Multilevel dual approach for pricing American style derivatives

Denis Belomestny[a] and John Schoenmakers[b]

[a] Duisburg-Essen University, Duisburg, Germany denis.belomestny@uni-due.de

[b] Weierstraß Institute (WIAS), Berlin, Germany schoenma@wias-berlin.de

, J. S. acknowledges the financial support by the DFG Research Center Matheon "Mathematics for Key Technologies" in Berlin.

In this article we propose a novel approach to reduce the computational complexity of the dual method for pricing American options. We consider a sequence of martingales that converges to a given target martingale and decompose the original dual representation into a sum of representations that correspond to different levels of approximation to the target martingale. By next replacing in each representation true conditional expectations with their Monte Carlo estimates, we arrive at what one may call a multilevel dual Monte Carlo algorithm. The analysis of this algorithm reveals that the computational complexity of getting the corresponding target upper bound, due to the target martingale, can be significantly reduced. In particular, it turns out that using our new approach, we may construct a multilevel version of the well-known nested Monte Carlo algorithm of Andersen and Broadie (2004) that is, regarding complexity, virtually equivalent to a non-nested algorithm. The performance of this multilevel algorithm is illustrated by a numerical example.

Keywords: optimal stopping, dual approach, multilevel Monte Carlo

AMS Subject Classification: 62L15, 65C05, 91B28

1 Introduction

Efficient methods for pricing high-dimensional American options have been a challenge for decades. While for a low or moderate dimensional underlying process, deterministic (PDE) based methods may be applicable, for higher dimensions Monte Carlo simulation based methods are virtually the only way out. Besides the absence of curse of dimensionality, Monte Carlo methods are quite popular because of their genericity. In the nineties a number of regression methods for constructing "good" exercise policies, hence price lower bounds, were introduced and studied in the literature (see Carriere (1996), Longstaff and Schwartz (2001), and Tsistsiklis and Van Roy (2001)). Among various other developments, we mention the stochastic mesh method of Broadie and Glasserman (2004), the quantization method by Bally and Pages (2003), and the policy iteration method by Kolodko and Schoenmakers (2006). The later method may

be effectively combined with the Longstaff-Schwartz approach as presented in Bender et al. (2008).

The aforementioned methods have in common that they provide an exercise policy for the American product. Based on this policy one may simulate a lower bound for its price. This is what is called the primal approach. A new direction in Monte Carlo simulation of American options was the dual approach, developed by Rogers (2002) and independently by Haugh and Kogan (2004), related to earlier ideas in Davis and Karatzas (1994). In this approach one looks for a "good" martingale rather than a "good" stopping time. Based on this martingale the price of an American derivative may be bounded from above. In fact, this price upper bound may be represented by a "look-back" option due to the difference of the cash-flow and the martingale. Meanwhile, several numerical algorithms for computing dual upper bounds have appeared in the literature. Probably one of the most popular ones is the method of Andersen and Broadie (2004) (see also Kolodko and Schoenmakers (2004) for an extension, and Chen and Glasserman (2007) for a more detailed analysis of the dual estimator). A drawback of this method is its computational complexity due to the need of nested Monte Carlo simulation however. As a remedy to this issue, Belomestny et al. (2009) proposed a dual simulation algorithm which does not require nested simulation and uses regression to approximate the integrand in a martingale representation. Kohler et al. (2010) proposed a regression based variance reduction method for the nested dual approach. Another non-nested regression based dual algorithm was proposed in Schoenmakers and Huang (2011) in the context of a study of optimal dual martingales (approximated by "low variance" martingales). Furthermore, in Belomestny (2011) an efficient dual algorithm is constructed which is based on convex optimization and empirical variance penalization. For another dual approach based on convex optimization see Desai, et al. (2010). The main goal of this paper is to enhance the efficiency of dual algorithms by using a "multi-level" idea in the spirit of Giles (2008). In Giles (2008) a multilevel Monte Carlo estimator is presented, which is based on approximate solutions of a stochastic differential equation given a sequence of different time discretisation steps. For instance, by this method the complexity of simulating a European option can be significantly reduced. In this paper we apply the multilevel idea to a sequence of martingales (rather than time discretisation steps). Based on this sequence of martingales we will construct a new multilevel dual estimator for the American option. As a special case we so obtain a multilevel version of the Andersen-Broadie algorithm. Under some assumptions we will prove that the complexity of this algorithm is (almost) equivalent to a non-nested Monte Carlo algorithm. As a byproduct of our complexity analysis we derive, to our knowledge for the first time, convergence rates of the Andersen-Broadie algorithm. In particular, our analysis reveals that, under some assumptions, the upper bias induced by inner simulations converges to zero at rate $O(1/k)$, with k being the number of inner simulations. The latter feature was observed empirically in the early literature (see, e.g., Kolodko and Schoenmakers (2004)), but has not yet got theoretical explanation.

The structure of the paper is as follows. In Section 2 we recap the primal and dual approaches for optimal stopping in the context of American options. The main setup and prerequisites including a key Theorem 5 are presented in Section 3. After a complexity analysis of the standard dual approach in Section 4 we present and analyze the multi-level dual estimator in Section 5. Section 6 concludes with a detailed numerical study of the multilevel version of the Andersen-Broadie algorithm.

2 Primal and Dual valuation of American options

Let $(Z_j)_{j\geq 0}$ be a nonnegative adapted process on a filtered probability space $(\Omega, \mathcal{F} = (\mathcal{F}_j)_{j\geq 0}, \mathrm{P})$ representing the discounted payoff of an American option, so that the holder of the option receives Z_j if the option is exercised at time $j \in \{0, ..., \mathcal{J}\}$ with $\mathcal{J} \in \mathbb{N}_+$. The pricing of American options can be formulated as a primal-dual problem. Let Y_j denote the time j solution to this problem. The primal representation corresponds to the following optimal stopping problems:

$$Y_j^* = \max_{\tau \in \mathcal{T}[j,\ldots,\mathcal{J}]} \mathrm{E}_{\mathcal{F}_j}[Z_\tau], \quad j = 0, \ldots, \mathcal{J},$$

where $\mathcal{T}[j, \ldots, \mathcal{J}]$ is the set of $\mathbb{F}$-stopping times taking values in $\{j, \ldots, \mathcal{J}\}$. During the nineties the primal approach was the only method available. More recently a quite different "dual" approach has been discovered by Rogers (2002) and Haugh and Kogan (2004). The next theorem summarizes their results.

Theorem 1 *Let $\mathcal{M}$ denote the space of adapted martingales, then we have the following dual representation for the value process Y_j^**

$$\begin{aligned} Y_j^* &= \inf_{M \in \mathcal{M}} \mathrm{E}_{\mathcal{F}_j} \left[\max_{s \in \{j,\ldots,\mathcal{J}\}} (Z_s - M_s + M_j) \right] \\ &= \max_{s \in \{j,\ldots,\mathcal{J}\}} (Z_s - M_s^* + M_j^*) \quad a.s., \end{aligned}$$

where

$$Y_j^* = Y_0^* + M_j^* - A_j^* \tag{1}$$

is the (unique) Doob decomposition of the supermartingale Y_j^*. That is, M^* is a martingale and A^* is an increasing process with $M_0 = A_0 = 0$ such that (1) holds.

