Solving the optimal stopping problem with reinforcement learning: an application in financial option exercise

1st Leonardo Kanashiro Felizardo
Dept. Electronic Systems
Escola Politécnica,
Universidade de São Paulo
São Paulo, Brazil
leonardo.felizardo@usp.br
0000-0002-2871-860X

2nd Elia Matsumoto
Sao Paulo School of Economics
Fundacao Getulio Vargas
São Paulo, Brazil
elia.matsumoto@fgv.br
0000-0003-1667-4221

3rd Emilio Del-Moral-Hernandez
Dept. Electronic Systems
Escola Politécnica,
Universidade de São Paulo
São Paulo, Brazil
emilio.delmoral@usp.br
0000-0003-4554-168X

Abstract—The optimal stopping problem is a category of decision problems with a specific constrained configuration. It is relevant to various real-world applications such as finance and management. To solve the optimal stopping problem, state-of-the-art algorithms in dynamic programming, such as the least-squares Monte Carlo (LSMC), are employed. This type of algorithm relies on path simulations using only the last price of the underlying asset as a state representation. In addition, the LSMC is designed for the valuation of options where risk-neutral probabilities can be employed to explain uncertainty. However, the general optimal stopping problem goals may not fit the requirements of the LSMC showing auto-correlated prices. We employ a data-driven method that uses Monte Carlo simulation to train and test artificial neural networks (ANN) to solve the optimal stopping problem. Using ANN to solve decision problems is not entirely new. We propose a different architecture that uses convolutional neural networks (CNN) to deal with the dimensionality problem that arises when we transform the whole history of prices into a Markovian state. We present experiments that indicate that our proposed architecture improves results over the previous implementations under specific simulated time series function sets. Lastly, we employ our proposed method to compare the optimal exercise of the financial options problem with the LSMC algorithm. Our experiments show that our method can capture more accurate exercise opportunities when compared to the LSMC. We have an outstandingly higher (above 974% improvement) expected payoff from these exercise policies under the many Monte Carlo simulations that used the real-world return database on the out-of-sample (test) data.

Index Terms—Reinforcement Learning, Least-squares Monte Carlo, Optimal stopping, Option pricing

I. INTRODUCTION

The optimal stopping problem is a relevant topic especially for option exercise (financial and real-options) [1]–[3]. The great variety of underlying assets and the distinct rules of each option created opportunities to explore many computational and mathematical tools for support. For instance, one crucial option rule is an early exercise option, such as the American options style. This class of options introduces the optimal stopping problem in the option exercise context and opens the doors for many mathematical and computational techniques to solve it [4]–[9]. The goal is always to find a policy that can make the best possible stop decision. Many decision problems can be framed as optimal stopping problems. The most famous application is the optimal exercise of an option to maximize an arbitrated utility function. The stopping problem can be interpreted as an exit or entry decision in an investment, such as a buy/sell action or an exercise action. In this context, it is common to deal with market uncertainties using real options, giving flexibility for investment and divestment decisions. Other applications can be many folds, such as financial market [1], [6], [10], inventory/production management [11], [12], investment decisions on technology/products [13], reservoir management for hydropower [14], real state [15], mergers and acquisitions [16], venture capital [17], advertising [18] and environmental compliance [19] (see [20] for more details).

Optimal stopping problems in finance usually assume a time series behavior, such as the geometric Brownian motion correctly describing the underlying asset dynamics. Methods such as the least-squares Monte Carlo (LSMC) [1] use path simulations to find optimal stopping points. According to [1], the LSMC aims to provide a pathwise approximation to an optimal stopping rule to maximize the value of the American Option. Many other articles formerly also employed the simulation approach [21]–[24] for American options pathwise simulation, employing the stratification and parameterization techniques to approximate the transition state density function or the exercise boundary. Although the conditional expectations based on the approximated policies are nonparametric, the simulated path used for the regression is not. The predominant model is the geometric Brownian motions that assume that the returns follow a normal distribution to generate the simulated paths. Each price is a Markov state, containing all the information for the subsequent returns. Assuming a Markovian framework is a practical modeling decision, it does not necessarily reflect the reality of many other problems. Studies such as [6], [8], [25] provide examples of how machine learning can help to solve the optimal decision problem under non-Markovian structures.
The main advantage is the non-dependency of multiple simulations and adjustments of the path generation model over the real data. An advantage of the ANN is the universal functional approximation capability [26]. Despite the mathematical proofs of the approximation capabilities [27], the use of better architectures can lead to improvement in specific task performances. The real-world probabilities may be estimated in some cases, as we further see, leading to better estimation of optimal stopping points. Finally, ANN can solve high-dimensional problems, sometimes when converting a non-Markovian state into a Markovian one. The use of more optimized architectures as the convolutional neural networks (CNN) [28], [29] brings feature extraction capabilities that help overcome the dimensionality problem. By using a set of adaptive parameters (filter), the CNN specializes in pattern recognition by learning the kernel coefficients using the back-propagation method [30]. This pattern recognition, which results from the adaptive filters, can synthesize the state with the relevant information for an ANN to approximate exercise policies.

