PRICING BERMUDAN OPTIONS VIA MULTI-LEVEL APPROXIMATION METHODS [1]

Denis Belomestny, Fabian Dickmann and Tigran Nagapetyan
Duisburg-Essen University and Fraunhofer ITWM

In this article we propose a novel approach to reduce the computational complexity of various approximation methods for pricing discrete time American or Bermudan options. Given a sequence of continuation values estimates corresponding to different levels of spatial approximation, we propose a multi-level low biased estimate for the price of the option. It turns out that the resulting complexity gain can be of order ε^{-1} with ε denoting the desired precision. The performance of the proposed multilevel algorithms is illustrated by a numerical example.

1 INTRODUCTION

Pricing of an American option usually reduces to solving an optimal stopping problem that can be efficiently solved in low dimensions via dynamic programming algorithm. However, many problems arising in practice (see e.g. Glasserman (2004)) have high dimensions, and these applications have motivated the development of Monte Carlo methods for pricing American option. Pricing American style derivatives via Monte Carlo is a challenging task, because it requires the backwards dynamic programming algorithm that seems to be incompatible with the forward structure of Monte Carlo methods. In recent years much research was focused on the development of fast methods to compute approximations to the optimal exercise policy. Eminent examples include the functional optimization approach of Andersen (2000), the mesh method of Broadie and Glasserman (1997), the regression-based approaches of Carriere (1996), Longstaff and Schwartz (2001), Tsitsiklis and Van Roy (1999), Egloff (2005) and Belomestny (2011). The complexity of the fast approximations algorithms depends on the desired precision ε in a quite nonlinear way that, in turn, is determined by some fine properties of the underlying exercise boundary and the continuation values (see, e.g., Belomestny (2011)). In some situations (e.g. in the case of the stochastic mesh method or local regression) this complexity is of order ε^{-3}, which is rather high. One way to reduce the complexity of the fast approximation methods is to use various variance reduction methods. However, the latter methods are often ad hoc and, more importantly, do not lead to provably reduced asymptotic complexity. In this paper we propose a generic approach which is able to reduce the order of asymptotic complexity

[1] This research was partially supported by the Deutsche Forschungsgemeinschaft through the SPP 1324 "Mathematical methods for extracting quantifiable information from complex systems" and by Laboratory for Structural Methods of Data Analysis in Predictive Modeling, MIPT, RF government grant, ag. 11.G34.31.0073.

and which is applicable to various fast approximation methods, such as global regression, local regression or stochastic mesh method. The main idea of the method is inspired by the pathbreaking work of Giles (2008) that introduced a multilevel idea into stochastics. As similar to the recent work of Belomestny et al (2012), we consider not levels corresponding to different discretization steps, but levels related to different degrees of approximation of the continuation values. For example, in the case of the Longstaff-Schwartz algorithm, the latter degree is basically governed by the number of basis functions and in the case of the mesh method by the number of training paths used to approximate the continuation values. The new multi-level approach is able to significantly reduce the complexity of the fast approximation methods leading in some cases to the complexity gain of order ε^{-1}. The paper is organised as follows. In Section 2 the pricing problem is formulated, the main assumptions are introduced and illustrated. In Section 3 the complexity analysis of a generic approximation algorithm is carried out. The main multi-level Monte Carlo algorithm is introduced in Section 4 where also its complexity is studied. In Section 5 we numerically test our approach for the problem of pricing Bermudan max-call options via mesh method. The proofs are collected in Section 6.

2 Main setup

An American option grants the holder the right to select the time at which to exercise the option, and in this differs from a European option that may be exercised only at a fixed date. A general class of American option pricing problems can be formulated through an $\mathbb{R}^d$ Markov process $\{X_t,\, 0 \le t \le T\}$ defined on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{0\le t\le T}, \mathrm{P})$. It is assumed that the process (X_t) is adapted to $(\mathscr{F}_t)_{0\le t\le T}$ in the sense that each X_t is $\mathscr{F}_t$ measurable. Recall that each $\mathscr{F}_t$ is a σ -algebra of subsets of Ω such that $\mathscr{F}_s \subseteq \mathscr{F}_t \subseteq \mathscr{F}$ for $s \le t$. We restrict attention to options admitting a finite set of exercise opportunities $0 = t_0 < t_1 < t_2 < \ldots < t_{\mathscr{J}} = T$, called Bermudan options. Then