Remark 2 *In Schoenmakers and Huang (2011) it is shown that in general there are infinitely many martingales M° such that*

$$Y_j^* = \max_{s \in \{j,\ldots,\mathcal{J}\}} (Z_s - M_s^\circ + M_j^\circ) \quad a.s.$$

Theorem 1 implies that, for an arbitrarily chosen martingale M with $M_0 = 0$, the value

$$\mathrm{E}\left[\max_{s \in \{0,\ldots,\mathcal{J}\}} (Z_s - M_s) \right]$$

defines an upper bound for the price of American option Y_0^*, and the upper bound will be tight if the chosen martingale M is close to the Doob martingale part M^* of the discounted true value process Y_j^*, which we shall refer to as the "optimal" martingale. Not surprisingly, finding such an optimal martingale is no easier than solving the original stopping problem. The so-called martingale duality approach aims at finding a martingale that approximates the "optimal" martingale and then use this approximation to compute an upper bound for the price of the American option by Monte Carlo. There are several methods known for approximating the "optimal" martingale M^*. In Andersen and Broadie (2004) the Doob martingale part of a given approximation to the Snell envelope is constructed by sub-simulations. In this way quite tight lower and upper bounds for a number of test examples were obtained. However, as we will see later on, the complexity of this method can be rather high, especially if a high precision of calculation is required. Therefore the complexity reduction of the dual algorithms is of primal importance and is one of the main aims of this paper.

3 Main setup and prerequisites

Let $(M^k)_{k\in\mathbb{N}}$ be a sequence of martingales starting at 0 ($M_0^k = 0$) with respect to an enlarged probability space $(\Omega, \mathcal{F}' = (\mathcal{F}'_j)_{j\geq 0}, P)$, where $\mathcal{F}_j \subset \mathcal{F}'_j$ for each j. It is supposed that the sequence $(M^k)_{k\in\mathbb{N}}$ converges in some sense to a target martingale M which is $\mathcal{F}$-adapted. We assume that

(AC) the numerical complexity of obtaining a single realization of M_j^k is of order $O(k)$ for each $j = 1, \ldots, \mathcal{J}$,

(AA) there exists an $\mathcal{F}$-adapted martingale M such that

$$\max_{j=0,\ldots,\mathcal{J}} \left|\mathrm{E}_{\mathcal{F}}\left[M_j^k - M_j\right]\right| \leq Ak^{-\alpha}, \quad \mathrm{E}_{\mathcal{F}}\left[\max_{j=0,\ldots,\mathcal{J}}(M_j^k - M_j)^2\right] \leq Bk^{-\beta},$$

almost surely for all $k \in \mathbb{N}$, some $\alpha > 0$, $\beta > 0$, $A > 0$ and $B > 0$.

Remark 3 *If $\mathcal{J}$ is finite, then*

$$\mathrm{E}_{\mathcal{F}}\left[\max_{j=0,\ldots,\mathcal{J}}(M_j^k - M_j)^2\right] \leq \sum_{j=1}^{\mathcal{J}} \mathrm{E}_{\mathcal{F}}\left[(M_j^k - M_j)^2\right] \leq \mathcal{J} \max_{j=0,\ldots,\mathcal{J}} \mathrm{E}_{\mathcal{F}}\left[(M_j^k - M_j)^2\right]$$

and the second condition in (AA) can be formulated in terms of expectations $\mathrm{E}_{\mathcal{F}}\left[(M_j^k - M_j)^2\right]$.

Our aim is to approximate the target upper bound

$$Y(M) := \mathrm{E}\left[\max_{j=0,\ldots,\mathcal{J}}(Z_j - M_j)\right] = \mathrm{E}[\mathcal{Z}(M)]$$

with

$$\mathcal{Z}(M) = \max_{j=0,\ldots,\mathcal{J}}(Z_j - M_j).$$

Note that any $\mathcal{F}$-stopping time τ is also an $\mathcal{F}'$-stopping time, so by Doob's sampling theorem $\mathrm{E}\left[M_\tau^k\right] = 0$, and we have

$$\begin{aligned} Y_0^* &= \sup_{\mathcal{F}\text{-stopping times } \tau} \mathrm{E}\left[Z_\tau\right] \\ &= \sup_{\mathcal{F}\text{-stopping times } \tau} \mathrm{E}\left[Z_\tau - M_\tau^k\right] \\ &\leq \mathrm{E}\left[\max_{j=0,\ldots,\mathcal{J}}(Z_j - M_j^k)\right] = Y(M^k), \end{aligned}$$

i.e., $Y(M^k)$ is an upper bound for any $k \in \mathbb{N}$. Note that $\widetilde{M}_j^k := \mathrm{E}_{\mathcal{F}_j}\left[M_j^k\right]$, $j = 0, \ldots, \mathcal{J}$, is an $\mathcal{F}$-martingale, and if M^k satisfies for each j,

$$\widetilde{M}_j^k = \mathrm{E}_{\mathcal{F}}\left[M_j^k\right],$$

then $Y(M^k)$ is upper biased with respect to $Y(\widetilde{M}^k)$. Indeed, by Jensen's inequality, we then have

$$Y(M^k) \geq \mathrm{E}\left[\max_{j=0,\ldots,\mathcal{J}}\left(Z_j - \mathrm{E}_{\mathcal{F}}\left[M_j^k\right]\right)\right] = \mathrm{E}\left[\max_{j=0,\ldots,\mathcal{J}}\left(Z_j - \widetilde{M}_j^k\right)\right] = Y(\widetilde{M}^k).$$

Introduce the random sets

$$\mathcal{Q} \;= \{j : Z_j - M_j = \mathcal{Z}(M)\}, \quad \mathcal{Q}^k \;= \left\{j : Z_j - M_j^k = \mathcal{Z}(M^k)\right\}, \quad k \in \mathbb{N},$$

and define the $\mathcal{F}$-measurable random variable

$$\Lambda := \min_{j \notin \mathcal{Q}} \left(\mathcal{Z}(M) - Z_j + M_j\right),$$

with $\Lambda := +\infty$ if $\mathcal{Q} = \{0, ..., \mathcal{J}\}$. Let us introduce the following two conditions.

(AL) Λ satisfies $\mathrm{E}[\Lambda^{-1}] < \infty$,

(AQ) $\#\mathcal{Q} = 1$ a.s.

Remark 4 *If $M = M^*$, then*

$$\begin{aligned}
\Lambda &= \min_{j \notin \mathcal{Q}} \left(Y_0^* - Z_j + M_j\right) \\
&= \min_{j \notin \mathcal{Q}} \left(Y_j^* - Z_j + A_j\right) \\
&\geq \min_{j \notin \mathcal{Q}} \left(Y_j^* - Z_j\right) \geq \min_{\{j : Y_j^* > Z_j\}} \left(Y_j^* - Z_j\right).
\end{aligned}$$

Hence, the condition (AL) is fulfilled if

$$\mathrm{P}(Y_j^* - Z_j < \delta | Y_j^* > Z_j) \lesssim \delta^{q+1}, \quad \delta \to 0 \tag{2}$$

for all $j = 1, \ldots, \mathcal{J}$ and some $q > 0$. The condition (2) bounds the probability of staying in the δ-vicinity of the exercise boundary $\{Y_j^* \leq Z_j\}$ in the case of continuation and is similar to the so-called margin condition in Belomestny (2011). As we will see, (2) leads to faster convergence rates of the standard Andersen-Broadie dual algorithm. It is easy to construct examples showing that the assumption (AL) is not always fulfilled. Take $\mathcal{J} = 1$, $Z_0 = 0$, $Z_1 = \zeta$ for some random variable ζ with absolutely continuous density, and $M_0 = M_1 = 0$. Then, obviously, $\Lambda = |\zeta|$ and $\#\mathcal{Q} = 1$ a.s., hence (AQ) holds. Now, if ζ has a density not vanishing at 0, then (AL) does not hold. If, on the contrary, the density of ζ behaves like $O(|z|^q)$ for some $q > 0$ as $|z| \to 0$, then (AL) holds (cf. (2)).