In this context, this paper proposes using convolutional neural networks to solve the optimal exercise problem in the option domain by using the learning procedures described in [6]. In their work, Becker et al. describe a method to train a neural network to predict an optimal stopping time sequence recursively produced by running a stochastic gradient ascent procedure.

We summarize the contributions of this paper as follows:

- Improvements on the original implementation of [6] addressing the dimensionality problem for a broader category of time series functions.
- Proposing a new approach to optimally exercise options in the context of financial options.
- Increasing the optimal stopping performance in the real-world data by transforming the historical information in a Markov state with the feature extraction of the CNN layer.
- We show that, by using a universal function approximation approach such as ANN, we can avoid the calibration process that can carry functional biases for optimal stop policies.

In Section II, we define our problem, and propose the solution for the problem in Section III. We conduct our experiments in Section IV and show the results. Finally, in Section V, we present the conclusions and considerations for future works.

II. PROBLEM DEFINITION

Financial institutions create different types of derivatives to hedge risks related to asset price fluctuation. A derivative is a contract with an underlying asset. That contract can have many formats, and one of them provides the investor the option to buy or sell the underlying asset for a predetermined price. This type of contract is called an option. The option can be the exercise for a positive payoff, or non-exercised, giving the investor zero value payoff. The payoff of an option depends on the type, put (right to sell in the future) or call (right to buy in the future). The optimal stopping, or optimal exercise problem, occurs when the investor has to decide the best moment to execute the option to maximize the payoff.

Furthermore, there are different options mechanisms. In the European-style options, the investor cannot early exercise the option, only at maturity. There is the case of the American-style option, whereby the investor can early exercise. For the American (index $a$) and European (index $e$) options, to calculate the payoff, we compare the strike price ($K$), the previously agreed price in the contract, and the asset price ($S$). To calculate the value of a European-style call option ($C$), we can use the Equation (1):

$$C_T^e = \max \{S_T - K, 0\}, \quad (1)$$

which is the maximum payoff in a period $T$ (maturity date). Analogous to the call option we can calculate the put option value using the Equation (2):

$$P_T^e = \max \{K - S_T, 0\}. \quad (2)$$

Unlike European-style, we can early exercise the American-style call option. If we consider an American-style call option, it is never optimal to exercise a non-dividend-paying stock early. Consequently, in the scenario of non-dividends, the price of a European call option is the same as the American call option ($C_T^e = C_T^a$). In the case of dividend payments, it may be optimal to exercise the option as it goes ex-dividend (the day that the dividend is announced). In turn, in the case of the American-style put options, as in Equation (3) the value of the put option cannot be less than the intrinsic value,

$$P_T^a \geq K - S_T, \quad (3)$$

it can be early exercised. While the intrinsic value is larger than the continuation value, we should not exercise the put option. Unlike the call option, the American-style put option has the optimal stopping point. Since the value of a stock is not negative, for European-style options, we have an upper bound defined by the present value of the strike price $PV(K) = Ke^{r(T-t)}$ and for the American-style put, an upper bound equals $K$ because American-style option can be exercised at any moment. The difference between the price of the European-style put option and the American-style put option is the option time-value.