$$Z_j := X_{t_j}, \quad j = 0, \ldots, \mathscr{J},$$

is a Markov chain. If exercised at time t_j, $j = 1, \ldots, \mathscr{J}$, the option pays $g_j(Z_j)$, for some known functions $g_0, g_1, \ldots, g_{\mathscr{J}}$ mapping $\mathbb{R}^d$ into $[0, \infty)$. Let $\mathscr{T}_j$ denote the set of stopping times taking values in $\{j, j+1, \ldots, \mathscr{J}\}$. A standard result in the theory of contingent claims states that the equilibrium price $V_j(z)$ of the Bermudan option at time t_j in state z given that the option was not exercised prior to t_j is its value under an optimal exercise policy:

$$V_j^*(z) = \sup_{\tau \in \mathscr{T}_j} \mathrm{E}[g_\tau(Z_\tau) | Z_j = z], \quad z \in \mathbb{R}^d.$$

A common feature of all fast approximation algorithms is that they deliver estimates $C_{k,0}(z), \ldots, C_{k,\mathscr{J}-1}(z)$ for the so called continuation values:

$$C_j^*(z) := \mathrm{E}[V_{j+1}^*(Z_{j+1}) | Z_j = z], \quad j = 0, \ldots, \mathscr{J} - 1, \tag{2.1}$$

based on the set of trajectories $(Z_0^{(i)}, \ldots, Z_{\mathscr{J}}^{(i)})$, $i = 1, \ldots, k$, all starting from one point, i.e., $Z_0^{(1)} = \ldots = Z_0^{(k)}$. In the case of the so-called regression methods and the mesh method, the estimates for the continuation values are obtained via the recursion (*dynamic programming principle*):

$$\begin{aligned} C_{\mathscr{J}}^*(z) &= 0, \\ C_j^*(z) &= \mathrm{E}[\max(g_{j+1}(Z_{j+1}), C_{j+1}^*(Z_{j+1}))|Z_j = z] \end{aligned}$$

combined with Monte Carlo: at $(\mathscr{J} - j)$ th step one estimates the expectation

$$\mathrm{E}[\max(g_{j+1}(Z_{j+1}), C_{k,j+1}(Z_{j+1}))|Z_j = z] \tag{2.2}$$

via regression (global or local) based on the set of paths

$$(Z_j^{(i)}, C_{k,j+1}(Z_{j+1}^{(i)})), \quad i = 1, \ldots, k,$$

where $C_{k,j+1}(z)$ is the estimate for $C_{j+1}^*(z)$ obtained in the previous step.

Based on the estimates $C_{k,0}(z), \ldots, C_{k,\mathscr{J}-1}(z)$ we can construct a lower bound (low biased estimate) for V_0^* using the (generally suboptimal) stopping rule:

$$\tau_k = \min\{0 \le j \le \mathscr{J} : g_j(Z_j) \ge C_{k,j}(Z_j)\}$$

with $C_{k,\mathscr{J}} \equiv 0$ by definition. Fix now a natural number n and simulate n new independent trajectories of the process Z. A low-biased estimate for V_0^* can be then defined as

$$V_0^{n,k} = \frac{1}{n} \sum_{r=1}^{n} g_{\tau_k^{(r)}}(Z_{\tau_k^{(r)}}^{(r)}) \tag{2.3}$$

with

$$\tau_k^{(r)} = \inf\{0 \le j \le \mathscr{J} : g_j(Z_j^{(r)}) \ge C_{k,j}(Z_j^{(r)})\}.$$

Thus any fast approximation approximation algorithm can be viewed as consisting of the following two steps.

Step 1 Construction of the estimates $C_{k,j}$, $j = 1, \ldots, J$, on k training paths.

Step 2 Construction of the low-biased estimate $V_0^{n,k}$ by evaluating functions $C_{k,j}$, $j = 1, \ldots, J$, on each of new n testing trajectories.

Let us now consider a generic family of the continuation values estimates $C_{k,0}(z), \ldots, C_{k,\mathscr{J}-1}(z)$ with the natural number k determining the quality of the estimates as well as their complexity. In particular we make the following assumptions.

(AP) For any $k \in \mathbb{N}$ the estimates $C_{k,0}(z), \ldots, C_{k,\mathscr{J}-1}(z)$ are defined on some probability space $(\Omega^k, \mathscr{F}^k, \mathrm{P}^k)$ which is independent of $(\Omega, \mathscr{F}, \mathrm{P})$.

