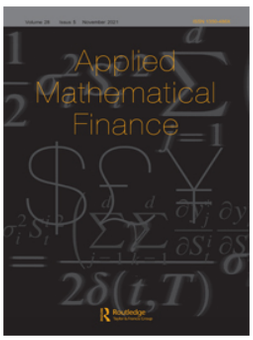


Routledge
Taylor & Francis Group

Applied Mathematical Finance

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ramf20

On a Neural Network to Extract Implied Information from American Options

Shuaiqiang Liu, Álvaro Leitao, Anastasia Borovykh & Cornelis W. Oosterlee

To cite this article: Shuaiqiang Liu, Álvaro Leitao, Anastasia Borovykh & Cornelis W. Oosterlee (2021) On a Neural Network to Extract Implied Information from American Options, Applied Mathematical Finance, 28:5, 449-475, DOI: 10.1080/1350486X.2022.2097099

To link to this article: https://doi.org/10.1080/1350486X.2022.2097099

Published online: 21 Jul 2022.

APPLIED MATHEMATICAL FINANCE
2021, VOL. 28, NO. 5, 449–475
https://doi.org/10.1080/1350486X.2022.2097099

Routledge
Taylor & Francis Group

On a Neural Network to Extract Implied Information from American Options

Shuaiqiang Liu [a], Álvaro Leitao [b,c], Anastasia Borovykh[d] and Cornelis W. Oosterlee [a,e]

[a]Applied Mathematics (DIAM), Delft University of Technology, Delft, The Netherlands; [b]Department of Mathematics, University of A Coruña, A Coruña, Spain; [c]CITIC research centre, University of A Coruña, A Coruña, Spain; [d]Imperial College London, London, UK; [e]Centrum Wiskunde & Informatica, Amsterdam, The Netherlands

ABSTRACT

Extracting implied information, like volatility and dividend, from observed option prices is a challenging task when dealing with American options, because of the complex-shaped early-exercise regions and the computational costs to solve the corresponding mathematical problem repeatedly. We will employ a data-driven machine learning approach to estimate the Black-Scholes implied volatility and the dividend yield for American options in a fast and robust way. To determine the implied volatility, the inverse function is approximated by an artificial neural network on the effective computational domain of interest, which decouples the offline (training) and online (prediction) stages and thus eliminates the need for an iterative process. In the case of an unknown dividend yield, we formulate the inverse problem as a calibration problem and determine simultaneously the implied volatility and dividend yield. For this, a generic and robust calibration framework, the Calibration Neural Network (CaNN), is introduced to estimate multiple parameters. It is shown that machine learning can be used as an efficient numerical technique to extract implied information from American options, particularly when considering multiple early-exercise regions due to negative interest rates.

ARTICLE HISTORY
Received 18 March 2022
Accepted 13 June 2022

KEYWORDS
Machine learning; American options; implied volatility; computational finance; negative interest rates

1. Introduction

So-called implied financial information, which is obtained from financial derivatives prices, is useful information for risk management, for the valuation of other financial derivatives, for hedging, and forecasting (Christoffersen, Jacobs, and Young Chang 2013; Won Seo and Sik Kim 2015). Different from historical volatility (which is computed from past known asset prices), implied volatility (which is computed from the market option prices) reflects the market implied uncertainty in the underlying asset prices. Similarly, implied dividend can be seen as a measure which indicates how much market participants expect an asset price will be reduced under the so-called pricing, or risk-neutral, measure. Compared to implied volatility, the implied dividend is typically a relatively weak

CONTACT Cornelis W. Oosterlee c.w.oosterlee@uu.nl

signal, but there are several applications for the implied dividend information. An accurate implied dividend yield may result in accurate theoretical option values, Greeks and implied volatilities. One may choose a specific trading strategy according to the difference between the market implied and announced actual dividends, see Hull (2019). The authors in Christoffersen, Jacobs, and Young Chang (2013), Bilson, Baum Kang, and Luo (2015), and Fodor, Stowe, and Stowe (2017) give evidence for the fact that implied dividends are a significant factor for forecasting actual dividend changes. In other words, implied dividend is shown to have important predictive power compared to historical dividends.

American options, i.e., options with early-exercise features, are commonly traded. The underlying asset may be any financial asset, such as currencies, commodities, bonds and stocks. With American options, the holder has the right (but not the obligation) to exercise the contract at any time before the contract's expiry, whereas a European option can only be exercised at the expiry time. The main difficulty is resolving the well-known free boundary due to the unknown early-exercise region. Valuation of American options requires more intensive computation. Computing the implied volatility and implied divided from observed European option prices has often been addressed in the literature, for example, in Brenner and Subrahmanyam (1988), Corrado and Miller (1996), Chance (1996), and Hull (2019). The computation of American option prices is generally more expensive than pricing European options, because of the early-exercise features. Deriving the implied information from American option values is therefore also a more challenging task (Kutner 1998; Achdou, Indragoby, and Pironneau 2004; Burkovska et al. 2018, 2019).

There are different ways to compute implied information. For European options, a closed-form expression may be derived, for example, to approximate the implied volatility in certain parameter ranges, see Brenner and Subrahmanyam (1988), Corrado and Miller (1996), and Chance (1996). Such expressions are typically based on a Taylor series expansion and on the analytical solution of the European option pricing model. One of the drawbacks, however, is that the resulting formulas are only accurate near at-the-money (ATM), and may give rise to inaccurate implied volatility values for deep in-the-money (ITM) and out-of-the-money (OTM) options. To estimate the implied dividend, a popular way is by means of the put-call parity, which holds, however, only for European options, and is not valid for American options due to the early-exercise premium (Fodor, Stowe, and Stowe 2017; Hull 2019). A second approach is by formulating the computation of the implied information as a minimization problem, which is then based on an iterative search technique. Traditionally, this methodology requires the repeated solution of the American option pricing problem before reaching the stop criterion. Under a minimization framework, there are essentially two popular approaches of determining the American option implied volatility. The first one is by means of a 'de-Americanization technique', which translates American option prices into the corresponding European prices (Jourdain and Martini 2002; Carr and Wu 2009; Burkovska et al. 2018, 2019). Significant pricing errors may arise due to inaccurate incorporation of the early-exercise premium. The higher the early-exercise premium, typically the larger the error can get. Taking dividends into consideration may further increase the error of the de-Americanization technique. A second approach is to conduct a direct calibration of the American pricing model, see Lagnado and Osher (1997), Kutner (1998), Achdou, Indragoby, and Pironneau (2004), Nardon and Pianca (2013), and Burkovska et al. (2019). Unlike the European options, the derivative of the option value with respect to the volatility does not have a closed-form expression

in the case of American options. In addition, other complicating factors, such as a negative interest rate (Hendrik Frankena 2016), may lead to complex-shaped early-exercise regions (e.g., we may encounter different continuation regions) (Donno, Palmowski, and Tumilewicz 2020).

Recently, artificial neural networks (ANNs) have been emerging as advanced computational techniques to obtain solutions to possibly complicated problems in computational finance, for example, in Han, Jentzen, and Weinan (2018), Liu, Oosterlee, and Bohte (2019), Buehler et al. (2019), Liu et al. (2019), Becker, Cheridito, and Jentzen (2019), and Lokeshwar, Bharadwaj, and Jain (2022). We refer to (Ruf and Wang 2020) for a review. Nowadays, deep neural networks are also used to speed up the computation of prices and sensitivities of American options, see Sirignano and Spiliopoulos (2018), Becker, Cheridito, and Jentzen (2020), Chen and Wan (2021), and Salvador, Oosterlee, and van der Meer (2021). To date, those neural network-based algorithms have not been computationally efficient to enable fast American option model calibration.

In order to accelerate the time-consuming solution of the American option pricing model, we can take advantage of supervised learning with ANNs. Training ANNs to learn the implied information from observed option prices is however non-trivial. For example, a steep gradient of the implied volatility with respect to the option price may cause inaccurate ANN results (Liu, Oosterlee, and Bohte 2019). Moreover, in the early-exercise region, the American option price and the volatility are not bijective, as the gradient of the option price with respect to volatility, Vega, equals zero in the early-exercise region. A robust, global optimization plays an important role when exploring the solution space. Consequently, the ANN efficiency may suffer from the global optimization. The paper (Liu et al. 2019) proposed a neural network-based calibration framework, i.e., Calibration Neural Networks (CaNN), to address the speed issue of model calibration with a global optimization algorithm. A recent study (Büchel et al. 2022) shows that the CaNN can achieve a competitive performance when dealing with financial market data. Usually, however, the implied dividend yield is also unknown, so that there are two open implied model parameters to determine, see for example, Cao (2005) and Kragt (2016).

In this work, we will employ the CaNN to extract the implied information from American option prices. Negative interest rates as well as a negative dividend yield are taken into consideration to cover a broad range of market conditions. We thus propose a data-driven machine learning method to address the American option implied information. More specifically, the proposed CaNN is composed of three components, an efficient option pricing method, a global optimization technique and an implementation which runs efficiently on a parallel computing platform. There are two separate stages in CaNN, the forward pass to learn the pricing model and the backward pass to estimate the model parameters. Here the forward pass of CaNN will give us two output quantities, the American put and call prices, which originate from one ANN.

The remainder of this paper is organized as follows. In Section 2, the mathematical American option pricing models are introduced. Furthermore, in Section 2.3, the implied volatility and implied dividend yield from American options are discussed. In Section 3, we describe the data-driven ANN to extract implied information from American options. When a dividend yield is already known, the CaNN simplifies as it can be used to compute the American option implied volatility by directly approximating the inverse function. In such case, an iterative numerical method is not needed. In other cases, the CaNN is

employed to determine both the implied dividend and implied volatility. In Section 4, numerical experiments are presented to demonstrate the performance of the proposed methods.

2. American Options

In this section, we will discuss the mathematical model used to price the American options, and the implied information in the market option prices.