Consequently, the problem is to find the critical point where the intrinsic value surpasses the continuation value of an American-style put option. In other terms, we want to find the point where the time-value is below zero. Finding the critical point is a challenging problem since we do not know the stock price movement. To find the optimal stopping point, we require finding the boundary region to exercise the option, which can be solved as an optimization problem. This optimization problem gets more complicated as we increase the complexity of the option mechanism, including more underlying asset and complex rules that govern the put price.
One popular option contract used in the market is the Bermudan-style option, in which it is only possible to exercise the option at pre-determined times. The Bermudan-style option is a discretization of the American-style options. As a method of approximating the market decision process, discretization of time can be a useful approach. We aim to solve the optimal stopping problem for the Bermudan-style options using artificial neural networks to approximate a function that solves the decision process. By analyzing state variable $s$, composed of the underlying asset price returns, interest $r$, strike price $K$, and payoff $p$, the model returns a decision (action) $a$ at each decision step of exercise or not the option. The optimal stopping problem of the form $\sup_{\pi} \mathbb{E} g(t, s_t)$ where $s = (s_n)_{n=0}^N$ where $N$ is the number of time-steps, $s \in S$ is an state, and $t \in T$ is the stopping time in the set of all stopping times. The definition given covers the case of Bermudan options in which we have finite stopping decisions with finite horizons. However, this definition is also relevant for the continuous case (such as American-style options) since its time-discretized version can approximate it.

The Brownian motion is one common assumption for the movement of asset prices in financial markets. For many reasons, it is convenient, i.e., option pricing. In this work, we are concerned only with the stopping problem from the perspective of the option owner. Therefore, we want to maximize the return when exercising the option and not calculate the option’s value. We try to approximate the real probabilities function from the probability space $(\Omega, \mathcal{F}, P)$, and horizon $[0, T]$, with $\Omega$ the sample space, $\mathcal{F}$ the sigma field of distinguishable events at time $T$, $P$ the probability function for each event in the event space. When solving the stopping problem, if the increments are correlated by following an underlying function, we need to keep track of all past movements to have a Markov state. The dimension of the observed variables is equal to the number of past time-steps $N$ used to approximate the continuous-time. We may suffer from a dimensionality problem as we try to solve the problem for longer time horizons with more precise time-discretization. Our proposed method deals with the dimensionality problem when we try to bring to the optimal stopping policy occurs iteratively starting from the terminal stopping decision $\pi_0(s_N) \equiv 1$, being the stopping policy approximated by ANN $\pi_0 : \mathbb{R}^d \rightarrow \{0, 1\}$. This process proceeds backward in time, from the terminal state to the first state. The neural network approximate try to approximate the optimal stopping policy $\pi^*$ by the $\pi_0$ which generates the stopping decision $[0, 1]$ by the forward computation where we get the output $v$ of the ANN by

$$v_j^{(l)}(s) = \sum_i w_j^{(l)}(s)y_i^{(l-1)}(s)$$

where $l$ is the neural network layer, $y_i$ is the output of neuron $i$ previous layer, $w$ is synaptic weights that compose parameters $\theta$ with bias $b$, $y$ is the result of the last output

$$y_j^{(l)} = \varphi_j(v_j(n))$$

where $\varphi$ is the activation function, such as the ReLU or sigmoid. The weights of the network are updated using the loss described in Equation 4.

$$w_j^{(l)}(n+1) = w_j^{(l)}(n)+\alpha \left[ \Delta w_j^{(l)}(n-1) \right] + \eta\delta_j^{(l)}(n)y_j^{(l-1)}(n)$$

where $\delta$ is the local gradient, $\eta$ is the learning rate and $\alpha$ is the momentum (check [31] Chapter 4 for more details). Under the assumption of a geometric Brownian motion simulation path, we would not require using the whole history of the price time series to generate the path simulation. n policy $\pi$ has parameters $\theta$ When we enter a series that can be auto-correlated, we require the whole history of prices to estimate the optimal stopping point. Fortunately, [6] brief explores this possibility using the fraction Brownian motion to generate the
time series and adapts his implementation to receive the whole price time series as input.

The main problem is that the multilayer perceptron architecture (MLP) has limitations related to overfitting the noisy signal. Up to this point, we focused on the multilayer perceptrons (MLP) formulation, but we employed a particular class of ANN called convolutional neural networks (CNN) [28]. Initial studies conceived the CNN to recognize two-dimensional shapes with invariance to translation, skewing, scaling, and other distinct distortions. As it is done for the MLP, the CNN learn in an supervised manner following the constrains on the feature extraction, feature mapping and subsampling [30]. We depict the proposed architecture for the optimal exercise problem in Fig. 1. First, we have textbfinput the set of time series considered to estimate the optimal stopping point where the first hidden layer performs convolution. The first hidden layer consists of six feature maps, with each feature map comprising 23 neurons, each neuron with a receptive field of size 3. The second hidden layer is similar to the first but with 22 neurons and a receptive size of 3. The third layer performs a reshape operation to serve as input for the subsequent MLP. The fourth and the fifth layers are dense layers, also called fully connected layers, with 50 neurons each. Our proposed architecture makes it possible to include multiple time series features as input, including each time series in a different channel, dealing with multiple time series as inputs.

