitsiklis & Van Roy (2001) used the dynamic programming equation on the value function $p_n^i = \max(g(x_n^i), c_{\theta_n}(x_n^i))$ and estimate directly the current price;
- Longstaff & Schwartz (2001) use the continuation value only for the decision to stop or continue and not for the estimation of the price:

$$\begin{aligned} p_N^i &:= g(x_N^i), \\ p_n^i &:= \begin{cases} g(x_n^i), & \text{if } g(x_n^i) \geq c_n(x_n^i), \\ \alpha p_{n+1}^i, & \text{otherwise.} \end{cases} \end{aligned} \tag{7}$$

  where $c_n(x_n^i)$ is the continuation value. This backward induction method is based on Longstaff & Schwartz (2001)'s results and is the most used in the field.

Originally, both versions in (Longstaff & Schwartz 2001, Tsitsiklis & Van Roy 2001) used a regression estimation of the conditional expectation by expressing the continuation value as $c_\theta(x_n^i) = \theta^T \phi(x_n^i)$, where $\phi = (\phi_1, \dots, \phi_K)$ is a set of K basis functions and $\theta \in \mathbb{R}^K$ are the weights. More recently, instead, Kohler et al. (2010), Lapeyre & Lelong (2021), Becker et al. (2020) have employed dynamic programming to approximate these with neural networks as:

$$f_\theta(x_n^i) \approx c_\theta(x_n^i) \tag{8}$$

Indicator function

In the present survey we will focus on the approximation methods involving the indicator function. Becker et al. (2019) introduced an expression for approximating the indicator function:

$$f_\theta(x_n^i) \approx \mathbf{1}_{\{g(x_n^i) \geq c(x_n^i)\}} \tag{9}$$

This allows to estimate the current price as:

$$p_n^i = g(x_n^i) f_{\theta_n}(x_n^i) + \alpha p_{n+1}^i (1 - f_{\theta_k}(x_n^i)) \tag{10}$$

The optimization involves directly the option price $\psi_n(\theta_n) = \frac{1}{m} \sum_{i=1}^m \alpha p_n^i$.

3 Optimal switching

Consider now an optimal switching problem, that is finding the optimal sequence of opening (starting) and closing (stopping) times of a multi-actions sequential process. Problems of this nature have arisen naturally in domains such as when evaluating the profitability of an investment in the natural resource industry where the production depends on the market price of a number of underlying commodities or assets (Brennan & Schwartz 1985). Since this early work, there have been proposed several extensions looking at different aspects such as the model describing the underlying commodity X as being a diffusion process for example (Knudsen et al. 1998, Zervos 2003) or looking at different choices for the filtration to which the stochastic process is adapted to (Hamadene & Hdhiri 2006, Hamadène & Jeanblanc 2007). This framework can also take into account of real-world approximation of constraints (e.g. resource allocation in a hospital facility) referred to as costs of opening, running, and closing the activities.

Definition 2: Optimal switching

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space with filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{T}}$ and integer-valued finite times $\mathbb{T} = \{0, 1, \ldots, T\}$ for the discretization of the process. In addition, the setup includes the following problem-specific elements:

- a discrete set of operational modes $\mathbb{I} = \{1, 2, \ldots, q\}$, where $2 \leq q < \infty$. In this case we consider $q = 2$ with $\mathbb{I} = \{\text{on}, \text{off}\}$, which could represent "start/stop" of a treatment, "admission/non-admission" to ICU
- a final reward received at time T for being in mode $i \in \mathbb{I}$, which is modelled by an $\mathcal{F}_T$-measurable real-valued random variable Γ_i
- a running reward (payoff rate per unit of time) received while in mode $i \in \mathbb{I}$, which is modelled by a real-valued adapted process as a mapping $\Psi_i(t, X_t) : \Omega \times [0, T] \to \mathbb{R}$
- a cost for switching from mode $i \in \mathbb{I}$ to $j \in \mathbb{I}$, which is modelled by a real-valued adapted process $\gamma_{i,j} : \Omega \times [0, T] \to \mathbb{R}$ to cover for the extra costs due to the change of the regime.