2.1. Problem Formulation

Although other pricing models would easily fit in our framework, for clarity we will concentrate on the Black-Scholes pricing framework. The underlying asset price thus follows a Geometric Brownian Motion (GBM) process, under the risk-neutral measure,

$$\mathrm{d}S(t) = (r-q)S(t)\,\mathrm{d}t + \sigma S(t)\,\mathrm{d}W^{\mathbb{Q}}(t), \quad S(0) = S_0, \tag{1}$$

where $S(t)$ is the underlying spot price at time t, and σ is the volatility parameter and $W(t)$ a Wiener process under the risk-neutral measure $\mathbb{Q}$, S_0 the starting point at time $t = 0$. The two parameters r and q can be interpreted in different ways. For example, r and q are the risk-less interest rate and dividend yield, respectively, for stocks. In the context of currencies, r and q may be two different interest rates, and q would represent the cost of carry in the case of commodities. In this paper, we stay with a stock option description, for convenience, but we will also discuss $q < 0$, which is found in commodity modelling. With a risk-less asset $B(t)$,

$$\mathrm{d}B(t) = rB(t)\,\mathrm{d}t, \tag{2}$$

the arbitrage-free value of an American option at time t is given by

$$V_{am}(t,S) = \sup_{u\in[0,T]} \mathbb{E}_t^{\mathbb{Q}}[e^{-r(T-t)}H(K,S(u)) \mid S(u)], \tag{3}$$

where $H(\cdot)$ is the payoff function, with strike price K; $\mathbb{E}_t^{\mathbb{Q}}$ represents the expectation under the risk-neutral measure $\mathbb{Q}$, with T being the maturity time. An optimal exercise boundary $S_t^* \equiv S^*(t)$, which depends on the time to maturity $T-t$, divides the domain into early-exercise (stopping) regions Ω_s and continuation (or holding) regions Ω_h. In general, early-exercise will be triggered when the discounted expected value drops below the value of exercising the option.

As an American option can be exercised anytime before the expiry time, a corresponding early-exercise premium should be added to the European option counterpart. For example, an American put option (Carr, Jarrow, and Myneni 1992) can be decomposed into the corresponding European put price and the early-exercise premium, i.e.,

$$V_{am}^P(t,S) = \mathbb{E}_t^{\mathbb{Q}}[e^{-r(T-t)}\max(K-S(T),0)] + \int_t^T \mathbb{E}_u^{\mathbb{Q}}[(rK-qS(u))\mathbb{1}_{\{S(u)\in\Omega_s\}}]\,\mathrm{d}u, \tag{4}$$

where Ω_s represents the stopping region, and the whole domain is $\Omega = \Omega_s + \Omega_h$ with Ω_h being the holding region. The first term in Equation (4) is indeed equivalent to the European Black-Scholes put solution. The above problem can be formulated as a Black-Scholes

inequality, on a domain Ω with a free boundary S_t^*. At the free boundary, we have,

$$V_{am}(t, S^*) = H(K, S_t^*), \quad \frac{\partial V_{am}}{\partial S} = \alpha, \tag{5}$$

where $\alpha = 1$ for American calls, $\alpha = -1$ for American puts, and S_t^* is the asset price for which the option value equals the payoff function. For American put options, the payoff function equals $H(K, S(t)) = \max(K - S(t), 0)$. The above conditions at the free boundary can be used to distinguish continuation from stopping regions, and we will use these conditions in Section 3.2. In this paper, the American Black-Scholes solution and corresponding pricing model are denoted by $V_{am} = BS_{am}(\sigma, S, K, t, T, r, q, \alpha)$, with the time to maturity $\tau := T - t$.

2.2. The Put-call Symmetry

The *put-call symmetry* relation holds for both European options and American options, and is given by,

$$V^P(t, S; K, T, \sigma, r, q) = V^C(t, K; S, T, \sigma, q, r). \tag{6}$$

The above relation allows us to value a call or put option by means of its counterpart, where the role of stock and cash values is interchanged. By swapping the strike with the spot price and the interest rate with the dividend yield, an American call value equals the corresponding American put. The relationship is also valid under negative discount rates (Donno, Palmowski, and Tumilewicz 2020). Because with Equation (6) we can get two option prices from one computation, only one function evaluation is required to compute American call and put prices. We will focus on American put options in the following sections, and compute American call options using the put-call symmetry relation.

There are two types of computational regions when dealing with American options, the continuation (or holding) and the early-exercise (or stopping) region. The continuation region boundaries are not known a-priori. Figure 1 illustrates the difference between European and American Black-Scholes option solutions, with two sets of parameters. The American put option price, in the case of no dividend payment, should not be less than the put payoff, while the value of a European put option may be less than the payoff before expiry. In Figure 1, there is only one early-exercise point.

However, two continuation regions may arise, when both the interest rate and dividend yield become negative (Battauz, Donno, and Sbuelz 2015; Donno, Palmowski, and Tumilewicz 2020) in the case of American options. Figure 2 presents an American put solution with two early-exercise points, so that the continuation regions are discontinuous. There are several reasons why we consider a negative dividend yield in our pricing model. The interest rate may be negative in practice. When we use the put-call symmetry to compute American calls by means of puts (the counterpart), we switch the interest rate and the dividend yield in the pricing equation. For that reason, the dividend yield may also be negative in the corresponding formula. An American option Black-Scholes model is also used in foreign exchange or commodity markets, where negative q may be interpreted from an economic point-of-view.

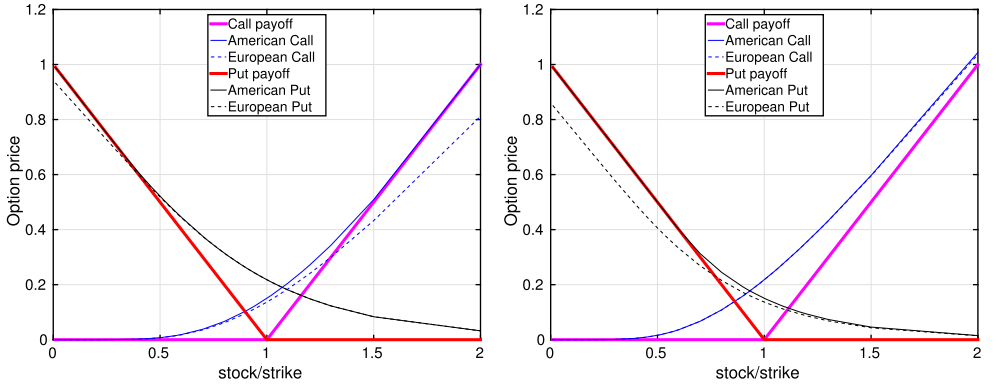


Figure 1. Left: American vs. European Black-Scholes put and call option prices ($r > q$): $r = 0.10$, $q = 0.04$. Right: American vs. European Black-Scholes put and call option prices ($r < q$): $r = 0.04$, $q = 0.10$. The option values are at initial time $t = 0.0$, with the expiry time $T = 1.5$, volatility $\sigma = 0.4$. (a) $r > q$ (b) $r < q$.

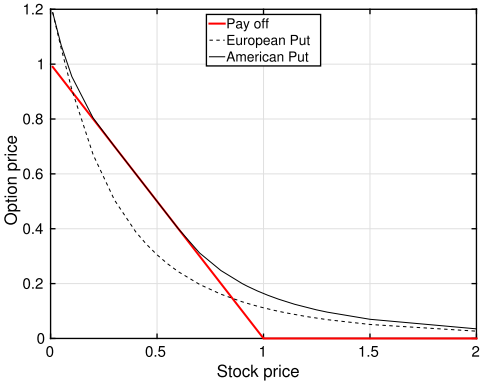


Figure 2. The setting: $r = -0.01$, $q = -0.06$, $\sigma = 0.2$, $T = 20$, $K = 1.0$. The option value in the solid black line hits the payoff function twice. The stopping region is between the two early-exercise points.

2.3. Implied Volatility and Dividend Yield

The information implied in option prices provides participants' expectations about future market conditions. We will discuss the implied volatility and implied dividend yield.

2.3.1. Implied Volatility

Implied volatility represents a specific measure of the future uncertainty from the market point of view. Mathematically, the implied volatility of an option is the level of volatility which, when inserted in the Black-Scholes pricing model, makes the market and model prices match. In that sense, the implied volatility computed from market option prices is viewed as an indication for the look-forward uncertainty of the underlying asset prices as estimated by market participants.

Computing the implied volatility can be formulated as an inverse problem. The American option's implied volatility is then written as,

$$\sigma^* = BS_{am}^{-1}(V_{am}^{mkt}; S, K, t, T, r, q, \alpha), \tag{7}$$

where $BS_{am}^{-1}(\cdot)$ denotes the inversion of the American option Black-Scholes pricing problem, and V_{am}^{mkt} is an American option price observed in the market, with S the underlying asset spot price at time t.

One often solves the implied volatility problem by means of a nonlinear root-finding method, and employs an iterative algorithm to obtain its solution. Given an American option market price, the implied volatility σ^* is determined by solving

$$V_{am}^{mkt} - BS_{am}(\sigma^*; S, K, t, T, r, q, \alpha) = 0. \tag{8}$$

Existence of σ^* can be guaranteed by the monotonicity of the Black-Scholes equation with respect to the volatility in the holding region. Unlike for European options, a closed-from expression for the derivative of the American option value with respect to the volatility is not available. Various solutions have been proposed to solve the implied volatility of American options, see, for example, Lagnado and Osher (1997), Kutner (1998), Achdou, Indragoby, and Pironneau (2004), and Burkovska et al. (2019). As stated in Kutner (1998), these solutions may have difficulties especially with deep in-the-money options. One of the reasons is that option prices are insensitive to the underlying volatility deep in the money. Gradient-free methods, like bisection, do not rely on gradient information, but they may converge slowly because of the stopping regions. An important aspect when extracting the implied volatility is that the derivative of the option price with respect to the volatility, the option's Vega, becomes zero in the stopping region for American call and put options. It is well-known that,

$$|\Delta| = \left|\frac{\partial V_{am}}{\partial S}\right| = 1, \quad \text{Vega} = \frac{\partial V_{am}}{\partial \sigma} = 0. \tag{9}$$

In other words, the American option prices do not depend on the volatility in the stopping regions. As shown in Figure 3, Vega is positive in the holding region and zero in the stopping region. Consequently,

$$\frac{\partial \sigma}{\partial V_{am}} = \frac{1}{\text{Vega}} \to \infty.$$

When we invert the American option Black-Scholes pricing problem in the stopping regions, there is no unique solution for the implied volatility. Therefore, the definition domain of Formula (7) should be the continuation region.

Remark: When gradient-based minimization algorithms (e.g., Newton's method) are used to extract implied information, an initial guess for the solution has to be specified. An inappropriate starting point may cause the algorithm to fall into a 'different' continuation region from the 'envisioned' region. Special rules have to be designed to help the algorithm reach the 'correct' continuation region and explore the solution space. This makes it challenging to define a suitable starting point for the minimization algorithm when inverting the pricing model.