(AC) For any $j = 1, \dots, \mathscr{J}$, the cost of constructing the estimate $C_{k,j}$ on k training paths, i.e., $C_{k,j}\big(Z_j^{(i)}\big)$, $i = 1, \dots, k$, is of order $k \times k^{\varkappa_1}$ for some $\varkappa_1 > 0$ and the cost of evaluating $C_{k,j}(z)$ in a new point $z \notin \{Z_j^{(1)}, \dots, Z_j^{(k)}\}$ is of order $k^{\varkappa_2}$ for some $\varkappa_2 > 0$.

(AQ) There is a sequence of positive real numbers γ_k with $\gamma_k \to 0$, $k \to \infty$ such that

$$\mathrm{P}^k\left(\sup_z \left|C_{k,j}(z) - C_j^*(z)\right| > \eta\sqrt{\gamma_k}\right) < B_1 e^{-B_2\eta}, \quad \eta > 0$$

for some constants $B_1 > 0$ and $B_2 > 0$.

Discussion

- Given (AC) the overall complexity of a fast approximation algorithm is proportional to
$$k^{1+\varkappa_1} + n \times k^{\varkappa_2}, \tag{2.4}$$
where the first term in (2.4) represents the cost of constructing the estimates $C_{k,j}$, $j = 1, \dots, J$, on training paths and the second one gives the cost of evaluating the estimated continuation values on n testing paths.

- Additionally, one usually has to take into account the cost of paths simulation. If the process X solves a stochastic differential equation and the Euler discretisation scheme with time step h is used to generate paths, then the term $k \times h^{-1} + n \times h^{-1}$ needs to be added to (2.4). In order to make the analysis more focused and transparent we do not take here the path generation costs into account.

Let us now illustrate the above assumptions for three well known fast approximation methods.

Example 1 (Global regression)**.** Fix a vector of real-valued functions $\psi = (\psi_1, \dots, \psi_M)$ on $\mathbb{R}^d$. Suppose that the estimate $C_{k,j+1}$ is already constructed and has the form

$$C_{k,j+1}(z) = \alpha_{j+1,1}^k \psi_1(z) + \ldots + \alpha_{j+1,M}^k \psi_L(z)$$

for some $(\alpha_{j+1,1}^k, \dots, \alpha_{j+1,M}^k) \in \mathbb{R}^M$. Let $\boldsymbol{\alpha}_j^k = (\alpha_{j,1}^k, \dots, \alpha_{j,M}^k)$ be a solution of the following least squares optimization problem:

$$\operatorname*{arginf}_{\boldsymbol{\alpha} \in \mathbb{R}^M} \sum_{i=1}^k \left[\zeta_{j+1,k}(Z_{j+1}^{(i)}) - \alpha_1 \psi_1(Z_j^{(i)}) - \ldots - \alpha_M \psi_M(Z_j^{(i)})\right]^2 \tag{2.5}$$

with $\zeta_{j+1,k}(z) = \max\left\{g_{j+1}(z), C_{k,j+1}(z)\right\}$, where $C_{k,j+1}$ is the estimate of C_{j+1}^* obtained in the previous step. Define the approximation for C_j^* via

$$C_{k,j}(z) = \alpha_{j,1}^k \psi_1(z) + \ldots + \alpha_{j,M}^k \psi_M(z), \quad z \in \mathbb{R}^d.$$

It is clear that all estimates $C_{k,j}$ are well defined on the cartesian product of k independent copies of $(\Omega, \mathscr{F}, \mathrm{P})$. The complexity $\mathrm{comp}(\boldsymbol{\alpha}_j^k)$ of computing $\boldsymbol{\alpha}_j^k$ is of order $k \cdot M^2 + \mathrm{comp}(\boldsymbol{\alpha}_{j+1}^k)$, since each $\boldsymbol{\alpha}_j^k$ is of the form $\boldsymbol{\alpha}_j^k = B^{-1} b$ with

$$B_{p,q} = \frac{1}{k} \sum_{i=1}^{k} \psi_p(Z_j^{(i)}) \psi_q(Z_j^{(i)})$$

and

$$b_p = \frac{1}{k} \sum_{i=1}^{k} \psi_p(Z_j^{(i)}) \zeta_{k,j+1}(Z_{j+1}^{(i)}),$$

$p, q \in \{1, \ldots, M\}$. Iterating backwardly in time we get $\mathrm{comp}(\boldsymbol{\alpha}_j^k) \sim (\mathscr{J} - j) \cdot k \cdot M^2$. Furthermore, it can be shown that the estimates $C_{k,0}(z), \ldots, C_{k,\mathscr{J}-1}(z)$ satisfy the assumption (AQ) with $\gamma_k = 1/k$, provided M increases with k at polynomial rate, i.e., $M = k^\rho$ for some $\rho > 0$ (see, e.g., Zanger (2013)). Thus, the parameters $\varkappa_1$ and $\varkappa_2$ in (AC) are given by 2ρ and ρ, respectively.