Algorithm 1: Training the CNN for the optimal stopping decision problem

```plaintext
input : S0, d, K, N, r, σ, δ
output: Trained parameters: w
1 initialization: model NN parameters: w0, max number of episodes: max_ep , batch size: b, choose the type of data generation: gen, max_price = 0
2 foreach episode in range(0, max_ep) do
  3 X, p, r, K ← gen(b, S0, d, K, N, r, σ, δ);
  4 gt ← p[:, -1];
  5 s = concatenate(X, p, r, K);
  6 hist_a = [];
  7 foreach step in range(N - 2, -1, -1) do
    8 a ← NN(w, s[step]);
    9 hist_a.append(a);
    10 loss = mean_a(p × a - gt × (1 - a));
    11 gt = where(a > 0, p, gt);
    12 opt_price ← mean(gt);
    13 w ← BackPropagation(w, loss);
```

A. The baseline method: LSMC

To compare our proposed method in a real-world case, we use the LSMC, a standard baseline method, to solve the optimal stopping problem. The LSMC is a mathematical approach based on simulation to approximate the value of the American-style option. In the LSMC method, we regress the ex-post realized payoffs from continuation on functions of the values of the state variables. With the conditional expectation function, we can obtain the optimal exercise strategy along each path [1].

We can list a few reasons why the LSMC maybe not be appropriate for the real-option or financial option optimal exercise problem:

- The LSMC tries to approximate the conditional expectation using a linear regression which can be inadequate since we are trying to approximate an arbitrary function that can be highly non-linear.
- In a non-Markovian model, we may have to project a high dimensional state variable into a vector of parameters β with hundreds of entries.
- Monte Carlo methods are model-dependent (model-based), which means we need to know the Markov decision process (MDP) to optimize the action. Sometimes we may not know the MDP, or it may be too complex to model easily.

Despite these arguments, the LSMC is well known for its broad applicability; hence, it can be considered a reasonable benchmark.

IV. EXPERIMENTAL VERIFICATION

We divide the experiments into two phases: First, we compare our approach against the [6]'s approach for different time series generation functions assuming that we are dealing with the case in which the asset price return does not depend only on the last asset price; In the second batch of the test, we use our proposed approach to solve the optimal exercises/stopping problem for real-option with the underlying asset being an energy market asset. To reproduce the results, we provide the code in: https://github.com/leokan92/optimal-stopping-cnn. We implemented all the algorithms using the Python language with Pytorch and executed it in a PC containing an AMD Ryzen 5 5600X 6-Core processor 3.70 GHz, an NVIDIA GeForce RTX 3060 GPU, and 32GB RAM.

A. Analysis for the path simulation approach

In the first batch of experiments, we compare our CNN approach to [6]'s approach using three different path generations: geometric Brownian motion, fractional Brownian motion, and harmonic composition with Gaussian noise. We calculate the average payoff in 32600 Monte Carlo simulations (paths). We specify the simulation parameters for each generation technique in Table I,

| Generator   | Parameters |
|-------------|------------|
| Geometric   | S0 = 120; K = 100; T = 3; r = 0.05; | |
| Brownian motion | δ = 0.1; σ = 0.1 | |
| Fractional Brownian motion | S0 = 120; K = 100; T = 3; r = 0.05; | |
| Gaussian noise | σ = 0.1; h = 0.7 | |
| Harmonic + Gaussian noise | S0 = 120; K = 100; T = 3; r = 0.05; ampl = 0.2; freq1 = 0.3; freq2 = 2 | |

TABLE I: Simulation generator parameters
where \( S_0 \) is the initial asset price, \( K \) is the strike price, \( T \) is the number of years, \( r \) is the risk-free rate, \( \delta \) is the dividend rate payment, \( \sigma \) the constant volatility value, \( h \) is the Hurst parameter used for fractional Brownian motion, \( ampl \) is the amplitude of both harmonics, and \( freq \) is the frequency, that is different for each harmonic. We explore how the exercise steps affect the overall results of both methods. Fig. 2d, Fig. 2a, and Fig. 2c depict how the average payoff evolves comparing our CNN approach against Becker’s original implementation. The comparison uses three different price series generation techniques.