Theorem 5 *Under assumption (AA) it holds*

$$|\,\mathrm{E}[\mathcal{Z}(M^k) - \mathcal{Z}(M)]| \leq Ck^{-\gamma}, \quad \mathrm{E}\left(\mathcal{Z}(M^k) - \mathcal{Z}(M)\right)^2 \leq Bk^{-\beta} \tag{3}$$

with $\gamma = \beta/2$ and some $C > 0$. If additionally the assumptions (AL) and (AQ) are satisfied, then (3) holds true with $\gamma = \min\{\alpha, \beta\}$.

Proof. On the one hand, it holds for each $k \in \mathbb{N}$, and $j_k^{\max} := \min \mathcal{Q}^k$

$$\begin{aligned} \mathcal{Z}(M^k) - \mathcal{Z}(M) &= \max_{j=0,\dots,\mathcal{J}}(Z_j - M_j^k) - \max_{j=0,\dots,\mathcal{J}}(Z_j - M_j) \\ &\leq M_{j_k^{\max}} - M_{j_k^{\max}}^k, \quad \text{a.s.}, \end{aligned} \tag{4}$$

and on the other hand, we get for each $k \in \mathbb{N}$ and $\mathcal{Q} =: \{j^{\max}\}$,

$$\mathcal{Z}(M^k) - \mathcal{Z}(M) \geq M_{j^{\max}} - M_{j^{\max}}^k. \tag{5}$$

By (4) and (5) we thus have

$$\mathrm{E}\left[\left(\mathcal{Z}(M^k) - \mathcal{Z}(M)\right)^2\right] \leq \mathrm{E}\,\mathrm{E}_{\mathcal{F}}\left[\max_{j=1,\dots,\mathcal{J}}\left(M_j - M_j^k\right)^2\right] \leq Bk^{-\beta}.$$

Further, by the Cauchy-Schwarz inequality we have similarly,

$$|\,\mathrm{E}[\mathcal{Z}(M^k) - \mathcal{Z}(M)]| \leq \mathrm{E}\left[\left\{\mathrm{E}_{\mathcal{F}}\left[\max_{j=1,\dots,\mathcal{J}}\left(M_j - M_j^k\right)^2\right]\right\}^{1/2}\right] \leq \sqrt{B}\cdot k^{-\beta/2}.$$

Let us now turn to the case where assumptions (AL) and (AQ) are fulfilled. From (4) we obtain for $k \in \mathbb{N}$,

$$\begin{aligned} \mathrm{E}_{\mathcal{F}}[\mathcal{Z}(M^k) - \mathcal{Z}(M)] &\leq \mathrm{E}_{\mathcal{F}}\left[M_{j_k^{\max}} - M_{j_k^{\max}}^k\right] \\ &= \mathrm{E}_{\mathcal{F}}\left[\left(M_{j_k^{\max}} - M_{j_k^{\max}}^k\right) 1_{j_k^{\max} \neq j^{\max}}\right] \\ &\quad + \mathrm{E}_{\mathcal{F}}\left[M_{j^{\max}} - M_{j^{\max}}^k\right] =: (I) + (II). \end{aligned}$$

Note that $Z_{j_k^{\max}} - M_{j_k^{\max}}^k \geq Z_{j^{\max}} - M_{j^{\max}}^k$ and hence

$$\{j_k^{\max} \neq j^{\max}\} \subset \left\{\max_{j=1,\dots,\mathcal{J}}(M_j - M_j^k + M_{j^{\max}}^k - M_{j^{\max}}) \geq \Lambda\right\}.$$

We thus have

$$\begin{aligned} \mathrm{P}_{\mathcal{F}}(j_k^{\max} \neq j^{\max}) &\leq \mathrm{P}_{\mathcal{F}}\left(\max_{j=1,\dots,\mathcal{J}}(M_j - M_j^k + M_{j^{\max}}^k - M_{j^{\max}}) \geq \Lambda\right) \\ &\leq \mathrm{P}_{\mathcal{F}}\left(\max_{j=1,\dots,\mathcal{J}}(M_j - M_j^k) \geq \Lambda/2\right) + \mathrm{P}_{\mathcal{F}}\left(M_{j^{\max}}^k - M_{j^{\max}} \geq \Lambda/2\right) \\ &\leq \mathrm{P}_{\mathcal{F}}\left(\max_{j=1,\dots,\mathcal{J}}(M_j - M_j^k) \geq \Lambda/2\right) + \mathrm{P}_{\mathcal{F}}\left(\max_{j=1,\dots,\mathcal{J}}(M_j^k - M_j) \geq \Lambda/2\right) \end{aligned}$$

By (AA) and the conditional Markov inequality it follows that

$$\mathrm{P}_{\mathcal{F}}\left(\max_{j=1,\dots,\mathcal{J}}(M_j - M_j^k) \geq \Lambda/2\right) \leq \frac{4B}{\Lambda^2}k^{-\beta}, \quad \mathrm{P}_{\mathcal{F}}\left(\max_{j=1,\dots,\mathcal{J}}(M_j^k - M_j) \geq \Lambda/2\right) \leq \frac{4B}{\Lambda^2}k^{-\beta}.$$

Hence

$$\mathrm{P}_{\mathcal{F}}(j_k^{\max} \neq j^{\max}) \leq \frac{8B}{\Lambda^2}k^{-\beta}$$

for all k. Furthermore, we have by (AA) and the Cauchy-Schwarz inequality

$$\begin{aligned}(I) &= \mathrm{E}_{\mathcal{F}}\left[\left(M_{j_k^{\max}} - M^k_{j_k^{\max}}\right) 1_{j_k^{\max} \neq j^{\max}}\right] \\ &\leq \sqrt{\mathrm{P}_{\mathcal{F}}(j_k^{\max} \neq j^{\max})}\sqrt{\mathrm{E}_{\mathcal{F}}\left[\max_{j=1,\dots,\mathcal{J}}\left(M_j - M_j^k\right)^2\right]} \leq \frac{2\sqrt{2}B}{\Lambda}k^{-\beta}.\end{aligned}$$

Combining (4) with (5) and using assumption (AA) again for the term (II), we arrive at the inequality

$$-Ak^{-\alpha} \leq \mathrm{E}[\mathcal{Z}(M^k) - \mathcal{Z}(M)] \leq 2\sqrt{2}B\,\mathrm{E}[\Lambda^{-1}]k^{-\beta} + Ak^{-\alpha} \leq Ck^{-\gamma}$$

with $\gamma = \min\{\alpha, \beta\}$ and some $C > 0$. ■

Remark 6 *For the Andersen-Broadie algorithm, where the sequence of martingales (M^k) is constructed using subsimulations, it is not difficult to show that assumption (AA) is fulfilled with $\alpha = \infty$ and $\beta = 1$. Hence, under the additional assumptions (AL) and (AQ) we have by Theorem 5 that the bias of the Andersen-Broadie estimator is of order $O(1/k)$. This rate was experimentally found in Kolodko and Schoenmakers (2004), but not proved by now to the best of our knowledge.*

4 Standard dual approach

Fix some natural numbers N and K, and consider the estimator

$$\begin{aligned}Y^{N,K} &= \frac{1}{N}\sum_{n=1}^{N} \max_{j=0,\dots,\mathcal{J}}(Z_j^{(n)} - M_j^{K,(n)}) \\ &=: \frac{1}{N}\sum_{n=1}^{N} \mathcal{Z}^{(n)}(M^K)\end{aligned}$$

based on a set of trajectories

$$\left\{(Z_j^{(n)}, M_j^{K,(n)}),\, n = 1, ..., N,\, j = 0, \dots, \mathcal{J}\right\}$$

of the vector process (Z, M^K).