A strategy α for the system will be a combination of two sequences:

- non decreasing sequence of $\mathbb{F}$-stopping times $(\tau_n)_{n\geq 1}$, $n \in \mathbb{N}\backslash\{0\}$, where at τ_n the production is switched from the current mode i to j. we also assume: $\tau_0 = t$ and $\tau_n \leq \tau_{n+1}$
- a sequence of indicators $(\iota)_{n\geq 1}$, $n \in \mathbb{N}\backslash\{0\}$, $\mathcal{F}_{\tau_n}$- measurable valued in $\mathbb{I}_q$. At time $t = \tau_n$ the system is switched from the current regime ι_{n-1} to ι_n, with $\iota_0 = i$.

We denote by $\mathcal{A}_{t,i}$ the set of admissible strategies to switch at time τ_n, $n \geq 1$, from the current regime ι_{n-1} to ι_n.

For any initial condition $(x, i) \in [0, T] \times \mathbb{I}_q$, and any control $\alpha = (\tau_n, \iota_n)_{n\leq 0} \in \mathcal{A}_{t,i}$, the total expected profit up to T for such strategy can be expressed as $J(\alpha; t, i)$. Given the costs and rewards elements listed above, the pofit for such strategy is:

$$J(\alpha : t, i) = \mathbb{E}\Big[\sum_{s=t}^{T-1} \Psi(X_t^{x,i}, I_t^i) + \Gamma - \sum_{n\leq 1} \gamma_{\iota_{n-1},\iota_n} \mathbf{1}_{\{\tau_n<T\}} | \mathcal{F}_n\Big] \tag{11}$$

The optimal switching problem is to maximize this expected total profit for all strategies

$\alpha \in \mathcal{A}_{t,i}$. For this purpose, we set the value function:

$$V_i(x) = \sup_{\alpha \in \mathcal{A}} J(\alpha : t, i) \qquad \forall \alpha \in \mathcal{A}_{t,i} \ \mathbb{P} \ a.s. \tag{12}$$

A switching control $\alpha^* \in \mathcal{A}_{t,i}$ is said to be optimal if it achieves the essential supremum in 12:

$$V_i(x) = J(\alpha^*; t, i) \geq J(\alpha; t, i) \qquad \forall \alpha \in \mathcal{A}_{t,i} \ \ \mathbb{P} - a.s. \tag{13}$$

We are interested in solving the optimal switching problem in 12 by following an analogous approach of Becker et al. (2019) and cast the problem as a stochastic control problem approximating the indicator function by a neural network.

4 Approximating the indicator function

We will follow the same approach proposed by Becker et al. (2019) to produce an estimate for the value in 12. This can be achieved in the following two steps.

4.1 Stopping decisions $0 - 1$

We reformulate the X-stopping times into a sequence of $0 - 1$ stopping decisions $f_n(X_n)$ for measurable functions $f_n : \mathbb{R}^d \to \{0, 1\}, n \in \mathbb{N}$. To optimally stop the Markov process X we make stopping decisions according to $f_n(X_n)$ and divide the problem 2 in a sequence of n stopping times:

$$V_n = \sup_{\tau \in \mathcal{T}_n} \mathbb{E}[g(\tau, X_\tau)] \tag{14}$$

where $\mathcal{T}_n$ denotes the family of all X-stopping times such that $n \leq \tau \leq N$.

Now we want to give an expression for the stopping times τ_n as a function of the $0 - 1$ stopping decisions $f_n(X_n)$. At $n = N$ there is, by construction, a terminal stopping decision where $\mathcal{T}_N = \{N\}$ because $f_N \equiv 1$. As a result the stopping time can be expressed as $\tau_n = N f_N$. Before maturity at time $0 \leq n < N - 1$, the stopping time can be written as:

$$\tau_n = \sum_{k=n}^{N} k f_k(X_k) \prod_{j=n}^{k-1} (1 - f_j(X_j)) \qquad \in \mathcal{T}_n \tag{15}$$

If on the one hand it is evident that τ_N is optimal for 2 when the process is at maturity, it should be proven that τ_n in 15 is optimal.