Fig. 2d shows that our method can achieve the same performance as the Becker implementation. In the Fig. 3 note that the training performance for both Becker’s and our implementation. Note that both algorithms stabilize in a few epochs when dealing with the geometric Brownian motion series and also that the exercise occurs mainly on the last exercise opportunity. The exercise concentrates on the last decision step if we choose a specific set of path simulation parameters. The Brownian motion parameters and exercise regions can be better explored (as in many other studies and is a branch of research [32], [33]) to understand why the networks choose some of the early exercise points. We did not notice any improvement or decrease in the training speed and average payoff by using the CNN approach for the Brownian motion simulation.

We also employed the fractional Brownian motion to compare the Hurst parameter equal to 0.7. In Fig. 2c we observe that the difference between both approaches (ours and the baseline) is not significant. In Fig. 2c, instead of the standard deviation, we use a 95% confidence interval, which is better to visualize since the standard deviation is significant and distorts the scale. When analyzing the training performance evolution by epoch, we do not perceive any improvement or decrease in learning speed.

In Fig. 2a, note that as the dimensionality grows, the more prominent is the convergence speed difference between our approach and Becker’s. Furthermore, Fig. 3 evidences that CNN implementation can achieve higher performances and much faster than the previous implementation. We recommend using our method to evaluate optimal stopping points for auto-correlated time series by covering a more extensive group of functions.

**B. Optimal stopping for real options**

In this part of the experiments, we simulate the returns based on the returns present in real-world asset returns data. We obtained the day-frequency price returns for the crude oil asset from NYMEX futures prices. We divide the data set into in-sample, with 7568 data points, and out-of-sample (testing) data for evaluation, with 1915 data points. In the in-sample data, we have a training and validation data division to control the overfitting of our model. We have a 0.7/0.3 proportion between training and validation data for the training set, giving 5630 points for training and 1764 data points for validation.

We employ the same framework used for the simulated data to perform the experiments for real-world data. The main difference is that all the price returns are generated but randomly picked from real-world data. We use these real-world returns to train our network instead of the simulated data. We build each path used for training given by \( \{S_0, S_0\rho_N^p, S_0\rho_{N-1}^p, ..., S_0\rho_1^p\} \), where \( i \) is a random number between the training max size minus \( N \) and 0, and \( \rho^p \) is the relative return. We defined the required parameters of the experiments as: \( S_0 = 100; K = 100; r = 0.05; N = 100 \).

In the first set of experiments, we test different exercise opportunity sizes and measure the average payoff in the testing data. We used the LSMC as a baseline model to evaluate our
TABLE II: Results with the returns statistics (mean, standard deviation) and the respective results of our proposed model and the baseline (LSMC) grouped by sector.

| Economic Sector | Mean Return | CNN Expected Payoff | LSMC Expected Payoff | % Improvement |
|-----------------|-------------|---------------------|----------------------|--------------|
| Communications  | -1.31e-03±0.59 | 4.54±0.55           | 0.73±1.37            | 668±40       |
| Consumer Discretionary | 2.79e-03±0.61 | 7.59±1.05           | 0.83±1.51            | 944±69       |
| Energy          | -7.90e-04±1.04 | 6.30±0.77           | 0.96±1.73            | 679±45       |
| Financial       | 2.31e-02±0.13  | 6.74±0.66           | 0.60±1.30            | 1134±51      |
| Materials       | 9.95e-03±0.89  | 10.88±1.33          | 1.38±2.39            | 833±57       |
| Technology      | 3.52e-02±1.01  | 14.35±1.35          | 0.93±1.84            | 1586±73      |

The proposed method solves the optimal stopping. The LSMC uses the discounted values to decide the optimal stopping point and fit the polynomial, unavailable in the out-of-sample data. Therefore, there is a gap in fairly comparing our proposed method and the LSMC. Suppose we use the out-of-sample data to fit the polynomial. In that case, we will be giving a fictional privileged scenario to the LSMC to find the optimal stopping points, even though we test in this scenario. The hypothesis is that we have underlying hidden functions in the time series, requiring more than just the last price and payoff to estimate the optimal stopping point. In other terms, that is a non-Markov state that we try to transform into a Markov state by giving history. The line graph presented in Fig. 2b shows the maximum theoretical stopping point, our approach, and the LSMC calculated stopping point using out-of-sample data.