Example 2 (Local regression)**.** Local polynomial regression estimates can be defined as follows. Fix some j such that $0 \leq j < \mathscr{J}$ and suppose that we want to compute the expectation in (2.2):

$$\mathrm{E}[\zeta_{j+1,k}(Z_{j+1}) | Z_j = z], \quad z \in \mathbb{R}^d$$

with $\zeta_{j+1,k}(z) = \max\left\{g_{j+1}(z), C_{k,j+1}(z)\right\}$. For some $\delta > 0$, $z \in \mathbb{R}^d$, an integer $l \geq 0$ and a function $K : \mathbb{R}^d \to \mathbb{R}_+$, denote by $q_{z,k}$ a polynomial on $\mathbb{R}^d$ of degree l (i.e. the maximal order of the multi-index is less than or equal to l) which minimizes

$$\sum_{i=1}^{k} \left[\zeta_{j+1,k}(Z_{j+1}^{(i)}) - q(Z_j^{(i)} - z)\right]^2 K\left(\frac{Z_j^{(i)} - z}{\delta}\right) \tag{2.6}$$

over the set of all polynomials q of degree l. The local polynomial estimator of order l for $C_j^*(z)$ is then defined as $C_{k,j}(z) = q_{z,k}(0)$ if $q_{z,k}$ is the unique minimizer of (2.6) and $C_{k,j}(z) = 0$ otherwise. The value δ is called a bandwidth and the function K is called a kernel function. In Belomestny (2011) it is shown that the local polynomial estimates $C_{k,0}(z), \ldots, C_{k,\mathscr{J}-1}(z)$ of degree l satisfy the assumption (AQ) with $\gamma_k = k^{-2\beta/(2\beta+d)}$ under β-Hölder smoothness of the continuation values $C_0^*(z), \ldots, C_{\mathscr{J}-1}^*(z)$, provided $\delta = k^{-1/(2l+d)}$. Since in general the summation in (2.6) runs over all k paths (see Figure 2.1) we have $\varkappa_1 = 1$ and $\varkappa_2 = 1$ in (AC).

Example 3 (Mesh Method)**.** In the mesh method of Broadie and Glasserman (2004) the continuation value C_j^* at a point z is approximated via

$$C_{k,j}(z) = \frac{1}{k} \sum_{i=1}^{k} \zeta_{k,j+1}(Z_{j+1}^{(i)}) \cdot w_{ij}(z), \tag{2.7}$$

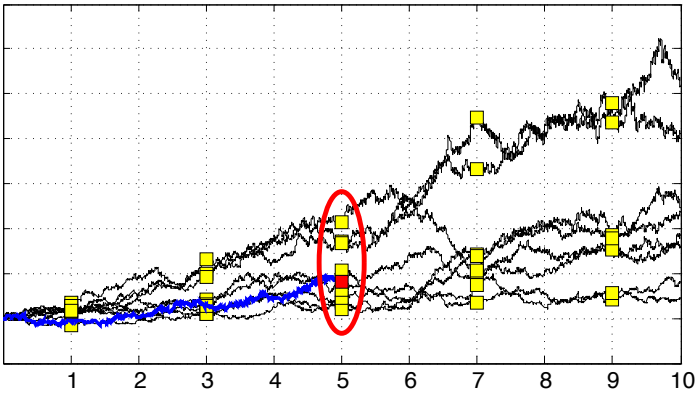

Figure 2.1: Local regression and mesh methods: in order to compute the continuation value estimate $C_{k,5}$ in a point (red) lying on a testing path (blue), all k points (yellow) on training paths at time 5 have to be used.

where $\zeta_{k,j+1}(z) = \max\left\{g_{j+1}(z), C_{k,j+1}(z)\right\}$ and

$$w_{ij}(z) = \frac{p_j(z, Z_{j+1}^{(i)})}{\frac{1}{k}\sum_{l=1}^{k} p_j(Z_j^{(l)}, Z_{j+1}^{(i)})},$$

where $p_j(x,\cdot)$ is the conditional density of Z_{j+1} given $Z_j = x$. Again the summation in (2.7) runs over all k paths. Hence $\varkappa_1 = 1$ in (AC) and for any $j = 0,\dots,\mathscr{J}-1$, the complexity of computing $C_{k,j}(z)$ in a point z not belonging to the set of training trajectories is of order k (see Figure 2.1), provided the transition density $p_j(x,y)$ is analytically known. For assumption (AQ) see, e.g., Agarwal and Juneja (2013).