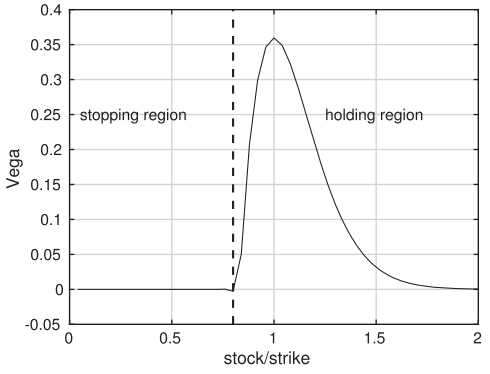


Figure 3. The Vega for American puts in different regions. The strike price is $K = 1.0$.

In our approach, we will use the ANNs to approximate the inverse function of Formula (7) on the continuation region, without relying on any iterative technique, which will be explained in Section 3.2.

2.3.2. Implied Dividend

Many companies pay a dividend to the share holder on the ex-dividend date, which causes the stock price to drop. The option prices are also impacted by the changes in the underlying stock price. Generally, option prices may rise (in the case of the put) or drop (the call) slightly due to the dividend payment. This dividend is called the actual dividend (denoted by δ), while *implied dividend* reflects how the market anticipates future dividend payments. It is extracted from option prices, and thus a quantity under the risk-neutral measure.

The difference between actual dividends and implied dividends is similar to that between historical and implied volatility. The two parameters reflect different market aspects. An implied dividend may be modelled by means of multiple components, for example,

$$q = \epsilon_r + \delta + b, \tag{10}$$

where ϵ_r reflects the difference between the employed and the market interest rate, δ the historical dividend and b the borrowing costs of the underlying asset. Some companies do not pay dividends, but the corresponding options still indicate a non-zero dividend, which may reflect the borrowing level of the stock, see Hull (2019). The borrowing costs are seen as a factor that influences the implied dividend as a function of the time or the strike price.

Our approach is to estimate the implied dividend and implied volatility at the same time, assuming the implied dividend is not constant over strike prices (Nardon and Pianca 2013). In the case of European stock options, the implied dividend can be estimated by the put-call *parity relation* (Hull 2019),

$$V^C_{eu}(t,S) - V^P_{eu}(t,S) = S(t)e^{-q\tau} - Ke^{-r\tau}, \tag{11}$$

so that,

$$q = -\frac{1}{\tau}\log\left(\frac{V^C_{eu} - V^P_{eu} + Ke^{-r\tau}}{S(t)}\right). \tag{12}$$

For American options, the put-call parity does not hold. In certain works (Fodor, Stowe, and Stowe 2017; Hull 2019) the authors employ Formula (12) to roughly estimate the implied dividend yield for American options. Obviously, this may result in inaccuracies when the put-call parity deviation gets large.

In order to eliminate this error, the authors in Guerrero (2017) take the early-exercise premium (EEP) into account for American options. For example, $V^C_{am} = V^C_{eu} + \Delta V^C$ and $V^P_{am} = V^P_{eu} + \Delta V^P$, where ΔV^C and ΔV^P stand for the American call and put early-exercise premiums, respectively. Given the early-exercise premiums ΔV^C and ΔV^P, Equation (11) can be used to calculate the implied dividend yield from the American option prices. A requirement is that the American and European options are available, under the same parameter set. It is however usually not easily possible to deduce the early-exercise premiums from market option prices.

Next, we will numerically investigate how early-exercise premiums affect the put-call parity. As mentioned, the American option price can be viewed as the sum of two components, the corresponding European option price and the early-exercise premium. Let $f(S(u); S(t))$ be the transition density function of $S(u)$ conditional on $S(t)$ for $u \geq t$. Then, Equation (4) can be rewritten as follows (Kuen Kwok 2008),

$$\begin{aligned} V^P_{am}(t,S) &= e^{-r(T-t)} \int_0^K (K - S(T)) f(S(T); S(t))\, \mathrm{d}S(T) \\ &\quad + \int_t^T e^{-r(u-t)} \int_{\Omega_s} (rK - qS(u)) f(S(u); S(t))\, \mathrm{d}S(u)\, \mathrm{d}u \\ &= V^P_{eu}(t,S) + \Delta V^P, \end{aligned} \tag{13}$$

with constant r and q, and where Ω_s is the stopping region. If the holder of an American put chooses to exercise the put option in the case of $S(u)$ being in the stopping region, he/she would gain interest $rKdt$ from the cash received, and lose dividend $qS(t)dt$ from selling the asset. Similarly, the American call price is made up of two components,

$$\begin{aligned} V^C_{am}(t,S) &= e^{-r(T-t)} \int_K^\infty (S(T) - K) f(S(T); S(t))\, \mathrm{d}S(T) \\ &\quad + \int_t^T e^{-r(u-t)} \int_{\Omega_s} (qS(u) - rK) f(S(u); S(t))\, \mathrm{d}S(u)\, \mathrm{d}u \\ &= V^C_{eu}(t,S) + \Delta V^C. \end{aligned} \tag{14}$$

Equations (13) and (14) can be substituted into the European put-call parity,

$$(V^C_{am} - \Delta V^C) - (V^P_{am} - \Delta V^P) = S(t)e^{-q\tau} - Ke^{-r\tau},$$

and the deviation from the put-call parity is found to be,

$$\Delta V^C - \Delta V^P = V^C_{am} - V^P_{am} - S(t)e^{-q\tau} + Ke^{-r\tau}. \tag{15}$$

We can measure the 'deviation' from the European put-call parity relation by EED $:=$ $\Delta V^C - \Delta V^P$, i.e., the difference between two EEPs. The larger the deviation, the more the American and European implied dividend yields will differ. For European options, EED $= 0$.

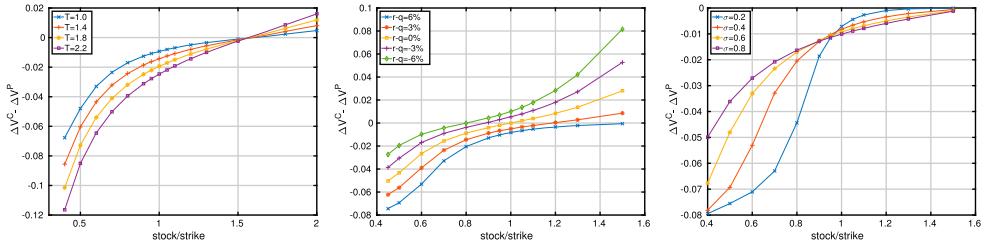


Figure 4. Left: EED versus the ratio of stock over strike price for different maturities T, with $\sigma = 0.4$, $r = 0.1$, $q = 0.05$, $K = 1.0$. Middle: EED versus the ratio of stock over strike price for different values $r-q$, with $T = 1.0$, $r = 0.1$, $\sigma = 0.4$, $K = 1.0$. Right: EED versus the ratio of stock over strike price for different volatilities σ, with $T = 1.0$, $r = 0.1$, $q = 0.04$, $K = 1.0$.

We can assess the corresponding early-exercise premium by calculating the difference between the European and American option prices. In the following figures, we will demonstrate how the early-exercise premiums vary with respect to the following factors, maturity time T (Figure 4(a)), difference between interest rate and dividend yield $r-q$ (Figure 4(b)), volatility σ (Figure 4(c)). Roughly speaking, the absolute deviation is monotonically increasing when an option goes deeper into the money (OTM or ITM), based on Figure 4. Therefore, significant errors may occur when using the European put-call parity to compute the implied dividends from American options.

Remark: The early-exercise premium (EEP) is not observable for most underlying securities, unless both American and European options with the same strike price and time to maturity are available. Empirical studies have been conducted to analyse how the EEP varies in the market using regression techniques, like in Su Chen (2018). The basic idea is to fit a function for the EEP and other observed factors, i.e.,

$$\text{EEP} = \beta_0 + \beta_1(r-q) + \beta_2(T-t) + \beta_3(S/K) + \beta_4(\sigma_{t-1}) + \epsilon$$

In the Swedish equity market, for example, the early-exercise premiums for American puts are empirically found to be positive (Engström and Nordén 2000), increasing with option moneyness, and decreasing with time to maturity and the underlying asset's volatility. For American-style currency options, it was observed (Poitras, Veld, and Zabolotnyuk 2009) that the early-exercise premiums equal approximately 5% for puts and 4.6% for calls. Our numerical results coincide with these empirical studies of EEP.

We will employ a model-based approach, i.e., the American option Black-Scholes pricing model, including a dividend yield, is inverted in order to extract the implied dividend. American options can be priced as follows,

$$\begin{cases} V^C_{am} = BS_{am}(\sigma^*, q^*; S_0, K, t=0, T, r, \alpha = 1), \\ V^P_{am} = BS_{am}(\sigma^*, q^*; S_0, K, t=0, T, r, \alpha = -1), \end{cases} \tag{16}$$

where, BS_{am} is the corresponding pricing model. Assuming the implied volatility and the implied dividend are the same for calls and puts with the same parameter set K, S_0, T and r,

there appears to be a unique solution of the system with two equations and two unknowns. However, this is not always the case, as option prices do not depend on the volatility in the stopping regions. The system of Equation (16) is usually formulated as a minimization problem and a numerical optimization algorithm is employed to search the solution space. Note that a local search-based optimization method will most likely not converge when traversing those early-exercise regions. For that reason, a global searcher is preferred.

3. Methodology

Artificial neural networks, ANNs, have been used to approximate the solution of European option pricing models, for instance, under the Black-Scholes and Heston models in Liu, Oosterlee, and Bohte (2019). Here, we will use the ANN to address the numerical solution of American options, and in particular use it for computing the implied information. For the inverse problem, when there is one parameter to calibrate, we can employ the ANN to build a mapping from the observed market data to the target parameter (as a unique mapping). When there are multiple parameters to calibrate, like the implied volatility and the implied dividend, the CaNN (Calibration Neural Network) (Liu et al. 2019) is more flexible to handle the minimization problem. The former case can be viewed as a simplified CaNN which approximates a single model parameter.