We also tested the LSMC using the training data to approximate the polynomials and test these approximated polynomials on the testing data. As expected, the results show worse results than the LSMC as the test performed out-of-sample data directly. We used the LSMC trained under in-sample data for the following tests and tested for out-of-sample data.

We tested this optimal stopping problem with put options for any underlying asset to extend the results on financial options. In this batch of experiments, we try to create a proxy for the possible underlying asset under different scenarios of financial options using stock prices from different economic sectors (communications, consumer discretionary, energy, financial, materials, and technology). For each economic sector, we selected five assets and divided them into an in-sample...
(training and validation) set with 3779 data points and an out-of-sample (testing) set with 1620 data points. Here, we try to show that despite the nature of the underlying stock, our method can, in most cases, outperform the LSMC method for optimal stopping decisions. Considering the mean return as the baseline, we have an average 974% increase, more than ten times improvement. Although this seems to be very high, since the mean return is usually a number next to zero, any improvement will be significant in relative terms. A relative performance improvement metric aims to present robustness in our results, not necessarily focusing on the magnitude of the performance increase. Table III shows the summary of results from our method and the baseline in different assets for distinct sectors. We can already perceive a significant increase in the average expected payoff from the table by using the CNN approach to outperform the LSMC in all the assets. The asset with the highest payoffs is also the underlying asset on which both models had the best performances. Table II we can extend the analysis to an economic sector comparison, which maintains the previous analysis. The Table II groups the results from Table III by calculating the mean value. The maximum expected payoff of our approach occurs for the material sector, not the sector with the highest asset returns.

V. CONCLUSIONS

In this article, we proposed a direct policy approximation based on [6]’s architecture but using a better ANN to overcome problems related to high dimensional state spaces. We compared and improved the original baseline model using synthetically generated time series using customized and improved ANN techniques. We applied this solution to a real-world set of problems showing the potential to better exercise real and financial options achieving outstanding results (974% increase, more than ten times improvement) compared with the literature performance baseline. This result can be extended to any problem related to optimal stopping in the financial and non-financial literature. The main gain of our proposed model may be in the realm of real options where the underlying asset has a behavior that can be predicted but is sometimes non-linear, following a wide range of possible functions. As an extension of this work, we propose using this same architecture in the optimal execution problem using the historical prices and the whole history book of negotiations. Our approach seems to be a powerful way to approximate policy function in the finite horizon problem of optimal stopping. Therefore, future researchers can use this algorithm to solve a broader set of problems.

ACKNOWLEDGMENT

This research was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES - Coordination for the Improvement of Higher Education Personnel, Finance Code 001, grant 88882.333380/2019-01), Brazil. The paper has benefited from discussions with Professor Paolo Brandimarte from the Politecnico di Torino.