Theorem 1: Theorem 1

For every $f_n : \mathbb{R}^d \to \{0,1\}, n \in \mathbb{N}$ such that $\tau_n \in \mathcal{T}_n$ given by 15

4.2 Neural network

To approach the problem, we iteratively approximate the optimal stopping decisions $f_n : \mathbb{R}^d \to \{0,1\}, n = \{1,\dots,N-1\}$, by a neural network $f^\theta : \mathbb{R}^d \to \{0,1\}$ with parameter $\theta \in \mathbb{R}^q$. We choose $\theta_N \in \mathbb{R}^q$ such that $f_N^\theta \equiv 1$ and determine $\theta_n \in \mathbb{R}^q$ for $n \leq N-1$ by recursion of the form:

$$\tau_{n+1} = \sum_{m=n+1}^{N} m f^{\theta_m}(X_m) \prod_{j=n+1}^{m-1} (1 - f^{\theta_j}(X_j)) \tag{16}$$

Since f^θ takes values in $\{0,1\}$, hence not appropriate for a gradient-descent optimization method, the neural network includes a layer performing a logistic transformation such that we have a continuous output function $F^\theta : \mathbb{R}^d \to (0,1)$.

The Neural network includes $(d+40)$ hidden units and comprises a combination of linear functions $a_i^\theta(x) = W_i x + b_i$, rectified linear activation functions $\phi_{q_i}(x_1,\dots,x_{q_i}) = (x_1^+,\dots,x_{q_i}^+$, and a final logistic sigmoid function $\psi(x) = 1/(1+e^{-x})$, **??**. The parameters that are being tuned are $\theta = \{W_1, W_2, W_3, b_1, b_2, b_3\} \in \mathbb{R}^q$ and $q = q_1(d+q_2+1)+2q_2+1$.

At each time step, for each epoch we compute F^{θ_n} using the θ_n from the previous epoch. Then, the parameter θ_n is update via backpropagation by the gradient of the loss function (Adam optimization algorithm (Kingma & Ba 2014)), which is specified as:

$$Loss = -\mathbb{E}[g(n,X_n)F^{\theta_n}(X_n) + g(\tau_{n+1}, X_{\tau_{n+1}})(1 - F^{\theta_n}(X_n))] \tag{17}$$

The aim is to determine $\theta_n \in \mathbb{R}^q$ so that the negative of the loss function is close to the supremum $\sup_{\theta\in\mathbb{R}^q} \mathbb{E}[g(n,X_n)F^\theta(X_n) + g(\tau_{n+1}, X_{\tau_{n+1}})(1 - F^\theta(X_n))]$. From the stopping probabilities F^{θ_n}, we retrieve f^{θ_n} as hard decisions $\{0,1\}$.

Once the optimal θ^n is computed for all time steps n it is straightforward to obtain the desired estimate for the value function in **??**, given some assumptions on the hyperparameters and payoff function. In the table below **??**, the resulting estimated values, consistent with Becker et al. (2019), are shown under the following scenario:

$$x_0^1 = \{90, 100, 110\}, \quad K = 90, \quad \sigma = 20,\% \quad d = 10\%, \quad r = 5\%, \quad T = 3, \quad N = 9, \quad \text{for } i = \{2, 3, 5, 10, 20,$$

5 An application of the optimal switching problem

Building upon the problem formulation presented by Becker et al. (2019), the optimal switching architecture will adopt the same probabilistic setup and similar assumptions as per being a finite-horizon discrete-time model pricing a Bermudan call option.