3.1. Artificial Neural Networks

It is well-known that neural networks are powerful function approximators. In a basic formulation, an ANN can be described as a composite function,

$$\mathrm{F}(\mathbf{x} \mid \boldsymbol{\theta}) = \hat{f}^{(\ell)}(\cdots \hat{f}^{(\Theta)}(\hat{f}^{(\Delta)}(\mathbf{x}; \boldsymbol{\theta}^{(\Delta)}); \boldsymbol{\theta}^{(\Theta)}); \cdots \boldsymbol{\theta}^{(\ell)}), \tag{17}$$

where $\mathbf{x}$ stands for the input variables, $\boldsymbol{\theta}$ for the hidden parameters (i.e., the weights and the biases in artificial neurons), ℓ for the total number of hidden layers, and $\hat{f}^{(\ell)}(\cdot)$ represents a hidden-layer function. The composite function, $\mathrm{F}(\cdot)$, depends on these hidden parameters and activation/transfer functions. Once the structure, i.e., the hidden parameters and transfer functions, is determined, the ANN in Equation (17) can be seen as a deterministic function. A popular approach for training neural networks is to employ first-order optimization algorithms to determine the values of the hidden parameters which will minimize the loss function. Gradient-based algorithms are often fast, but it may be difficult to calculate the gradients for a large test set. Stochastic gradient descent algorithms (SGD) randomly select a portion of the data set, to compute the gradient, and address the memory issue. SGD and its variants (like Adam) are thus preferable to train the ANNs on big data sets. In a data-driven supervised learning context, the objective function is given as follows,

$$\arg\min_{\boldsymbol{\theta}} L(\boldsymbol{\theta} \mid (\mathbf{X}, \mathbf{Y})), \tag{18}$$

given the input-output pairs $(\mathbf{X}, \mathbf{Y})$ and a user-defined loss function $L(\boldsymbol{\theta})$. Thus, the ANN will be trained to approximate the function of interest in a certain norm, e.g., the l^2-norm. More details can be found in the book (Goodfellow, Bengio, and Courville 2016). Next, we will discuss how the ANNs are used to approximate the inverse or pricing functions.

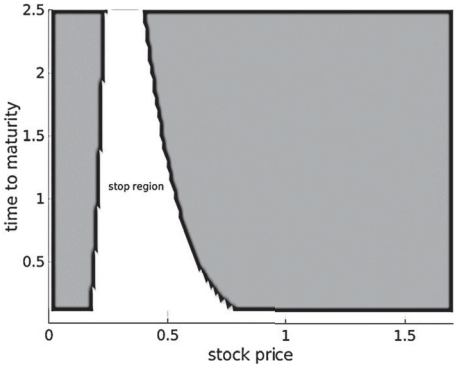


Figure 5. Schematic diagram: An example of two continuation regions for an American put. The shaded area represents the holding region, while the white area represents the stopping region. There are two isolated continuation regions. Here the strike price is fixed $K = 1$.

3.2. ANN for Implied Volatility

When we focus solely on extracting the implied volatility, the basic technique is to employ the ANN to approximate the inverse function of the American Black-Scholes model on a suitable effective definition domain Ω_h,

$$\begin{aligned}\sigma^* &= BS_{am}^{-1}(V_{am}^{mkt}; S, K, t, T, r, q, \alpha) \\ &\approx \mathrm{NN}(V_{am}^{mkt}; S, K, t, T, r, q, \alpha), \quad [V, S, K, t, T, r, q] \in \Omega_h. \end{aligned} \tag{19}$$

The ANN is trained based on the above known model variables, which are observable in the market, to approximate the unique target variable σ^*.

3.2.1. Definition Domain Selection

The continuation regions are not known initially or are complicated so that there is no analytic formula to describe them. However, the early-exercise regions can be found implicitly in a data-driven approach. In our method, we wish to only train the neural network on the points in the continuation region. Overall, our aim is to find the inverse function of the American-style pricing model on the continuation region, represented by the shaded domain in Figure 5. There is an off-line stage, where the continuation regions are determined, and an on-line stage where we use the problem parameters to compute the implied volatility. We build the mapping function via the ANN in the 'irregular' continuation region. The resulting trained ANN solver is typically much faster to determine the implied volatility than an iterative numerical solver.

First, random parameter values are generated as ANN samples in the entire input domain Ω, followed by detecting the parameter samples that are in the early-exercise region Ω_s according to Equation (5).

Second, we use a robust version of the COS method (Fang and Oosterlee 2009) to calculate American option values while generating the data set during the off-line ANN stage. The COS method is an efficient numerical technique and provides the derivative information (i.e., the Greeks) essentially without any additional costs. The basic idea in Fang and Oosterlee (2009) was to approximate the American option value by solving a small series

of Bermudan options with different numbers of exercises opportunities, and subsequently apply a four-point Richardson extrapolation based on these four option prices. With a large variety of asset and option parameters, we need to make sure that the option pricing technique is robust, i.e., it should work under all occurring parameter sets during training. This might imply a large integration domain within the COS method, and a relatively large number of Fourier terms in the cosine expansion. We generalized the above COS method to deal with two separate early-exercise regions, which are encountered when pricing American options with negative interest rates. More details can be found in the Appendix.

Third, we obtain the approximate continuation region, $\Omega_h = \Omega - \Omega_s$, as shown in Figure 5. There are two indicators to detect the samples in the early-exercise region, the difference between the option value and the payoff, and the option's sensitivity Vega in Equation (9). To control the numerical errors, threshold values are prescribed for the two indicators. A threshold ϵ_1 is set for the difference between the payoff function value and the generated option value,

$$V_{am}(t, S) - H(K, S(t)) > \epsilon_1. \tag{20}$$

The appropriate training samples for the continuation region are selected based on Formula (20). We also set a threshold ϵ_2 for the value of Vega,

$$\text{Vega} > \epsilon_2, \tag{21}$$

with $\epsilon_1 \in \mathcal{R}^+$ and $\epsilon_2 \in \mathcal{R}^+$. As early-exercise takes place with options that are ITM, the above two criteria only apply to ITM samples. In principle, Criterion (20) is equivalent to Criterion (21), but for robustness reasons, both will be enforced to mitigate the influence of numerical errors. However, Vega may be unavailable in the prediction phase. In that case, only Criterion (20) will be implemented. Note that the shape of the stopping region also depends on the other model parameters. For example, the number of the early-exercise points associated with the stopping region decreases from two to one, when the dividend yield increases from a negative value to a positive one. Here for the illustration purpose, in Figure 5, we plot a stop region just as a function of stock price and time to maturity.

Let $\hat{\Omega}_s$ represent the region where the generated samples do not meet the requirements (20) and (21), and the remaining region $\hat{\Omega}_h = \Omega - \hat{\Omega}_s$. The effective definition domain is found numerically by taking $\Omega_h \approx \hat{\Omega}_h$. The procedure requires additional computations, but these happen in the off-line stage without affecting the on-line approximation.

3.2.2. Gradient-squashing and Option Prices

Generally, ANNs are not accurate when functions with steep gradients need to be approximated. Therefore, we adapt the requested output function. To obtain the implied volatility from the option prices, we need to employ the so-called gradient-squashing technique as proposed in Liu, Oosterlee, and Bohte (2019). We subtract the intrinsic value from the American option price to obtain the corresponding *time value*. Next, a brief derivation of computing the intrinsic value of an American option with the dividend yield is given.

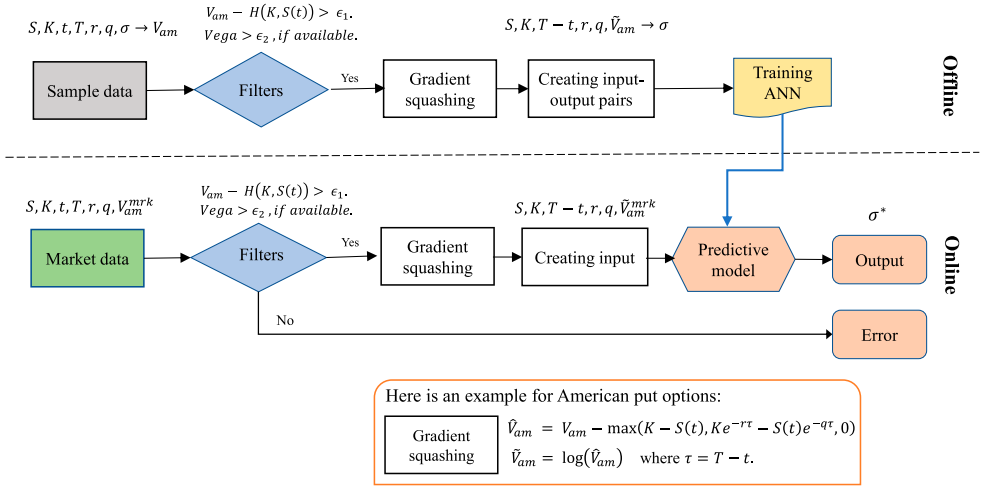


Figure 6. A flowchart of the ANN-based method to compute the implied volatility from American option.

Taking the put as an example, recall the put-call parity for European options,

$$V_{eu}^{C} - S(t)e^{-q\tau} = V_{eu}^{P} - Ke^{-r\tau}.$$

The following lower bound can then be deduced,

$$V_{eu}^{P}(t,S) = V_{eu}^{C}(t,S) + Ke^{-r\tau} - S(t)e^{-q\tau} \geq Ke^{-r\tau} - S(t)e^{-q\tau}, \tag{22}$$

where the right-hand side is called the European option's intrinsic value. As an American option is at least as expensive as its European counterpart, we have

$$V_{am}^{P}(t,S) \geq V_{eu}^{P}(t,S) \geq Ke^{-r\tau} - S(t)e^{-q\tau}. \tag{23}$$

Additionally, American option prices should not be worth less than the pay-off function at any time, as for example shown in Figure 2, and the time value of an American put is computed by

$$\hat{V}_{am}^{P} = V_{am}^{P}(t,S) - \max(K - S(t), Ke^{-r\tau} - S(t)e^{-q\tau}, 0). \tag{24}$$

After this step, the gradient-squashing technique (Liu, Oosterlee, and Bohte 2019) is applied as follows,

$$\tilde{V}_{am}^{P} = \log(\hat{V}_{am}^{P}). \tag{25}$$

The gradient-squashing technique for computing the implied volatility using the ANNs, means taking the logarithm of the time value to obtain a quantity that can be well approximated with ANNs, because its gradient is not too steep to approximate accurately.

As shown in Figure 6, with help of the sample filters and the gradient squashing technique, the American implied volatility can be easily computed using neural networks.