Complexity analysis

By Theorem 5, for some $C > 0$ it holds

$$\begin{aligned}\mathrm{E}\left[Y^{N,K} - Y(M)\right]^2 &\leq N^{-1}\,\mathrm{Var}(\mathcal{Z}(M^K)) + CK^{-2\gamma} \\ &=: N^{-1}v_K + CK^{-2\gamma}, \quad K \to \infty.\end{aligned}$$

In order to get $\sqrt{\mathrm{E}\left[Y^{N,K} - Y(M)\right]^2} \leq \varepsilon$, we may take

$$K = \frac{(2C)^{1/2\gamma}}{\varepsilon^{1/\gamma}}$$

and then

$$N = \frac{2v_K}{\varepsilon^2}.$$

Assuming that v_K is non-increasing, the required computation time for reaching accuracy ε, hence the complexity, is then given, up to a constant, by

$$\begin{aligned}\mathcal{C}^{N,K}(\varepsilon) := NK &= \frac{2v_K}{\varepsilon^2}\frac{(2C)^{1/2\gamma}}{\varepsilon^{1/\gamma}} \\ &\lesssim \frac{v_{\left\lfloor \frac{(2C)^{1/2\gamma}}{\varepsilon^{1/\gamma}} \right\rfloor}}{\varepsilon^{2+1/\gamma}}.\end{aligned}$$

with $\lfloor x \rfloor$ denoting the largest integer which is smaller than or equal to x.

In the usual case, where M is the Doob martingale of some approximation Y to the Snell envelope Y^*, we have that $\mathrm{Var}(\mathcal{Z}(M)) > 0$ and $v_K \to v_\infty \neq 0$ leading to $\mathcal{C}^{N,K}(\varepsilon)$ of order $O(1/\varepsilon^{2+1/\gamma})$.

Remark 7 *If* $\mathrm{Var}(\mathcal{Z}(M)) = 0$ *(e.g., the target martingale* M *is the Doob martingale of* Y^**) we have (under the assumptions of Theorem 5),*

$$v_K = \mathrm{Var}(\mathcal{Z}(M^K)) \leq \mathrm{E}(\mathcal{Z}(M^K) - \mathcal{Z}(M))^2 \leq BK^{-\beta},$$

and as a result

$$\mathcal{C}^{N,K}(\varepsilon) \lesssim \frac{B}{(2C)^{\beta/2\gamma}}\frac{1}{\varepsilon^{2+(1-\beta)/\gamma}}.$$

That is, if $\beta \geq 1$ *the complexity is even less than or equal to the complexity of the plain Monte Carlo algorithm !*

5 Multilevel dual algorithm

Fix some natural number $L > 0$. Let $\mathbf{k} = (k_0, \ldots, k_L)$ be a sequence of natural numbers satisfying $1 \leq k_0 < k_1 < \ldots < k_L$. Write

$$\begin{aligned}Y(M^{k_L}) &= Y(M^{k_0}) + \sum_{l=1}^{L}[Y(M^{k_l}) - Y(M^{k_{l-1}})] \\ &= \mathrm{E}[\mathcal{Z}(M^{k_0})] + \sum_{l=1}^{L}\mathrm{E}[\mathcal{Z}(M^{k_l}) - \mathcal{Z}(M^{k_{l-1}})].\end{aligned} \tag{6}$$

For a given sequence $\mathbf{n} = (n_0, \ldots, n_L)$ with $1 \leq n_0 < \ldots < n_L$, we first simulate the initial set of trajectories

$$\left\{\left(Z_j^{(i)}, M_j^{k_0,(i)}\right), \quad i = 1, ..., n_0, \quad j = 0, \ldots, \mathcal{J}\right\}$$

of the vector process (Z, M^{k_0}), and then for each level $l = 1, \ldots, L$ *independently* a set of trajectories

$$\left\{\left(Z_j^{(i)}, M_j^{k_{l-1},(i)}, M_j^{k_l,(i)}\right), \quad i = 1, ..., n_l, \quad j = 0, \ldots, \mathcal{J}\right\}$$

of the vector process $(Z, M^{k_{l-1}}, M^{k_l})$. Based on these simulations we consider the following approximation of (6)

$$Y^{\mathbf{n},\mathbf{k}} := \frac{1}{n_0}\sum_{i=1}^{n_0} \mathcal{Z}^{(i)}(M^{k_0}) + \sum_{l=1}^{L} \frac{1}{n_l}\sum_{i=1}^{n_l} \left[\mathcal{Z}^{(i)}(M^{k_l}) - \mathcal{Z}^{(i)}(M^{k_{l-1}})\right]$$

with $\mathcal{Z}^{(i)}(M^k) := \max_{j=0,\dots,\mathcal{J}} \left(Z_j^{(i)} - M_j^{k,(i)}\right)$, $i = 1, \dots, n_l$, $k \in \mathbb{N}$, where $M^{k,(i)}$ denotes the i-th simulated trajectory of the martingale M^k.

Complexity analysis

For the bias of the multilevel estimator we obtain by Theorem 5

$$\left|\mathrm{E}\left[Y^{\mathbf{n},\mathbf{k}}\right] - Y(M)\right| = \left|\mathrm{E}\left[\mathcal{Z}(M^{k_L})\right] - \mathrm{E}\left[\mathcal{Z}(M)\right]\right| \leq C k_L^{-\gamma} \tag{7}$$

and for the variance we have

$$\mathrm{Var}\left[Y^{\mathbf{n},\mathbf{k}}\right] = n_0^{-1}\,\mathrm{Var}[\mathcal{Z}(M^{k_0})] + \sum_{l=1}^{L} \frac{1}{n_l}\,\mathrm{Var}\left[\mathcal{Z}(M^{k_l}) - \mathcal{Z}(M^{k_{l-1}})\right]. \tag{8}$$

Note that for any $l > 0$,

$$\begin{aligned}
\mathrm{Var}\left[\mathcal{Z}(M^{k_l}) - \mathcal{Z}(M^{k_{l-1}})\right] &\leq \mathrm{E}\left[\left(\mathcal{Z}(M^{k_l}) - \mathcal{Z}(M^{k_{l-1}})\right)^2\right] \\
&\leq 2\,\mathrm{E}\left[\left(\mathcal{Z}(M^{k_l}) - \mathcal{Z}(M)\right)^2\right] \\
&\quad + 2\,\mathrm{E}\left[\left(\mathcal{Z}(M^{k_{l-1}}) - \mathcal{Z}(M)\right)^2\right] \\
&\leq 2(Bk_l^{-\beta} + Bk_{l-1}^{-\beta}) \leq 4Bk_l^{-\beta} \leq \widetilde{B}k_l^{-\beta},
\end{aligned}$$

by Theorem 5. For notational convenience we assume that $\widetilde{B}$ is such that $\mathrm{Var}[\mathcal{Z}(M^{k_0})] \leq \widetilde{B}k_0^{-\beta}$.

The following complexity theorem is in some sense similar to the complexity theorem in Giles (2008), although its context is entirely different (duality bias due to number of inner simulations versus European option bias due to SDE discretisation).