5.1 *Training*

5.2 *Testing - Lower bound*

To approach the estimation of this iteration, we can start by providing a simulation of the process, hence assume the following forms for the various cost and benefit terms:

- the running benefit is the payoff of a Bermudan max-call option, as defined in the previous section 2 in the form of a power utility function

$$\Psi_i(x) = f_i(x)^{\kappa_i}, \qquad 0 < \kappa_i < 1 \tag{18}$$

- the terminal benefit is the payoff of a Bermudan max-call option at expiration date T
- the switching costs can be defined as:

$$\gamma_{0,1} = e^{-rt}\big(\max_{i\in\{1,\dots,\}} x^i - K)\big)^+ + \delta \qquad 0 < \delta < 1\gamma_{1,0} = -e^{-rt}\big(\max_{i\in\{1,\dots,\}} x^i - K)\big)^+ \tag{19}$$

Borrowing notation from section 2, we can identify the expression under the expectation in **??** as the discounted profit $G_t^i = g(x, \alpha)$. From here, we follow the same approach suggested by Becker et al. (2019) and shown above in 2. In a similar fashion we first assume that a switching mode occurs at termination to ease computation of the iteration. It is clear that the iteration term in **??**, Y_τ^j, requires the computation of Y when the system switches to the other regime, hence increasing the complexity of the space. Preliminary results can be found in

5.3 *Testing - Upper bound*

References

Ajdari, A., Niyazi, M., Nicolay, N. H., Thieke, C., Jeraj, R. & Bortfeld, T. (2019), 'Towards optimal stopping in radiation therapy', *Radiotherapy and Oncology* **134**, 96–100.

Becker, S., Cheridito, P. & Jentzen, A. (2019), 'Deep optimal stopping', *Journal of Machine Learning Research* **20**, 74.

Becker, S., Cheridito, P. & Jentzen, A. (2020), 'Pricing and hedging american-style options with deep learning', *Journal of Risk and Financial Management* **13**(7), 158.

Brennan, M. J. & Schwartz, E. S. (1985), 'Evaluating natural resource investments', *Journal of business* pp. 135–157.

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

Hamadene, S. & Hdhiri, I. (2006), 'On the starting and stopping problem with brownian and independant poisson noise', *Preprint Université du Maine, Le Mans* .

Hamadène, S. & Jeanblanc, M. (2007), 'On the starting and stopping problem: application in reversible investments', *Mathematics of Operations Research* **32**(1), 182–192.

Haugh, M. B. & Kogan, L. (2004), 'Pricing american options: a duality approach', *Operations Research* **52**(2), 258–270.

Kingma, D. P. & Ba, J. (2014), 'Adam: A method for stochastic optimization', *arXiv preprint arXiv:1412.6980* .

Knudsen, T. S., Meister, B. & Zervos, M. (1998), 'Valuation of investments in real assets with implications for the stock prices', *SIAM journal on control and optimization* **36**(6), 2082–2102.

Kohler, M., Krzyżak, A. & Todorovic, N. (2010), 'Pricing of high-dimensional american options by neural networks', *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics* **20**(3), 383–410.

Lapeyre, B. & Lelong, J. (2021), 'Neural network regression for bermudan option pricing', *Monte Carlo Methods and Applications* **27**(3), 227–247.

Liyanage, Y. W., Zois, D.-S., Chelmis, C. & Yao, M. (2019), Automating the classification of urban issue reports: an optimal stopping approach, *in* 'ICASSP 2019-2019 IEEE

International Conference on Acoustics, Speech and Signal Processing (ICASSP)', IEEE, pp. 3137–3141.

Longstaff, F. A. & Schwartz, E. S. (2001), 'Valuing american options by simulation: a simple least-squares approach', *The review of financial studies* **14**(1), 113–147.

Peskir, G. & Shiryaev, A. (2006), *Optimal stopping and free-boundary problems*, Springer.

Rao, A. & Jelvis, T. (n.d.), 'Foundations of reinforcement learning with applications in finance'.

Tsitsiklis, J. N. & Van Roy, B. (2001), 'Regression methods for pricing complex american-style options', *IEEE Transactions on Neural Networks* **12**(4), 694–703.

Zervos, M. (2003), 'A problem of sequential entry and exit decisions combined with discretionary stopping', *SIAM Journal on Control and Optimization* **42**(2), 397–421.