3.3. Determining Implied Dividend and Implied Volatility

When the American option implied dividend yield is unknown, we will determine both implied volatility and implied dividend simultaneously by means of the CaNN calibration methodology. We assume the implied volatility and the implied dividend are identical for American calls and puts with the same K, S_0, T and r values,

$$\begin{cases} V_{am}^{C,mkt} - BS_{am}(\sigma^*, q^*; S_0, K, t = 0, T, r, \alpha = 1) = 0, \\ V_{am}^{P,mkt} - BS_{am}(\sigma^*, q^*; S_0, K, t = 0, T, r, \alpha = -1) = 0, \end{cases} \tag{26}$$

so that there are two unknown parameters to calibrate, implied volatility σ^* and the implied dividend yield q^*, given a pair of American option prices, $V_{am}^{C,mkt}$ and $V_{am}^{P,mkt}$. The above system is reformulated as a minimization problem,

$$\underset{\sigma^* \in R^+, q^* \in R}{\text{argmin}} \ (BS_{am}(\sigma^*, q^*; \alpha = 1) - V_{am}^{C,mkt})^2 + (BS_{am}(\sigma^*, q^*; \alpha = -1) - V_{am}^{P,mkt})^2. \tag{27}$$

We employ a fast, generic and robust calibration framework, the CaNN (Calibration Neural Networks) developed in Liu et al. (2019). The basic idea of the methodology is to convert the calibration of the model parameters into an estimation of the neural network's hidden units. The reason for this is that both model calibration and training ANNs (here supervised learning) can be reduced to solving an optimization problem according to Equations (18) and (27). It can take advantage of parallel GPU computing to speed up the computations, which enables us to employ a global optimization technique to search the solution space. The gradient-free optimization algorithm, here Differential Evolution (DE), typically does not get stuck in local minima or in the stopping region. Another benefit of DE is that it is an inherently parallel technique where populations of possible optimal solutions can be evaluated simultaneously within the ANN.

There is a calibration (the backward pass) in the CaNN, in addition to the training and testing stages. The training is to determine suitable weights and biases in the hidden layers to map model input to output, while the calibration estimates the model input parameters to optimally match a given output. We will view the different stages as one framework, and just change the learnable units between the original layers (i.e., the hidden, output and input layers) within the different stages. More specifically, the CaNN consists of two passes. The forward pass, including the training and testing stages, approximates the American Black-Scholes prices. We have developed one neural network providing two output values in the forward pass, the American call and the put prices, as illustrated in Figure 7. The backward pass, on the other hand, aims to find the two parameters, (σ^*, q^*), to match the two observed American option prices, $V_{am}^{P,mkt}$ and $V_{am}^{C,mkt}$, with strike price K, maturity time T, spot price S_0, interest rate r.

In Figure 7, black circles represent the frozen (unchanged) parameters during each epoch, and open circles represent the parameters to be optimized. For example, in the training phase, the black circles represent input-output pairs from the training dataset, while the open circles represent the hidden weights and biases. In the calibration phase, the black circles correspond to the already trained hidden weights and biases, as well as the quotes (strike price, stock price, option price, maturity time, etc) observed in the market, while the open circles stand for the implied volatility and implied dividend yield, which are the to-be-calibrated parameters in Formula (28).

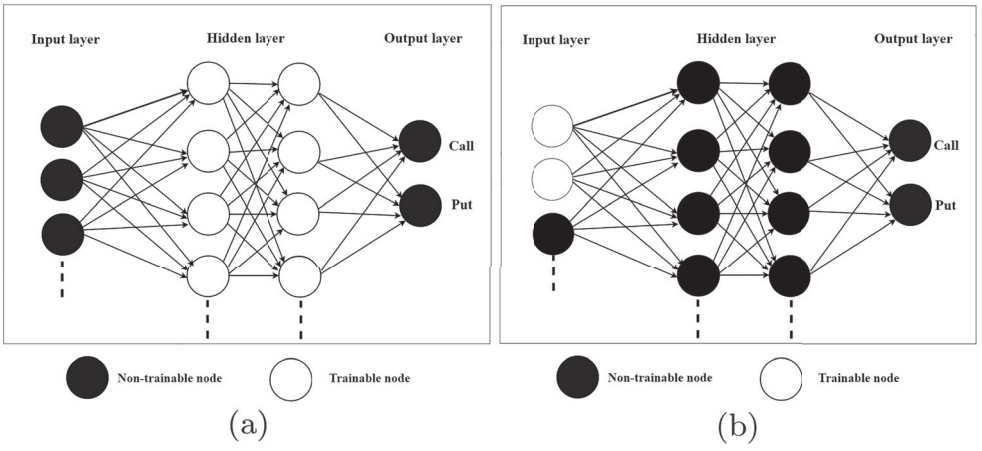


Figure 7. Left: In the forward pass of the CaNN, the output layer gives us two option prices. Right: During the calibration, the CaNN estimates the two parameters, implied volatility and implied dividend, in the original input layer. (a) Training (b) Calibration.

This is different from the network in Liu et al. (2019), where one neural network corresponds to one output quantity. Therefore Equation (27) is written as an objective function for the model calibration, as follows,

$$\underset{\sigma^*,q^*}{\operatorname{argmin}}(\mathrm{NN}(\sigma^*, q^*; \alpha = 1) - V_{am}^{C,mkt})^2 + (\mathrm{NN}(\sigma^*, q^*; \alpha = -1) - V_{am}^{P,mkt})^2, \qquad (28)$$

where $\sigma^* \in R^+, q^* \in R$. Formula (28) is used as the loss function for the backward pass in the CaNN. The (group-based) DE global optimization can be implemented in an efficient way together with ANNs, see Figure 8, on a central processing unit (CPU) or a graphics processing unit (GPU). Figure 8 illustrates the parallel computing aspect of the calibration in the context of Differential Evolution and Neural Networks. In such way, the for-loop is replaced by an efficient matrix multiplication, and searching the implied information globally can be completed efficiently.

Remark: Because of a generic calibration framework, the CaNN can easily deal with more complex situations, for example, the objective function (28) including more than a pair of American price quotes which share the same implied dividend, for example, as in the work (Cao 2005).

3.4. The ANN Configuration

Our chosen ANN architecture in the forward pass constitutes four hidden layers and two parallel output layers. Some particularly useful operations in deep neural networks, e.g., dropout and batch normalization, did not bring any significant benefits to our ANN. The proposed configuration has been demonstrated to be able to fit the pricing model with acceptable accuracy in Liu et al. (2019). We choose the activation function Softplus in this

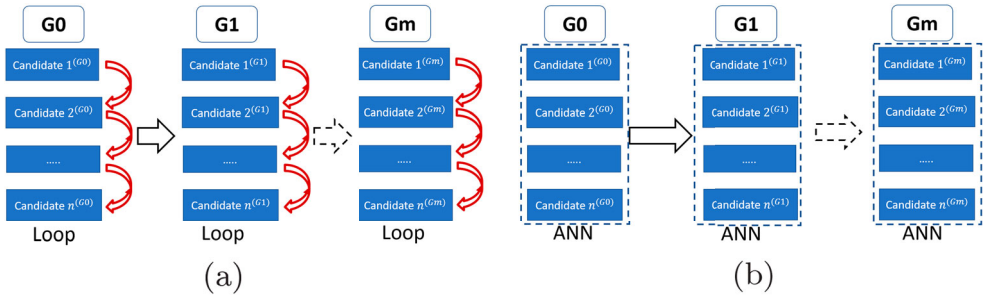


Figure 8. The global optimizer DE runs in parallel within CaNN. 'Gm' represents the *m*-th generation, where there are *n* candidates of to-be-optimized parameters, i.e., *n* sets of open parameters in the pricing model. These *n* sets of model parameters are independent, thus can be processed by the ANN simultaneously, instead of a for-loop, to reduce the computation time. (a) Conventional DE (b) Parallel DE.

Table 1. The ANN configuration.

Hyper-parameters	Options
Hidden layers	4
Neurons (each layer)	200
Activation	Softplus
Initialization	Glorot_uniform
Optimizer	Adam
Batch size	1024

paper, i.e.,

$$\tilde{f}(x) = \log(1 + e^x),$$

as its smooth derivative fits well to the smoothness of the pricing function, especially in the continuation region. According to the universal approximation theorem, a one-layer based ANN can be used to approximate any continuous function to any desired precision, but with the rate which linearly depends on the number of neurons involved. The depth of the ANN (i.e., the number of hidden layers) can increase the function's representation accuracy exponentially, but deep ANNs are difficult to implement in parallel (e.g., a hidden layer has to wait for output signals of a previous one), resulting in long computation time. Considering the approximation power and the computation efficiency, we choose four hidden layers and 200 neurons in each layer. Both ANNs in the forward and backward passes of the CaNN will make use of the hyper-parameters that are shown in Table 1. In the backward pass, the DE algorithm, a gradient-free global optimizer, replaces Adam. These values have shown to result in a robust neural network which converges well for a variety of problem parameters.

4. Numerical Results

This section presents numerical experiments for using the ANN to extract implied information from American options. We begin with the simplified CaNN, focussing on the implied volatility, and later employ the CaNN to extract implied volatility and dividend.

Note that we use the COS method described in the Appendix to compute American option prices for the training data set and for the simulated market data in this section.

4.1. Computing Implied Volatility

In this section, we approximate only implied volatility. We use one forward pass to approximate the inverse function, from an American option price to its corresponding implied volatility, assuming the other parameters to be known.

Without loss of generality, we use a fixed spot price $S_0 = 1.0$. Then, the input for the ANN is made up of five parameters $\{\log(\hat{V}^P_{am}), K, r, q, \tau\}$. The two thresholds in Equations (20) and (21) are $\epsilon_1 = 0.0001$ and $\epsilon_2 = 0.001$ for the data set.

The ANN is trained with American put options to optimize the weights of the ANN-based solver for the computation of the implied volatility from the American options. We define the measures as follows:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \quad \text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \quad \text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i},$$

where y represents the true values of the American option prices, and $\hat{y}$ represents the ANN approximated values, with n being the number of samples. During training, the MSE is used to find the weights and biases, while the other two measures MAE and MAPE are monitored. The goodness of fit, R^2, is also provided, which describes the closeness from the predicted values to the true values. By using different measures, we evaluate the quality of approximation from different angles.

After the model input parameters are sampled (here by LHS) over the specific domain, the COS method is used to solve the corresponding American Black-Scholes pricing model, resulting in the data collection $\{S_0, K, \tau, r, q, \sigma, V_{am}\}$, where σ is considered the ground-truth value. Afterwards, the variable σ is placed into the output layer of the ANN, as the implied volatility $\sigma^* \equiv \sigma$ for the data collection. Meanwhile, the other variables in the collection are included in the input layer of the ANN, and more details are in Table 2. The validation samples help avoid over-fitting during training the ANN. The test samples, which the ANN did not encounter during training, are subsequently used to evaluate the generalization performance of the trained ANN. There are around one million samples, with 80% being used as training, 10% as validation, 10% as test samples. The learning rate is halved every 400 epochs during training. After 4000 epochs, the training and validation losses have converged.