Theorem 8 *(complexity theorem) Suppose that $k_l = k_0\kappa^l$ for some integer $k_0, \kappa > 1$, and $l = 0, \dots, L$. Assume (AA) and suppose that $\gamma = \min(\alpha, \beta/2) \geq 1/2$. Fix some $0 < \varepsilon < 1$ and set*

$$L = \left\lceil \frac{-\ln \frac{k_0^\gamma \varepsilon}{C\sqrt{2}}}{\gamma \ln \kappa} \right\rceil. \tag{9}$$

With $\lceil x \rceil$ denoting the first integer which is larger than or equal to x, let

$$n_l = \begin{cases} \left\lceil 2\varepsilon^{-2}\widetilde{B}k_0^{-\beta}\kappa^{L(1-\beta)/2}(1-\kappa^{-(1-\beta)/2})^{-1}\kappa^{-l(1+\beta)/2}\right\rceil, & \beta < 1, \\ \left\lceil 2\varepsilon^{-2}\widetilde{B}(L+1)k_0^{-1}\kappa^{-l}\right\rceil, & \beta = 1, \\ \left\lceil 2\varepsilon^{-2}\widetilde{B}k_0^{-\beta}(1-\kappa^{-(\beta-1)/2})^{-1}\kappa^{-l(1+\beta)/2}\right\rceil, & \beta > 1. \end{cases}$$

Then

$$\sqrt{\mathrm{E}[Y^{\mathbf{n},\mathbf{k}} - Y(M)]^2} \le \varepsilon, \tag{10}$$

while the computational complexity of the estimator $Y^{\mathbf{n},\mathbf{k}}$ is, up to a constant, given by

$$\mathcal{C}^{\mathbf{n},\mathbf{k}}(\varepsilon) := \sum_{l=0}^{L} k_l n_l = \begin{cases} O(\varepsilon^{-2-(1-\beta)/\gamma}), & \beta < 1, \\ O(\varepsilon^{-2}\ln^2\varepsilon), & \beta = 1, \\ O(\varepsilon^{-2}), & \beta > 1. \end{cases}$$

Proof. Due to (7) and (9), we have in any case

$$\left|\mathrm{E}\left[Y^{\mathbf{n},\mathbf{k}}\right] - Y(M)\right| \le C k_0^{-\gamma}\kappa^{-L\gamma} = \varepsilon/\sqrt{2}. \tag{11}$$

i) Case $\beta < 1$: We have by (8),

$$\begin{aligned}
\mathrm{Var}\left[Y^{\mathbf{n},\mathbf{k}}\right] &\le \widetilde{B}\sum_{l=0}^{L}\widetilde{B}^{-1}2^{-1}\varepsilon^2 k_0^{\beta}\kappa^{-L(1-\beta)/2}(1-\kappa^{-(1-\beta)/2})\kappa^{l(1+\beta)/2}k_0^{-\beta}\kappa^{-\beta l} \\
&= 2^{-1}\sum_{l=0}^{L}\varepsilon^2\kappa^{-L(1-\beta)/2}(1-\kappa^{-(1-\beta)/2})\frac{\kappa^{(L+1)(1-\beta)/2}-1}{\kappa^{(1-\beta)/2}-1}\frac{\kappa^{-(1-\beta)/2}}{\kappa^{-(1-\beta)/2}} \\
&= 2^{-1}\sum_{l=0}^{L}\varepsilon^2\kappa^{-L(1-\beta)/2}\left(\kappa^{(L+1)(1-\beta)/2}-1\right)\kappa^{-(1-\beta)/2} \\
&= 2^{-1}\varepsilon^2\left(1-\kappa^{-(L+1)(1-\beta)/2}\right) \le \varepsilon^2/2,
\end{aligned}$$

and (10) straightforwardly follows from (11). The order of $\mathcal{C}^{\mathbf{n},\mathbf{k}}(\varepsilon)$ follows from the estimate

$$\begin{aligned}
\sum_{l=0}^{L} k_l n_l &\le \sum_{l=0}^{L} k_0\kappa^l\left(2\varepsilon^{-2}\widetilde{B}k_0^{-\beta}\kappa^{L(1-\beta)/2}(1-\kappa^{-(1-\beta)/2})^{-1}\kappa^{-l(1+\beta)/2}+1\right) \\
&= \sum_{l=0}^{L}\kappa^l\left(2\varepsilon^{-2}\widetilde{B}k_0^{1-\beta}\kappa^{L(1-\beta)/2}(1-\kappa^{-(1-\beta)/2})^{-1}\kappa^{-l(1+\beta)/2}+k_0\right) \\
&= k_0^{1-\beta}\left(2\varepsilon^{-2}\widetilde{B}\kappa^{L(1-\beta)/2}\left(\kappa^{L(1-\beta)/2}-\kappa^{-(1-\beta)/2}\right)+k_0\frac{\kappa^{L+1}-1}{\kappa-1}\right) \\
&\le 2\varepsilon^{-2}\widetilde{B}k_0^{1-\beta}\kappa^{L(1-\beta)}+k_0\kappa^{L+1} \\
&\le 2\varepsilon^{-2}\widetilde{B}k_0^{1-\beta}\kappa^{\left(\frac{-\ln\frac{k_0^\gamma\varepsilon}{C\sqrt{2}}}{\gamma\ln\kappa}+1\right)(1-\beta)} + k_0\kappa^{\frac{-\ln\frac{k_0^\gamma\varepsilon}{C\sqrt{2}}}{\gamma\ln\kappa}+2} \\
&= 2\widetilde{B}\frac{\left(C\sqrt{2}\right)^{(1-\beta)/\gamma}}{\varepsilon^{2+(1-\beta)/\gamma}}\kappa^{(1-\beta)} + \frac{\left(C\sqrt{2}\right)^{1/\gamma}}{\varepsilon^{1/\gamma}}\kappa^2 \\
&= O(\varepsilon^{-2-(1-\beta)/\gamma}), \quad \varepsilon\to 0
\end{aligned}$$

(note that $\gamma \ge 1/2$).

ii) Case $\beta = 1$: By a straightforward calculation, we obtain again

$$\text{Var}\left[Y^{\mathbf{n},\mathbf{k}}\right] \leq \varepsilon^2/2,$$

as well as (10) via (11). For $\mathcal{C}^{\mathbf{n},\mathbf{k}}(\varepsilon)$ we now have

$$\begin{aligned}
\mathcal{C}^{\mathbf{n},\mathbf{k}}(\varepsilon) &\leq \sum_{l=0}^{L} k_0\kappa^l \left(2\varepsilon^{-2}\widetilde{B}(L+1)k_0^{-1}\kappa^{-l}+1\right) \\
&= \sum_{l=0}^{L} \left(2\varepsilon^{-2}\widetilde{B}(L+1)+k_0\kappa^l\right) \\
&= 2\varepsilon^{-2}\widetilde{B}(L+1)^2 + k_0\frac{\kappa^{L+1}-1}{\kappa-1} \\
&\leq 2\varepsilon^{-2}\widetilde{B}(L+1)^2 + k_0\kappa^{L+1} \\
&\leq 2\varepsilon^{-2}\widetilde{B}\left(\frac{-\ln\frac{k_0^\gamma\varepsilon}{C\sqrt{2}}}{\gamma\ln\kappa}+2\right)^2 + \frac{\left(C\sqrt{2}\right)^{1/\gamma}}{\varepsilon^{1/\gamma}}\kappa^2 \\
&= O(\varepsilon^{-2}\ln^2\varepsilon), \quad \varepsilon\to 0
\end{aligned}$$

since $\gamma \geq 1/2$.