REFERENCES

[1] F. A. Longstaff and E. S. Schwartz, “Valuing American Options by Simulation: A Simple Least-Squares Approach,” *The Review of Financial Studies*, vol. 14, no. 1, pp. 113–147, 01 2001.
[2] M. Hanfeld and S. Schlüter, “Operating a swing option on today’s gas markets – how least squares monte carlo works and why it is beneficial,” *Zeitschrift für Energiewirtschaft*, vol. 41, no. 2, pp. 137–150, Apr. 2017.
[3] S. Nadarajah, F. Margot, and N. Secomandi, “Comparison of least squares monte carlo methods with applications to energy real options,” *European Journal of Operational Research*, vol. 256, no. 1, pp. 196–204, 2017.
[4] G. Peskir and A. Shiryaev, *Optimal stopping and free-boundary problems*. Basel: Birkhäuser Basel, 2006, pp. 123–142.
[5] D. Egloff, ”Monte Carlo algorithms for optimal stopping and statistical learning.” *The Annals of Applied Probability*, vol. 15, no. 2, pp. 1396 – 1432, 2005.
[6] S. Becker, P. Cheridito, and A. Jentzen, “Deep optimal stopping,” *Journal of Machine Learning Research*, 2019.
[7] J. Ruf and W. Wang, “Neural networks for option pricing and hedging: A literature review,” *SSRN Electronic Journal*, 2019.
[8] K. Andersson and C. W. Oosterlee, “A deep learning approach for computations of exposure profiles for high-dimensional bermudan options,” *Applied Mathematics and Computation*, vol. 408, p. 126332, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0096300321004215
[9] S. BECKER, P. CHERIDITO, A. JENTZEN, and T. WELTI, “Solving high-dimensional optimal stopping problems using deep learning,” *European Journal of Applied Mathematics*, vol. 32, no. 3, p. 470–514, 2021.
[10] M. Broadie, P. Glasserman, and Z. Ha, *Pricing American Options by Simulation Using a Stochastic Mesh with Optimized Weights*. Boston, MA: Springer US, 2000, pp. 26–44.
[11] K. D. Miller, “Knowledge inventories and managerial myopia,” *Strategic Management Journal*, vol. 23, no. 8, pp. 689–706, 2002.
[12] W.-K. C. Xinmin Huang, Na Song, “A real option approach to optimal inventory management of retail products,” *Journal of Industrial & Management Optimization*, vol. 8, no. 2, pp. 379–389, 2012.
[13] E. Pennings and O. Lint, “The option value of advanced r & d,” *European Journal of Operational Research*, vol. 103, no. 1, pp. 83–94, 1997.
[14] S. Steinhschneider and C. Brown, “Dynamic reservoir management with real-option risk hedging as a robust adaptation to nonstationary climate,” *Water Resources Research*, vol. 48, no. 5, 2012.
[15] L. Bulan, C. Mayer, and C. T. Somerville, “Irreversible investment, real options, and competition: Evidence from real estate development,” *Journal of Urban Economics*, vol. 65, no. 3, pp. 237–251, 2009.
[16] D. Hackforth and E. Morellec, “Stock returns in mergers and acquisitions,” *The Journal of Finance*, vol. 63, no. 3, pp. 1213–1252, 2008.
| Economic Sector | Asset | Mean Return | CNN Expected Payoff | LSMC Expected Payoff | % Improvement |
|-----------------|-------|-------------|---------------------|---------------------|---------------|
| Communications  | BCE   | 0.002±0.50  | 3.07±0.32           | 0.37±0.84           | 826±38        |
| Communications  | VZ    | 0.006±0.61  | 3.36±0.36           | 0.79±1.30           | 424±28        |
| Communications  | T     | -0.003±0.47 | 3.55±0.40           | 0.66±1.26           | 539±32        |
| Communications  | IPG   | 0.008±0.41  | 6.67±0.81           | 0.64±1.27           | 1042±64       |
| Communications  | DISH  | -0.019±0.94 | 6.07±0.85           | 1.19±2.16           | 510±39        |
| Consumer Discretionary | ABEV | -0.002±0.10 | 4.97±0.55           | 1.15±1.89           | 433±29        |
| Consumer Discretionary | LEN | 0.030±1.34  | 9.65±1.20           | 0.74±1.43           | 1310±84       |
| Consumer Discretionary | F    | -0.003±0.20 | 7.18±1.07           | 0.67±1.23           | 1078±87       |
| Consumer Discretionary | BTI  | -0.009±0.75 | 3.97±0.48           | 0.58±1.02           | 687±47        |
| Consumer Discretionary | GPS  | -0.006±0.66 | 12.20±1.96          | 1.01±1.97           | 1212±100      |
| Energy | BP   | -0.005±0.61 | 4.85±0.65           | 0.66±1.32           | 736±49        |
| Energy | COP  | -0.002±1.17 | 7.83±1.00           | 1.18±2.15           | 663±47        |
| Energy | CVX  | 0.002±1.80  | 5.07±0.63           | 0.61±1.27           | 827±49        |
| Energy | COG  | -0.008±0.50 | 5.05±0.65           | 1.18±1.95           | 427±33        |
| Energy | LNI  | 0.010±1.14  | 8.68±0.92           | 1.17±1.96           | 741±47        |
| Financial | ACNB | 0.005±0.70  | 6.12±0.63           | 0.70±1.61           | 875±39        |
| Financial | ORI  | 0.007±0.31  | 5.80±0.59           | 0.53±1.18           | 1097±50       |
| Financial | RE   | 0.047±3.56  | 5.66±0.48           | 0.51±1.05           | 1104±46       |
| Financial | AXP  | 0.045±1.87  | 7.72±0.69           | 0.56±1.23           | 1375±56       |
| Financial | C    | 0.012±1.19  | 8.40±0.91           | 0.69±1.42           | 1219±64       |
| Materials | DD   | 0.010±1.33  | 6.76±0.83           | 0.92±1.67           | 738±50        |
| Materials | BHP  | 0.016±0.95  | 9.37±0.93           | 1.86±3.13           | 504±30        |
| Materials | SCCO | 0.020±0.89  | 11.60±1.17          | 0.99±1.72           | 1172±68       |
| Materials | OLN  | 0.014±0.64  | 15.64±2.23          | 1.42±2.51           | 1103±89       |
| Materials | MOS  | -0.009±0.66 | 11.03±1.46          | 1.70±2.94           | 648±50        |
| Technology | INTC | 0.013±1.01  | 7.38±0.73           | 0.63±1.32           | 1168±55       |
| Technology | HPQ  | 0.007±0.40  | 9.61±0.99           | 0.67±1.47           | 1428±67       |
| Technology | AMD  | 0.051±1.19  | 31.41±3.01          | 2.16±4.01           | 1452±75       |
| Technology | LOGI | 0.070±1.16  | 15.69±1.25          | 0.60±1.22           | 2599±102      |
| Technology | ARW  | 0.037±1.26  | 7.68±0.80           | 0.60±1.19           | 1281±67       |