Table 3 and Figure 9 present the performance of the trained ANN. The test performance is close to the train performance, suggesting that the trained ANN generalizes well for unseen data, as shown in Table 3. The ANN predicted implied volatility values approximate the true values accurately for both the train and test datasets, as is indicated by the R^2 measure value in Figure 9. Moreover, the online prediction stage for the American option implied volatility, requiring only the evaluation of the trained ANN, is much faster than traditional iterative root-finding algorithms.

It is observed that the trained model performance tends to decrease when the pricing model parameters gets close to the upper or lower bounds of the values in Table 2. In other

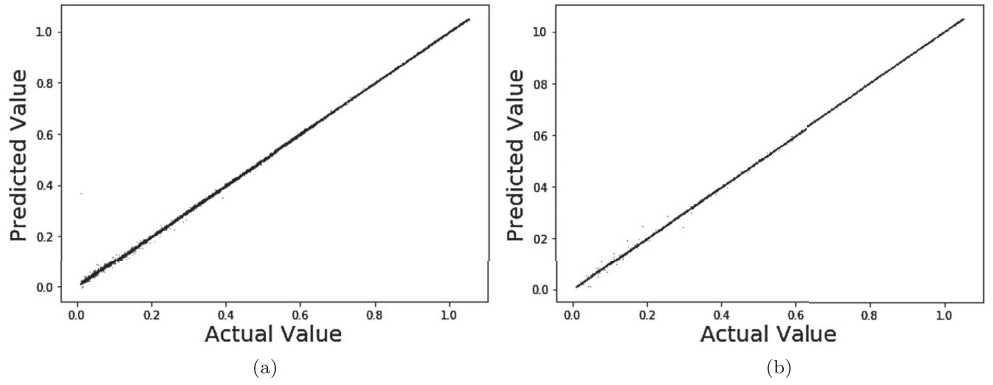
Table 2. Training dataset for American options under the Black-Scholes model; Although the spot price is fixed $S_0 = 1$, one would scale the strike and option prices if needed.

ANN	Parameters	Value range	Employed method
ANN Input	Strike, K	[0.60, 1.40]	LHS
	Time value, log $(\hat{V}^P_{am})$	$(-11.51, -0.24)$	COS
	Time to maturity, τ	[0.05, 3.0]	LHS
	Interest rate, r	$[-0.05, 0.10]$	LHS
	Dividend yield, q	$[-0.05, 0.10]$	LHS
ANN output	Implied volatility, σ^*	(0.01, 1.05)	LHS

Notes: Here we use $\tau := T - t$ instead of the two variables T and $t = 0$. The upper bound of the American put prices is set to 1.8. LHS stands for Latin Hypercube Sampling.

Table 3. Multiple measures are used to evaluate the performance.

–	MSE	MAE	MAPE	R^2
Training	$4.33 \cdot 10^{-7}$	$2.44 \cdot 10^{-4}$	$1.11 \cdot 10^{-3}$	0.999994
Testing	$4.60 \cdot 10^{-7}$	$2.51 \cdot 10^{-4}$	$1.15 \cdot 10^{-3}$	0.999993



(a) (b)

Figure 9. Predicted versus ground-truth implied volatility value on the test data set. Here the actual values are prescribed. Left: $R^2 = 0.999994$; Right: $R^2 = 0.999993$. (a) Training (b) Testing.

words, outliers are most likely to appear near the boundary. Thus the training data set is recommended to have a wider parameter range than the test range of interest.

4.2. Computing Implied Information

Next, we will approximate the implied volatility and the implied dividend yield simultaneously from the observed American option prices using the CaNN. The CaNN is based on a forward and a backward pass, that are implemented in a sequential way.

Here we extend the original CaNN by using one forward pass to approximate two American option prices, a put and a call value. The input for the neural network in the forward pass consists of $(S_0, K, r, q, \sigma, \tau)$, and the output comprises a pair of American prices, that is $(\tilde{V}^P_{am}, \tilde{V}^C_{am})$. As the neural network gives us two output variables, the loss function of the

Table 4. Training data set for the forward pass. We fix $S_0 = 1$, and sample strike prices K to generate different moneyness levels.

ANN	Parameters	Value range	Method
Forward input	Strike, K	[0.45, 1.55]	LHS
	Time to maturity, τ	[0.08, 3.05]	LHS
	Risk-free rate, r	[−0.1, 0.25]	LHS
	Dividend yield, q	[−0.1, 0.25]	LHS
	Implied volatility, σ	(0.01, 1.05)	LHS
Forward output	American put, V^P_{am}	(0, 1.8)	COS
	American call, V^C_{am}	(0, 1.2)	COS

Notes: The total number of the data samples is nearly one million, with 80% training, 10% validation, 10% test samples. Note that the American call option prices are obtained through the put-call symmetry, where the corresponding put prices are first computed by the COS method.

Table 5. The performance of the CaNN forward pass with two outputs.

–	Option	MSE	MAE	MAPE	R^2
Training	call	1.40×10^{-7}	3.00×10^{-4}	1.25×10^{-3}	0.9999965
	put	2.54×10^{-7}	4.24×10^{-4}	1.64×10^{-3}	0.9999959
Testing	call	1.43×10^{-7}	3.02×10^{-4}	1.27×10^{-3}	0.9999964
	put	2.55×10^{-7}	4.26×10^{-4}	1.64×10^{-3}	0.9999959

forward pass includes two components,

$$\text{MSE} = \frac{1}{2n}\sum_{i=1}^{n}\{(\tilde{V}^P_{am,i} - V^{P,mod}_{am,i})^2 + (\tilde{V}^C_{am,i} - V^{C,mod}_{am,i})^2\}, \tag{29}$$

where $V^{P,mod}_{am}$ and $V^{C,mod}_{am}$ stand for the American put and call prices, respectively, that are generated by the American option Black-Scholes model. This rule also applies to MAE and MAPE. There are four hidden layers with 200 neurons each layer, as shown in Table 1. The total number of hidden parameters is 122,202, and the loss function, Equation (29), is used to update the hidden layers of the CaNN during the training stage. The training data set is constructed according to the parameter ranges in Table 4, where the COS method is used to compute American option prices for both the training data set and the simulated market data. Based on a converging and accurate training phase, the performance of the CaNN's forward pass is presented in Table 5. The results, for both calls and puts, are satisfactory, achieving very high levels of precision in all the considered measures.

After performing the forward pass, in the CaNN's backward (calibration) pass the implied parameters are determined, in our current setting. Supposing each option quote in the market includes American call and put prices, the idea behind the backward pass is to determine two parameters (σ^*, q^*) within the American Black-Scholes model to optimally match the pair of market option prices, given the interest rate r, maturity time T, strike price K, and spot price S_0. The objective function for the calibration procedure is found in Formula (28), which is equivalent to Criterion (29) in the case of $n = 1$. In practice, the market price is taken to be the mid-price of the bid and ask prices.

Table 6. Examples of using CaNN to extract implied volatility and implied dividend.

K/S_0	T	r	$\sigma^{\dagger}$	$q^{\dagger}$	C_{am}^{mkt}	P_{am}^{mkt}	σ^{*}	q^{*}
1.0	0.5	−0.04	0.1	0.06	0.0146	0.0597	0.099	0.059
1.1	0.5	−0.04	0.2	−0.06	0.0255	0.1181	0.198	−0.061
1.0	0.75	0.0	0.3	−0.02	0.1119	0.0976	0.300	−0.020
1.2	1.0	−0.04	0.4	0.08	0.0603	0.3810	0.40	0.080
0.8	1.0	0.02	0.3	0.02	0.2322	0.03472	0.299	0.020
0.7	1.25	0.0	0.4	−0.04	0.3886	0.0378	0.399	−0.040

†Represents the prescribed values.*Represents the calibrated values.

Table 7. The parameter range for the systemic experiment.

Parameter	Interval	Step	Number
σ	[0.10, 0.45]	0.05	8
q	[−0.06, 0.08]	0.02	8
K	[0.70, 1.20]	0.10	6
τ	[0.50, 1.50]	0.25	5
r	[−0.04, 0.06]	0.02	6
P_{am}^{mkt}	[0.70, 1.20]	–	9271
C_{am}^{mkt}	[0.70, 1.20]	–	9271

Remark: The objective function, under the CaNN, can also be defined differently taking into account the bid-ask spread, Delta (the sensitivity of an option's value to the underlying asset price) weighting, and other factors. This is however out of our scope here.

In order to evaluate the approach, we prescribe model parameters and investigate how accurately the CaNN can recover them. Table 6 presents a set of examples, including many different scenarios, e.g., ITM and OTM scenario's, are considered. The results in Table 6 suggest that the CaNN can recover the implied volatility and implied dividend highly accurately from our 'artificial market option data'. Even when interest rates and dividend yields are negative, the CaNN recovers the true values without any stalling of the convergence. The method's robustness may be attributed to the robust numerical solver generating accurate option prices for a wide range of model parameters, the designed neural network providing sufficient approximation capacity, and the gradient-free optimizer (i.e., DE) to globally search the solution space.

For the DE, a search interval is required for each parameter, and we provide $q \in [-0.08, 0.1]$ and $\sigma^{*} \in (0, 1.0)$. During the mutation operation in the DE, the population size of each generation is taken as a small number, here 10. We choose 'best1bin' as the mutation strategy, that is, the best candidate of the previous generation enters the mutation. During the crossover stage, the crossover possibility is set to 0.7. During the selection stage, all new trial candidates with the objective function can be processed in parallel, and the DE convergence tolerance is set to 0.01. As the number of calibration parameters is only two, the computation time on a CPU is around 0.37 seconds using the sequential DE and is less than 0.1 second using the parallel version of the DE method.

Furthermore, a systemic test is conducted to assess the averaged performance over a large number of cases. We generate equally-spaced samples over a certain interval according to Table 7, but remove the samples that are connected to early-exercise region option prices. The experiment ends up with 9271 test cases. The results in Table 8 suggest that the proposed approach performs very well under a wide variety of option market conditions

Table 8. With a CPU (Intel i5) and a GPU(NVIDIA Tesla P100), the averaged performance of the CaNN estimating implied volatility and implied dividend based on 9271 different test cases.

Absolute deviation		Computational cost	
$\lvert\sigma^{\dagger} - \sigma^{*}\rvert$	6.65×10^{-3}	**CPU** time (seconds)	0.08
$\lvert q^{\dagger} - q^{*}\rvert$	8.56×10^{-4}	**GPU** time (seconds)	0.04

Note: The averaged number of function evaluations is 1060.

at a reduced computational cost. The calibration speed is due to the efficient forward pass and the parallel, gradient-free DE optimizer. Basically the forward pass serves as a fast numerical solver for the American pricing model. Additionally, with two output prices, the forward pass requires half of the computational costs.