iii) Case $\beta > 1$: For the variance we have

$$\begin{aligned}
\text{Var}\left[Y^{\mathbf{n},\mathbf{k}}\right] &\leq \widetilde{B}\sum_{l=0}^{L} 2^{-1}\varepsilon^2\widetilde{B}^{-1}k_0^\beta(1-\kappa^{-(\beta-1)/2})\kappa^{l(1-\beta)/2}k_0^{-\beta} \\
&= 2^{-1}\varepsilon^2(1-\kappa^{-(\beta-1)/2})\frac{1-\kappa^{(L+1)(1-\beta)/2}}{1-\kappa^{(1-\beta)/2}} \\
&= 2^{-1}\varepsilon^2\left(1-\kappa^{(L+1)(1-\beta)/2}\right) \leq \varepsilon^2/2
\end{aligned}$$

and (10) holds in view of (11). Finally, we derive

$$\begin{aligned}
\mathcal{C}^{\mathbf{n},\mathbf{k}}(\varepsilon) &\leq \sum_{l=0}^{L} k_0\kappa^l \left(2\varepsilon^{-2}\widetilde{B}k_0^{-\beta}(1-\kappa^{-(\beta-1)/2})^{-1}\kappa^{-l(1+\beta)/2}+1\right) \\
&\leq 2\varepsilon^{-2}\widetilde{B}k_0^{1-\beta}\sum_{l=0}^{L}(1-\kappa^{-(\beta-1)/2})^{-1}\kappa^{l(1-\beta)/2} + k_0\kappa^{L+1} \\
&= 2\varepsilon^{-2}\widetilde{B}k_0^{1-\beta}\left(1-\kappa^{(L+1)(1-\beta)/2}\right) + k_0\kappa^{L+1} \\
&\leq 2\varepsilon^{-2}\widetilde{B}k_0^{1-\beta} + \frac{\left(C\sqrt{2}\right)^{1/\gamma}}{\varepsilon^{1/\gamma}}\kappa^2 \\
&= O(\varepsilon^{-2}\ln^2\varepsilon), \quad \varepsilon\to 0,
\end{aligned}$$

since $\gamma \geq 1/2$. ■

6 Numerical example: Bermudan max-call

This is a benchmark example studied in Glasserman (2003), Haugh and Kogan (2004) and Rogers (2002) among others. Specifically, a model with D identical distributed assets is considered where each underlying has dividend yield δ. The risk-neutral dynamics of the assets are given by

$$dX_t^d = (r-\delta)X_t^d dt + \sigma X_t^d dW_t^d, \quad d = 1, ..., D, \tag{12}$$

where W_t^d, $d = 1, ..., D$, are independent one dimensional Brownian motions and r, δ, σ are constants. As in the mentioned literature we take $r = 0.05$, $\delta = 0.1$, and $\sigma = 0.2$. At any time $t \in \{T_0, ..., T_{\mathcal{J}}\}$ the holder of the option may exercise it and receive the payoff

$$Z_j = h(X_{T_j}) = (\max(X_{T_j}^1, ..., X_{T_j}^D) - \kappa)^+.$$

We consider an example when $T_j = jT/\mathcal{J}$, $j = 0, ..., \mathcal{J}$, with $T = 3$ and $\mathcal{J} = 9$. In our implementation study we first construct a family of stopping rules $\tau_j : \Omega \to \{j, \dots, \mathcal{J}\}$ by the Longstaff-Schwartz algorithm. This basically boils down to choosing a basis $(\phi_k(t,x)$, $k = 1, \dots, K)$ and estimating vectors of regression coefficients $(\alpha_l \in \mathbb{R}^K$, $l = 0, \dots, \mathcal{J})$. Once the coefficients $\{\alpha_l\}$ are estimated, we can define

$$\tau_j := \min\{j \le l \le \mathcal{J} : \alpha_l^\top \phi(T_l, X_{T_l}) \le Z_l\}$$

and

$$Y_j := \mathrm{E}_{\mathcal{F}_j}[Z_{\tau_j}], \quad j = 1, \dots, \mathcal{J}.$$

In our example with $D = 2$ the following $K = 4$ basis functions are used:

$$\psi_1(t,x) = 1, \quad \psi_2(t,x) = x^1, \quad \psi_3(t,x) = x^2, \quad \psi_4(t,x) = h(x).$$

We stress that stopping rules $\{\tau_j\}$ are estimated only once and remain fixed thereafter. The target martingale M is thus defined by $M_0 = 0$,

$$M_j = M_{j-1} + \mathrm{E}_{\mathcal{F}_j}[Z_{\tau_j}] - \mathrm{E}_{\mathcal{F}_{j-1}}[Z_{\tau_i}], \quad j = 1, \dots, \mathcal{J}, \tag{13}$$

hence, following the description of the Andersen and Broadie (2004) algorithm in Glasserman (2003), we have

$$\begin{aligned} M_j &= \sum_{p=1}^{j} \left(\mathrm{E}_{\mathcal{F}_p}[Z_{\tau_p}] - \mathrm{E}_{\mathcal{F}_{p-1}}[Z_{\tau_p}]\right) \\ &= \sum_{p=1}^{j} \left\{1_{\tau_p = p}\left(Z_p - \mathrm{E}_{\mathcal{F}_{p-1}}[Z_{\tau_p}]\right) + 1_{\tau_p > p} 1_{p < \mathcal{J}}\left(\mathrm{E}_{\mathcal{F}_p}[Z_{\tau_{p+1}}] - \mathrm{E}_{\mathcal{F}_{p-1}}[Z_{\tau_p}]\right)\right\}. \end{aligned} \tag{14}$$

In order to approximate the target martingale M corresponding to (an outer) trajectory X we construct a sequence of martingales M^k, where each M^k is a martingale constructed via (14) by replacing the one-step conditional expectations $\mathrm{E}_{\mathcal{F}_{p-1}}[Z_{\tau_p}]$, $p = 1, \dots, \mathcal{J}$, with their Monte Carlo approximations based on k (inner) trajectories. Fix some $k_0, n_0, L \in \mathbb{N}$, set $\kappa = 2$, and define

$$\begin{aligned} n_l &= n_0 \cdot \kappa^{-l}, \\ k_l &= k_0 \cdot \kappa^l \end{aligned}$$

for any $l \in \mathbb{N}$. Then the numerical complexity of the multi-level estimate $Y^{\mathbf{n},\mathbf{k}}$ is given, up to a constant, by

$$\mathcal{C}_{ML} = n_0 k_0 + \sum_{l=1}^{L} n_l (k_l + k_{l-1}) = n_0 k_0 \left(1 + L(1+\kappa^{-1})\right).$$

Let us compute now the variance of $Y^{\mathbf{n},\mathbf{k}}$. We have

$$\begin{aligned}\operatorname{Var}[Y^{\mathbf{n},\mathbf{k}}] &= n_0^{-1} \operatorname{Var}[\mathcal{Z}(M^{k_0})] + n_l^{-1} \sum_{l=1}^{L} \operatorname{Var}(\mathcal{Z}(M^{k_l}) - \mathcal{Z}(M^{k_{l-1}})) \\ &= n_0^{-1}\sigma^2(k_0) + n_0^{-1}k_0^{-1} \sum_{l=1}^{L} \mathfrak{V}_l,\end{aligned}$$

where $\sigma^2(k) := \operatorname{Var}[\mathcal{Z}(M^k)]$, $k \in \mathbb{N}$ and $\mathfrak{V}_l := k_l \cdot \operatorname{Var}(\mathcal{Z}(M^{k_l}) - \mathcal{Z}(M^{k_{l-1}}))$, $l = 1, \ldots, L$. Based on 100000 paths of the process $\mathcal{Z}(M^{k_l}) - \mathcal{Z}(M^{k_{l-1}})$, $l = 1, \ldots, L$, for different values k_0, we found $\mathfrak{V}_l$ to be bounded by $\mathfrak{V}_\infty = 1000$ for all $l = 1, \ldots, L$, where $\mathfrak{V}_\infty$ does not depend on k_0, but only on κ. Hence

$$\operatorname{Var}[Y^{\mathbf{n},\mathbf{k}}] \le n_0^{-1}\sigma^2(k_0) + n_0^{-1}k_0^{-1} L \mathfrak{V}_\infty.$$