[17] Y. Li, “Duration analysis of venture capital staging: A real options perspective,” Journal of Business Venturing, vol. 23, no. 5, pp. 497–512, 2008.

[18] B. Kamrad, S. S. Lele, A. Siddique, and R. J. Thomas, “Innovation diffusion uncertainty, advertising and pricing policies,” European Journal of Operational Research, vol. 164, no. 3, pp. 829–850, 2005, recent Advances in Scheduling in Computer and manufacturing Systems.

[19] G. Cortazar, E. S. Schwartz, and M. Salinas, “Evaluating environmental investments: A real options approach,” Management Science, vol. 44, no. 8, pp. 1059–1070, 1998.

[20] D. M. Lander and G. E. Pinches, “Challenges to the practical implementation of modeling and valuing real options,” The Quarterly Review of Economics and Finance, vol. 38, no. 3, Part 2, pp. 537–567, 1998.

[21] P. Bossaerts, “Simulation estimators of optimal early exercise,” Graduate School of Industrial Administration, 1989.

[22] J. A. Tilley, “Valuing american options in a path simulation model,” Transactions of the Society of Actuaries, 1993.

[23] J. F. Carriere, “Valuation of the early-exercise price for options using simulations and nonparametric regression,” Insurance: Mathematics and Economics, vol. 19, no. 1, pp. 19–30, 1996.

[24] M. Broadie and P. Glasserman, “Pricing american-style securities using simulation,” Journal of Economic Dynamics and Control, vol. 21, no. 8, pp. 1323–1352, 1997, computational financial modelling.

[25] L. Goudenège, A. Molen, and A. Zanette, “Machine learning for pricing american options in high-dimensional markovian and non-markovian models,” Quantitative Finance, vol. 20, no. 4, pp. 573–591, 2020.

[26] V. Kůrková, “Kolmogorov’s theorem and multilayer neural networks,” Neural Networks, 1992.

[27] G. Cybenko, “Approximation by superpositions of a sigmoidal function,” Mathematics of Control, Signals, and Systems, 1989.

[28] D. H. Hubel and T. N. Wiesel, “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex,” The Journal of physiology, vol. 160, no. 1, pp. 106–154, Jan 1962. 14449617[pmid]. [Online]. Available: https://doi.org/10.1113/jphysiol.1962.sp006837.

[29] K. Fukushima, “Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position,” Biological Cybernetics, vol. 36, no. 4, pp. 193–202, Apr 1980.

[30] Y. LeCun and Y. Bengio, Convolutional Networks for Images, Speech, and Time Series. Cambridge, MA, USA: MIT Press, 1998, p. 255–258.

[31] S. S. Haykin, Neural networks and learning machines, 3rd ed. Upper Saddle River, NJ: Pearson Education, 2009.

[32] S. Villeneuve, “Exercise regions of american options on several assets,” Finance and Stochastics, vol. 3, no. 3, pp. 295–322, May 1999. [Online]. Available: https://doi.org/10.1007/s007800050064.

[33] T. L. Lai and T. W. Lim, “Exercise regions and efficient valuation of american lookback options,” Mathematical Finance, vol. 14, no. 2, pp. 249–269, 2004.