5. Conclusion

We studied the problem of pricing American options and extended a data-driven machine learning method to extract the implied volatility and implied dividend yield from observed market American option prices in a fast and robust way.

For computing the American implied volatility, we explained that the domain for the inverse function should be equivalent to the continuation regions of the American options. The ANN-based approach builds an approximating function and addresses complex boundaries of the definition domain, by means of the different off-line and on-line stages. More specifically, we used two conditions to classify the random data samples in the domain in the off-line stage. The definition domain is represented by data points which lie in the continuation regions. Subsequently, a neural network was trained on those samples to approximate the inverse function. This data-driven approach avoids an iterative algorithm which may suffer from convergence problems. Due to the off-line definition of the domain, our approach also successfully dealt with negative interest rates and dividend yields, where two early-exercise regions may appear. In short, the offline-online decoupling brings much flexibility.

Furthermore, we presented a method for finding simultaneously implied dividend and implied volatility from American options using a calibration approach. The CaNN, which consists of an efficient solver and a fast global optimizer, is employed to carry out the calibration procedure. As a result, the early-exercise premiums, which the European option put-call parity relation fails to deal with, are handled successfully. The parallel global optimizer prevents the CaNN from stopping in the early-exercise regions and allows to achieve a good quality solution in a short amount of time. The numerical experiments demonstrate that the CaNN is able to accurately extract multiple pieces of implied information from American options. A continuous dividend yield is considered in this paper, and it should be feasible to extend the approach to deal with time-dependent or discrete dividends.

Acknowledgments

We would also like to thank Dr.ir Lech Grzelak for valuable suggestions, as well as Dr. Damien Ackerer for fruitful discussions. The author S. Liu would like to thank the China Scholarship Council (CSC) for the financial support.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Shuaiqiang Liu http://orcid.org/0000-0003-2697-1932
Álvaro Leitao http://orcid.org/0000-0002-3442-4587
Cornelis W. Oosterlee http://orcid.org/0000-0002-7322-4094

References

Achdou, Yves, Indragoby, Govindaraj, and Pironneau, Olivier. 2004. "Volatility Calibration with American Options." *Methods and Applications of Analysis* 11 (4): 533–556.
Büchel, Patrick, Kratochwil, Michael, Nagl, Maximilian, and Rösch, Daniel. 2022. "Deep Calibration of Financial Models: Turning Theory Into Practice." *Review of Derivatives Research* 25: 109–136.
Battauz, Anna, Donno, Marzia De, and Sbuelz, Alessandro. 2015. "Real Options and American Derivatives: The Double Continuation Region." *Management Science* 61 (5): 1094–1107.
Becker, Sebastian, Cheridito, Patrick, and Jentzen, Arnulf. 2019. "Deep Optimal Stopping." *Journal of Machine Learning Research* 20 (74): 1–25.
Becker, Sebastian, Cheridito, Patrick, and Jentzen, Arnulf. 2020. "Pricing and Hedging American-Style Options with Deep Learning." *Journal of Risk and Financial Management* 13 (7): 158.
Bilson, John F. O., Baum Kang, Sang, and Luo, Hong. 2015. "The Term Structure of Implied Dividend Yields and Expected Returns." *Economics Letters* 128: 9–13.
Brenner, Menachem, and Subrahmanyam, Marti G. 1988. "A Simple Formula to Compute the Implied Standard Deviation." *Financial Analysts Journal* 44 (5): 80–83.
Buehler, H., Gonon, L., Teichmann, J., and Wood, B. 2019. "Deep Hedging." *Quantitative Finance* 19 (8): 1271–1291.
Burkovska, Olena, Gass, Maximilian, Glau, Kathrin, Mahlstedt, Mirco, Schoutens, Wim, and Wohlmuth, Barbara. 2018. "Calibration to American Options: Numerical Investigation of the De-Americanization Method." *Quantitative Finance* 18 (7): 1091–1113.
Burkovska, Olena, Glau, Kathrin, Mahlstedt, Mirco, and Wohlmuth, Barbara. 2019. "Complexity Reduction for Calibration to American Options." *Journal of Computational Finance* 23: 25–60.
Cao, Qi. 2005. "Computation of Implied Dividend Based on Option Market Data." MSc Thesis, Delft University of Technology, The Netherlands.
Carr, Peter, Jarrow, Robert, and Myneni, Ravi. 1992. "Alternative Characterizations of American Put Options." *Mathematical Finance* 2 (2): 87–106.
Carr, Peter, and Wu, Liuren. 2009. "Stock Options and Credit Default Swaps: A Joint Framework for Valuation and Estimation." *Journal of Financial Econometrics* 8 (4): 409–449.
Chance, Don M. 1996. "A Generalized Simple Formula to Compute the Implied Volatility." *Financial Review* 31 (4): 859–867.
Chen, Yangang, and Wan, Justin W. L. 2021. "Deep Neural Network Framework Based on Backward Stochastic Differential Equations for Pricing and Hedging American Options in High Dimensions." *Quantitative Finance* 21 (1): 45–67.
Christoffersen, Peter, Jacobs, Kris, and Young Chang, Bo. 2013. "Chapter 10 – Forecasting with Option-Implied Information." In Volume 2 of *Handbook of Economic Forecasting*. Amsterdam: Elsevier.
Corrado, Charles J., and Miller, Thomas W. 1996. "A Note on a Simple, Accurate Formula to Compute Implied Standard Deviations." *Journal of Banking and Finance* 20 (3): 595–603.
Donno, Marzia De, Palmowski, Zbigniew, and Tumilewicz, Joanna. 2020. "Double Continuation Regions for American and Swing Options with Negative Discount Rate in Lévy Models." *Mathematical Finance* 30 (1): 196–227.
Engström, Malin, and Nordén, Lars. 2000. "The Early Exercise Premium in American Put Option Prices." *Journal of Multinational Financial Management* 10 (3): 461–479.

Fang, Fang, and Oosterlee, Cornelis W. 2009. "Pricing Early-exercise and Discrete Barrier Options by Fourier-cosine Series Expansions." *Numerische Mathematik* 114 (1): 27.

Fodor, Andy, Stowe, David L., and Stowe, John D. 2017. "Option Implied Dividends Predict Dividend Cuts: Evidence From the Financial Crisis." *Journal of Business Finance and Accounting* 44 (5-6): 755–779.

Goodfellow, Ian, Bengio, Yoshua, and Courville, Aaron. 2016. *Deep Learning*. Cambridge, MA: MIT Press.

Guerrero, Robert. 2017. "Essays on Implied Dividends." PhD diss., UQ Business School, The University of Queensland.

Han, Jiequn, Jentzen, Arnulf, and Weinan, E. 2018. "Solving High-dimensional Partial Differential Equations Using Deep Learning." *Proceedings of the National Academy of Sciences* 115 (34): 8505–8510.

Hendrik Frankena, Luuk. 2016. "Pricing and Hedging Options in A Negative Interest Rate Environment." MSc thesis, Delft University of Technology, The Netherlands.

Hull, John C. 2019. "Derivative Securities: Options." Lecture 3. Accessed June 30, 2019. https://www.math.nyu.edu/faculty/avellane/DSLecture3.pdf.

Jourdain, B., and Martini, C. 2002. "Approximation of American Put Prices by European Prices Via an Embedding Method." *The Annals of Applied Probability* 12 (1): 196–223.

Kragt, Jac. 2016. "Option Implied Dividends." Working Paper, SSRN.

Kuen Kwok, Yue. 2008. *Mathematical Models of Financial Derivatives*. 2nd ed. Berlin: Springer Verlag.

Kutner, George W. 1998. "Determining the Implied Volatility for American Options Using the QAM." *The Financial Review* 33 (1): 119–130.

Lagnado, Ronald, and Osher, Stanley. 1997. "A Technique for Calibrating Derivative Security Pricing Models: Numerical Solution of An Inverse Problem." *Journal of Computational Finance* 1: 13–25.

Li, Weiping, and Chen, Su. 2018. "The Early Exercise Premium in American Options by Using Nonparametric Regressions." *International Journal of Theoretical and Applied Finance* 21 (7): 1850039.

Liu, Shuaiqiang, Borovykh, Anastasia, Grzelak, Lech A., and Oosterlee, Cornelis W. 2019. "A Neural Network-Based Framework for Financial Model Calibration." *Journal of Mathematics in Industry* 9 (1): 9.

Liu, Shuaiqiang, Oosterlee, Cornelis W., and Bohte, Sander M. 2019. "Pricing Options and Computing Implied Volatilities Using Neural Networks." *Risks* 7 (1): 16.

Lokeshwar, Vikranth, Bharadwaj, Vikram, and Jain, Shashi. 2022. "Explainable Neural Network for Pricing and Universal Static Hedging of Contingent Claims." *Applied Mathematics and Computation* 417: 126775.

Nardon, Martina, and Pianca, Paolo. 2013. "Extracting Information on Implied Volatilities and Discrete Dividends from American Options Prices." *Journal of Modern Accounting and Auditing* 9 (1): 112–129.

Poitras, Geoffrey, Veld, Chris, and Zabolotnyuk, Yuriy. 2009. "European Put-Call Parity and the Early Exercise Premium for American Currency Options." *Multinational Finance Journal* 13 (1/2): 39–54.

Ruf, Johannes, and Wang, Weiguan. 2020. "Neural Networks for Option Pricing and Hedging: A Literature Review." *Journal of Computational Finance* 24 (1): 1–46.

Salvador, Beatriz, Oosterlee, Cornelis W., and van der Meer, Remco. 2021. "Financial Option Valuation by Unsupervised Learning with Artificial Neural Networks." *Mathematics* 9 (1): 46.

Sirignano, Justin, and Spiliopoulos, Konstantinos. 2018. "DGM: A Deep Learning Algorithm for Solving Partial Differential Equations." *Journal of Computational Physics* 375: 1339–1364.

Won Seo, Sung, and Sik Kim, Jun. 2015. "The Information Content of Option-implied Information for Volatility Forecasting with Investor Sentiment." *Journal of Banking and Finance* 50: 106–120.

Zhang, Bowen, and Oosterlee, Cornelis W. 2012. "Fourier Cosine Expansions and Put-Call Relations for Bermudan Options." In *Numerical Methods in Finance*, 323–350. Berlin, Heidelberg: Springer Berlin Heidelberg.