Fix some $n > 0$, then the computational costs of the Andersen-Broadie algorithm based on n outer simulations and k inner simulations are proportional to

$$\mathcal{C}_{AB} = nk.$$

The variance of the estimate $Y^{n,k}$ is given by

$$\operatorname{Var}(Y^{n,k}) = n^{-1} \operatorname{Var}(\mathcal{Z}(M^k)) = n^{-1}\sigma^2(k).$$

We have estimated $\sigma^2(k)$ using 10^6 replications of the r.v. $\mathcal{Z}(M^k)$ for $k = 1, \ldots, 100$. As one can see from Figure 1 (right), $\sigma(k) \to \sigma_\infty \approx 5.2$ as $k \to \infty$. The bias of $Y^{n,k}$, can be written as

$$\mathcal{B}(k) := Y^{\infty,k} - Y^{\infty,\infty}, \quad k \in \mathbb{N}.$$

In Figure 1(left) the values $\ln \mathcal{B}(k)$ are plotted on the log-scale, where in order to estimate the limit values $Y^{\infty,k}$ and $Y^{\infty,\infty} \approx 8.3231$, we use 10^6 outer paths and 1000 inner paths. By fitting a straight line to the data $(\ln(k), \ln(\mathcal{B}(k)))$, we obtain

$$0 < \mathcal{B}(k) < \mu_\infty k^{-1}, \quad k > 10,$$

with $\mu_\infty = \exp(2.5)$. Note that the multi-level estimate $Y^{\mathbf{n},\mathbf{k}}$ has, by construction, the same bias as Y^{n_L,k_L}. Now we choose L, n_0, k_0 and n, k in such a way that the overall accuracy of both AB and ML dual estimates measured by $\sqrt{\operatorname{Var}[Y^{n,k}] + \mathcal{B}^2(k)}$ and $\sqrt{\operatorname{Var}[Y^{\mathbf{n},\mathbf{k}}] + \mathcal{B}^2(k_L)}$, respectively, is bounded by ε. In the case of the AB dual estimate, it is enough to require

$$\left(\frac{\mu_\infty}{k}\right)^2 = n^{-1}\sigma^2(k) = \frac{\varepsilon^2}{2}$$

which leads to

$$n = \frac{2\sigma_\infty^2}{\varepsilon^2}, \quad k = \frac{\sqrt{2}\mu_\infty}{\varepsilon}$$

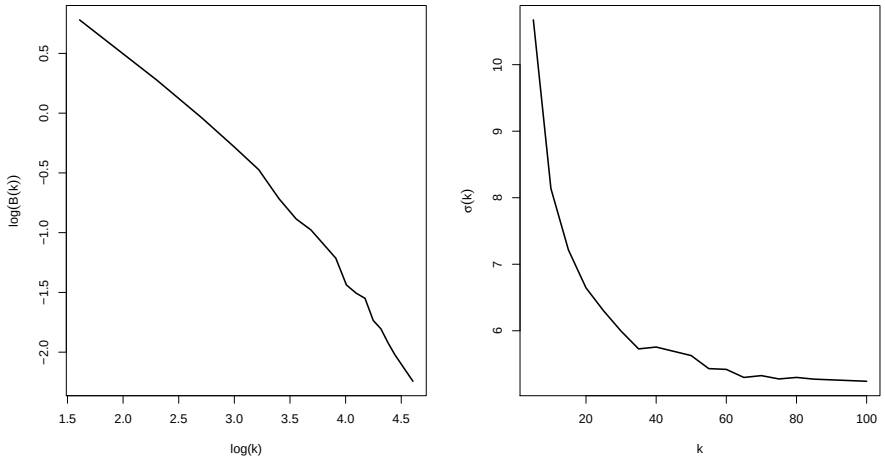

Figure 1: The functions $\ln \mathcal{B}(k)$ and $\sigma(k)$.

for k large enough. For the ML algorithm we assume that

$$n_0^{-1}\sigma^2(k_0) = n_0^{-1}k_0^{-1}L\mathfrak{V}_\infty = \frac{\varepsilon^2}{4}, \tag{15}$$

$$\frac{\mu_\infty}{k_L} = \frac{\varepsilon}{\sqrt{2}}. \tag{16}$$

From (15) we have

$$n_0 = 4\varepsilon^{-2}\sigma^2(k_0), \quad k_0 = L\mathfrak{V}_\infty\sigma^{-2}(k_0)$$

and, due to (16), the parameter L is to be defined as (approximate) solution of the equation

$$L = \ln\left(\frac{\sqrt{2}\mu_\infty}{\varepsilon k_0}\right)\ln^{-1}\kappa = \ln\left(\frac{\sqrt{2}\mu_\infty\sigma^2(k_0)}{\varepsilon L\mathfrak{V}_\infty}\right)\ln^{-1}\kappa.$$

Hence we solve

$$L = \ln\left(\frac{\sqrt{2}\theta}{\varepsilon L}\right)\ln^{-1}\kappa \tag{17}$$

with $\theta = \mu_\infty\sigma_\infty^2/\mathfrak{V}_\infty$, provided k_0 is large enough. We thus have asymptotically,

$$L = L(\varepsilon) \asymp \ln^{-1}\kappa \ln\left(\frac{\sqrt{2}\theta}{\varepsilon}\right), \quad \varepsilon \to 0.$$

Let us finally compare the complexities of AB and ML algorithms for n_0, k_0, n, k, large enough. It holds

$$\begin{aligned}\mathcal{R}(\varepsilon) = \frac{\mathcal{C}_{AB}(\varepsilon)}{\mathcal{C}_{ML}(\varepsilon)} &= \frac{nk}{n_0k_0\left(1 + L(1+\kappa^{-1})\right)} \\ &= \frac{\theta/\sqrt{2}}{L\left(1 + L(1+\kappa^{-1})\right)\varepsilon},\end{aligned}$$

where L solves (17). As can be seen from the above formulas, the quantity θ has a big impact on $\mathcal{R}$. In our example we have $\theta \approx 0.3$. The resulting functions L and $\mathcal{R}$ are reported in Figure 2.

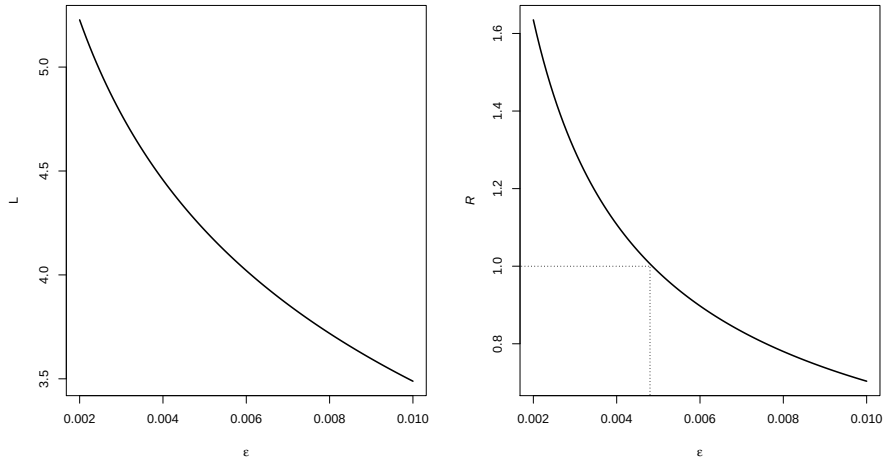

Figure 2: The values L and R as functions of the precision ε.