Appendix. Pricing American options by the COS method

For the valuation of American options, we use Richardson extrapolation on a series of Bermudan options with an increasing number of exercise opportunities. We start with a Bermudan option, where the holder has the right to exercise the contract at pre-specified dates before maturity. With t_0 being initial time, we assume there are M pre-specified exercise dates, and have $\{t_1, \ldots, t_M\}$ as the collection of all exercise dates. The regular time interval reads $\Delta t := (t_m - t_{m-1})$, $t_0 < t_1 < \cdots < t_M = T$. With the help of the risk-neutral valuation, we arrive at the pricing formula for a Bermudan option with M exercise dates, for $m = M, M-1, \ldots, 2$:

$$\begin{cases} c(t_{m-1}, x) & = e^{-r\Delta t} \int_{\mathbb{R}} V(t_m, y) f(y \mid x)\, \mathrm{d}y, \\ V_{ber}(t_{m-1}, x) & = \max\left(h(t_{m-1}, x), c(t_{m-1}, x)\right), \end{cases} \tag{A1}$$

followed by

$$V_{ber}(t_0, x) = e^{-r\Delta t} \int_{\mathbb{R}} V(t_1, y) f(y \mid x)\, \mathrm{d}y. \tag{A2}$$

where $f(\cdot)$ is the conditional density function, and the state variables x and y are the log-prices and separately defined as

$$x := \log(S(t_{m-1})/K) \quad \text{and} \quad y := \log(S(t_m)/K),$$

where $S(t_m)$ stands for the stock price at time t_m, and K for the strike price. Functions $V(t, x)$, $c(t, x)$ and $h(t, x)$ represent the option value, the continuation value and the log-price payoff at time t, respectively, for example,

$$h(T, x) = \max\left[\alpha K(e^x - 1), 0\right], \quad \alpha = \begin{cases} 1 & \text{for a call,} \\ -1 & \text{for a put.} \end{cases} \tag{A3}$$

A.1 Pricing Bermudan Options

The COS method is generally based on employing a Fourier cosine expansion to approximate the density function on a truncated domain $[a, b]$,

$$f(y \mid x) \approx \frac{2}{b-a} {\sum_{j=0}^{N-1}}' \Re\left(\varphi\left(\frac{j\pi}{b-a}; x, \Delta t\right) \exp\left(-i\frac{ak\pi}{b-a}\right)\right) \cos\left(k\pi \frac{y-a}{b-a}\right), \tag{A4}$$

where $\varphi(u; x, t)$ represents the characteristic function of the log-asset price, $x := \log(S(t)/K)$, and the notation ${\sum}'$ means that the first term in the summation is weighted by one-half.

With the centre of the interval $x_0 := \log(S_0/K)$, the integration range, $[a, b]$, is defined as follows,

$$[a, b] := \left[(\xi_1 + x_0) - L\sqrt{\xi_2}, (\xi_1 + x_0) + L\sqrt{\xi_2}\right], \tag{A5}$$

where L is a user-defined parameter to achieve a certain integration accuracy, and parameters ξ_i represent the corresponding cumulants of the underlying stochastic process, see Zhang and Oosterlee (2012).

We define the following formula,

$$\phi(t, u) := \varphi(u; x = 0, t). \tag{A6}$$

For the Black-Scholes dynamics in Formula (1) under the log-asset price, we have

$$\phi(t, u) = \exp\left(iut\left(r - q - \frac{1}{2}\sigma^2\right) - \frac{1}{2}\sigma^2 u^2 t\right),$$

$$\xi_1 = \left(r - q - \frac{1}{2}\sigma^2\right)t, \quad \xi_2 = \frac{1}{2}\sigma^2 t, \quad \xi_4 = 0.$$

The continuation value in (A1), which resembles a European option between two consecutive exercise dates, can be computed through the COS formula,

$$c(t_{m-1},x)=e^{-r\Delta t}\sum_{k=0}^{N-1}{}'\,\Re\left(\phi\left(\Delta t,\frac{k\pi}{b-a}\right)e^{ik\pi\frac{x-a}{b-a}}\right)\mathcal{V}_k(t_m), \tag{A7}$$

where N is the number of Fourier cosine items. The $\mathcal{V}_k(t_m)$ terms are the so-called option coefficients, to be computed depending on the early-exercise region.

An early-exercise point, x_m^*, at time t_m, is a point where the continuation value equals the payoff, i.e., $c(x_m^*,t_m)=H(x_m^*,t_m)$. We will first derive the induction formula for $\mathcal{V}_k(t_1)$ for a single early-exercise point, and then extend it to the case of two early-exercise points. The early-exercise point is determined by means of a root-finding algorithm, for example, Newton's method. With x_m^*, the option coefficients $\mathcal{V}_k(t_m)$ can be split into two components: One on the interval $[a,x_m^*]$ and the other on $(x_m^*,b]$ (i.e., on the holding or stopping region),

$$\mathcal{V}_k(t_m)=\begin{cases}C_k(a,x_m^*,t_m)+G_k(x_m^*,b), & \text{for a call,}\\ G_k(a,x_m^*)+C_k(x_m^*,b,t_m), & \text{for a put,}\end{cases} \tag{A8}$$

for $m=M-1,M-2,\ldots,1$. When $t_m=t_M$ at the terminal time,

$$\mathcal{V}_k(t_M)=\begin{cases}G_k(0,b), & \text{for a call,}\\ G_k(a,0), & \text{for a put.}\end{cases} \tag{A9}$$

With the COS method, we have

$$G_k(x_1,x_2):=\frac{2}{b-a}\int_{x_1}^{x_2}h(t_m,x)\cos\left(k\pi\frac{x-a}{b-a}\right)\mathrm{d}x, \tag{A10}$$

and

$$C_k(x_1,x_2,t_m):=\frac{2}{b-a}\int_{x_1}^{x_2}c(t_m,x)\cos\left(k\pi\frac{x-a}{b-a}\right)\mathrm{d}x, \tag{A11}$$

for $k=0,1,\ldots,N-1$ and $m=1,2,\ldots,M$, there are analytic solutions for $G_k(x_1,x_2)$ in (A10), since the payoff function $h(x,t_m)$ is known. The terms $C_k(x_1,x_2,t_m)$ can be computed in $O(N\log_2 N)$ operations under the Black-Scholes dynamics.

At times t_m, $m=1,2,\ldots,M$, from Equations (A1) and (A7), we obtain an approximation for $c(t_m,x)$. Afterwards, $c(t_m,x)$ is inserted into (A11). Interchanging summation and integration gives the following coefficients, $C_k(x_1,x_2,t_m)$:

$$C_k(x_1,x_2,t_m):=e^{-r\Delta t}\sum_{j=0}^{N-1}{}'\,\Re\left(\phi\left(\Delta t,\frac{j\pi}{b-a}\right)\mathcal{V}_j(t_{m+1})\cdot H_{k,j}(x_1,x_2)\right), \tag{A12}$$

where $H_{k,j}(x_1,x_2)$ is computed in the following integrals,

$$H_{k,j}(x_1,x_2)=\frac{2}{b-a}\int_{x_1}^{x_2}e^{ij\pi\frac{x-a}{b-a}}\cos\left(k\pi\frac{x-a}{b-a}\right)\mathrm{d}x.$$

With the help of basic calculus, the term $H_{k,j}(x_1,x_2)$ can be further divided into two parts,

$$H_{k,j}(x_1,x_2)=-\frac{i}{\pi}(H^s_{k,j}(x_1,x_2)+H^c_{k,j}(x_1,x_2)).$$

Because $H^s_{k,j}(x_1,x_2)=H^s_{k+1,j+1}(x_1,x_2)$ and $H^c_{k,j}(x_1,x_2)=H^c_{k+1,j-1}(x_1,x_2)$, we get a Toeplitz and Hankel structure in the matrices H_s and H_c, respectively. Therefore the Fast Fourier Transform can be employed for highly efficient matrix-vector multiplication, and the resulting computational complexity of $C_k(x_1,x_2,t_m)$ is reduced to $O(N\log_2 N)$. In addition, the Greeks of American

Black-Scholes option prices can be easily approximated based on Equation (A7), for example,

$$\text{Vega} \approx \frac{\partial c(t_{m-1},x)}{\partial \sigma} = e^{-r\Delta t} \sum_{k=0}^{N-1}{}' \Re\left(\frac{\partial \phi\left(\Delta t, \frac{k\pi}{b-a}\right)}{\partial \sigma} e^{ik\pi \frac{x-a}{b-a}} \right) \mathcal{V}_k(t_m), \tag{A13}$$

Here we extend this Bermudan COS method to deal with two early-exercise points. Suppose there are at most two early-exercise points $x^*_{m1} < x^*_{m2}$ at time t_m. Thus, we have three intervals while computing $\mathcal{V}_k(t_m)$, that is, $[a, x^*_{m1}]$, $[x^*_{m1}, x^*_{m2}]$, and $[x^*_{m2}, b]$. Take a Bermudan put as an example,

$$\mathcal{V}_k(t_m) = C^1_k(a, x^*_{m1}, t_m) + G_k(x^*_{m1}, x^*_{m2}) + C^2_k(x^*_{m2}, b, t_m), \tag{A14}$$

for $m = M-1, M-2, \ldots, 1$ and $\mathcal{V}_k(t_M) = G_k(a, 0)$ when $t_m = t_M$. There are two continuation values C^1_k and C^2_K correspondingly. In such case, two roots x^*_{m1} and x^*_{m2} are computed, with different, smartly selected, starting points for the Newton's method.

A.2 Pricing American Options

The extrapolation-based pricing method combined with the COS computations for Bermudan options to price American options has been described in Fang and Oosterlee (2009), and a brief description is given in this section.

Let $V_{ber}(M)$ denote the value of a Bermudan option with M exercise dates, considering maturity T and $\Delta t = T/M$ which is a time interval between two consecutive exercise dates, the American option value V_{am} can be approximated by applying the following 4-point Richardson extrapolation scheme,

$$V_{am}(\ell) \approx \frac{1}{21}\left(64V_{ber}(2^{\ell+3}) - 56V_{ber}(2^{\ell+2}) + 14V_{ber}(2^{\ell+1}) - V_{ber}(2^{\ell})\right), \tag{A15}$$

where parameter ℓ determines the number of the exercise dates considered for each Bermudan option involved.

Remark: The binomial trees technique is also a suitable candidate for American option valuation, since it can accurately deal with two early-exercise points when pricing American options. However, the COS method provides faster computation and the Greeks come without any extra cost.