By examining Figure 2 we conclude that (for this example) the use of the ML algorithm with $L > 0$ is advantageous as soon as $\varepsilon < 5 * 10^{-3}$, i.e., in situations where relatively high precision is needed. The efficiency of ML dual approach can be further enhanced by using the same set of k_l inner trajectories to construct both $M_j^{k_{l-1},(i)}$ and $M_j^{k_l,(i)}$. This modification leads to much smaller values of $\mathfrak{V}_\infty$ ($\mathfrak{V}_\infty = 350$ and consequently $\theta = 0.87$ in our example). The resulting functions $L(\varepsilon)$ and $\mathcal{R}(\varepsilon)$ are depicted in Figure 3.

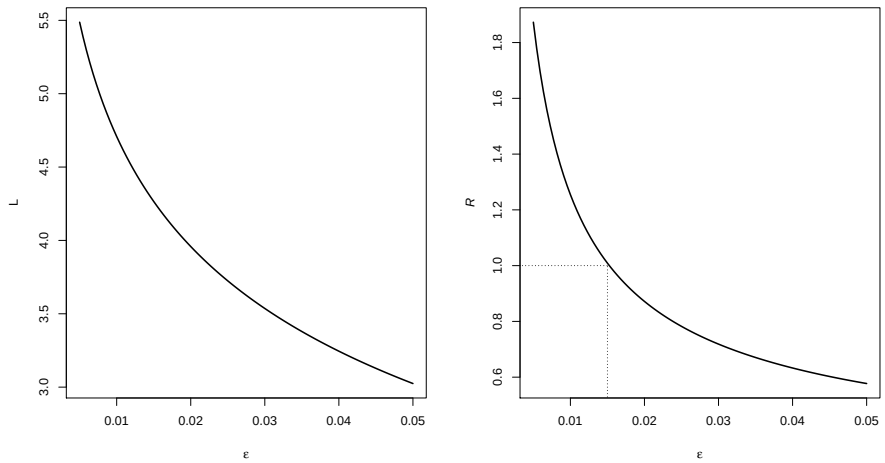

Figure 3: The values L and R as functions of the precision ε.

Hence the modified version of the ML algorithm can be recommended already if $\varepsilon < 0.015$.

Remark 9 *The max-call example studied above is somehow advantageous for the standard nested dual algorithm because of the inequality* $\mathcal{B}(k) \leq \mu_\infty k^{-1}$. *In a more general situation where the assumptions (AL) and/or (AQ) do not hold, we have at least* $\mathcal{B}(k) \leq \mu_\infty k^{-1/2}$ *leading to*

$$\mathcal{R}(\varepsilon) = \frac{\theta}{L\left(1 + L(1+\kappa^{-1})\right)\varepsilon^2}$$

with $\theta = \mu_\infty^2 \sigma_\infty^2 / \mathfrak{V}_\infty$ *and* L *solving*

$$L = \ln\left(\frac{2\theta}{\varepsilon^2 L}\right) \ln^{-1}(\kappa).$$

As a result, the multilevel dual approach may become superior to the standard dual approach for the values of ε *much larger than* 0.015. *For example, if* $\theta = 0.87$ *as before, we get that* $\mathcal{R}(\varepsilon) > 1$ *for all* $\varepsilon \leq 0.18$.

Concluding remarks

We have presented a multilevel version of the dual approach for valuation of American derivatives. Unlike the multilevel method by Giles (2008) who considers, in the context of pricing European options, different levels of time discretisation for the numerical SDE solution of the underlying process, we consider different levels of approximation to a target martingale in the context of dual pricing of American options. In this respect we underline that, in this paper, the trajectories of the underlying process are assumed to be simulated *exactly*. This assumption is made partially to keep the presentation and the message of the paper as clear as possible. From a practical point of view, by taking the time discretisation step small enough, the additional error due to an approximate solution of the underlying SDE, can be kept much smaller than the typically required Monte Carlo accuracy. Anyway, incorporation and analysis of the time discretisation error in the multilevel approach to dual pricing of American options will be an interesting subject for a subsequent paper.

References

L. Andersen and M. Broadie (2004). A primal-dual simulation algorithm for pricing multidimensional American options. *Management Sciences*, **50**(9), 1222-1234.

V. Bally, G. Pages (2003). A quantization algorithm for solving multidimensional discrete optimal stopping problem. Bernoulli, **9**(6), 1003-1049.

D. Belomestny (2011). Pricing Bermudan options using regression: optimal rates of convergence for lower estimates, to appear in *Finance and Stochastics*.

D. Belomestny (2011). Solving optimal stopping problems by empirical dual optimization and penalization. Available at http://www.uni-due.de/~hm0124/publications.php?cat=10 .

D. Belomestny, C. Bender and J. Schoenmakers (2009). True upper bounds for Bermudan products via non-nested Monte Carlo. *Math. Financ.*, **19**(1), 53–71.

C. Bender, A. Kolodko, J. Schoenmakers (2008). Enhanced policy iteration for American options via scenario selection. Quant. Finance **8**, No. 2, 135–146.

M. Broadie, P. Glasserman (2004). A stochastic mesh method for pricing high-dimensional American options. Journal of Computational Finance, **7**(4), 35–72.

J. Carriere (1996). Valuation of early-exercise price of options using simulations and nonparametric regression. Insurance: Mathematics and Economics, **19**, 19–30.

N. Chen and P. Glasserman (2007). Additive and Multiplicative Duals for American Option Pricing. Finance and Stochastics, **11**, 153–179.

M.H.A. Davis, I. Karatzas (1994). A Deterministic Approach to Optimal Stopping. In: Probability, Statistics and Optimisation (ed. F.P. Kelly). NewYork Chichester: John Wiley & Sons Ltd., 455–466.

V. Desai, V. Farias and C. Moallemi (2010). Pathwise optimization for optimal stopping problems. Preprint.

M.B. Giles (2008). Multilevel Monte Carlo path simulation. Operations Research 56(3):607-617.

P. Glasserman (2003). Monte Carlo Methods in Financial Engineering. Springer.

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

M. Kohler, A. Krzyzak und H. Walk (2010). Upper bounds for Bermudan options on Markovian data using nonparametric regression and a reduced number of nested Monte Carlo steps. *Statistics and Decision*, **26**, 275–288.

A. Kolodko, J. Schoenmakers (2004). Upper bounds for Bermudan style derivatives. Monte Carlo Methods and Appl., **10**, No. 3–4, 331-343.

A. Kolodko, J. Schoenmakers (2006). Iterative construction of the optimal Bermudan stopping time. Finance and Stochastics, **10**, 27–49.

F.A. Longstaff, E.S. Schwartz (2001). Valuing American options by simulation: a simple least-squares approach. Review of Financial Studies, **14**, 113–147.

L.C.G. Rogers (2002). Monte Carlo valuation of American options. *Mathematical Finance*, **12**, 271-286.

J. Schoenmakers, J. Huang (2011). Optimal dual martingales, their analysis and application to new algorithms for Bermudan products. Available at SSRN: http://ssrn.com/abstract=1825944, initial version WIAS Preprint 1574 (2010).

J. Tsitsiklis, B. Van Roy (2001). Regression methods for pricing complex American style options. IEEE Trans. Neural. Net., **12**, 694–703.