3 Complexity analysis of $V_0^{n,k}$

We shall use throughout the notation $A \lesssim B$ if A is bounded by a constant multiple of B, independently of the parameters involved, that is, in the Landau notation $A = O(B)$. Equally $A \gtrsim B$ means $B \lesssim A$ and $A \sim B$ stands for $A \lesssim B$ and $A \gtrsim B$ simultaneously.

In order to carry out the complexity analysis of the estimate (2.3) we need the so-called "margin" or boundary assumption.

(AM) There exist constants $A > 0$, $\delta_0 > 0$ and $\alpha > 0$ such that

$$\mathrm{P}\left(|C_j^*(Z_j) - g_j(Z_j)| \le \delta\right) \le A\delta^\alpha$$

for all $j = 0,\dots,\mathscr{J}$, and all $\delta < \delta_0$.

Remark 4. Assumption (AM) provides a useful characterization of the behavior of the continuation values (C_j^*) and payoffs (g_j) near the exercise boundary $\partial\mathscr{E}$ with

$$\mathscr{E} = \left\{(j,x) : g_j(x) \geq C_j^*(x)\right\}.$$

In the situation when all functions $C_j^* - g_j$, $j = 0,\dots,\mathscr{J}-1$, are smooth and have non-vanishing derivatives in the vicinity of the exercise boundary, we have $\alpha = 1$. Other values of α are possible as well, see Belomestny (2011).

Let us now turn to the properties of the estimate $V_0^{n,k}$. While the variance of the estimate $V_0^{n,k}$ is given by

$$\mathrm{Var}[V_0^{n,k}] = \mathrm{Var}[g_{\tau_k}(Z_{\tau_k})]/n, \tag{3.1}$$

its bias is analyzed in the following theorem.

Theorem 5. *Suppose that (AP), (AM) and (AQ) hold with some $\alpha > 0$, and all functions g_j are uniformly bounded, i.e.,*

$$|g_j(x)| \leq G, \quad x \in \mathbb{R}^d.$$

for some constants $G > 0$. Then it holds

$$\left|V_0^* - \mathrm{E}[V_0^{n,k}]\right| \lesssim \gamma_k^{(1+\alpha)/2}, \quad k \to \infty.$$

The next theorem gives an upper estimate for the complexity of $V_0^{n,k}$.

Theorem 6. *Let assumptions (AP), (AC), (AQ) and (AM) hold with*

$$\gamma_k = k^{-\mu}, \quad k \in \mathbb{N}$$

for some $\mu > 0$. Then for any $\varepsilon > 0$ the choice

$$k^* = \varepsilon^{-\frac{2}{\mu(1+\alpha)}}, \quad n^* = \varepsilon^{-2}$$

leads to

$$\mathrm{E}\left[V_0^{n^*,k^*} - V_0^*\right]^2 \leq \varepsilon^2,$$

and the complexity of the estimate $V_0^{n^*,k^*}$ is bounded from above by $\mathscr{C}_{n^*,k^*}(\varepsilon)$ with

$$\mathscr{C}_{n^*,k^*}(\varepsilon) \lesssim \varepsilon^{-2\cdot\max\left(\frac{x_1+1}{\mu(1+\alpha)}, 1+\frac{x_2}{\mu(1+\alpha)}\right)}, \quad \varepsilon \to 0. \tag{3.2}$$

Discussion Theorem 6 implies that the complexity of $V_0^{n^*,k^*}$ is always larger than ε^{-2}. In the case $\varkappa_1 = 1$ and $\varkappa_2 = 1$ (mesh method or local regression) we get

$$\mathscr{C}_{n^*,k^*}(\varepsilon) \lesssim \varepsilon^{-2\max\left(\frac{2}{\mu(1+\alpha)}, 1+\frac{1}{\mu(1+\alpha)}\right)} \tag{3.3}$$

Furthermore, in the most common case $\alpha = 1$ the bound (3.3) simplifies to

$$\mathscr{C}_{n^*,k^*}(\varepsilon) \lesssim \varepsilon^{-2\max\left(\frac{1}{\mu}, 1+\frac{1}{2\mu}\right)}.$$

Since for all regression methods and the mesh method $\mu \le 1$, the asymptotic complexity is always larger than ε^{-3}. In the next section we present a multilevel approach that can reduce the asymptotic complexity down to ε^{-2} in some cases.

4 Multilevel approach

Fix some natural number L and let $\mathbf{k} = (k_0, k_1, \dots, k_L)$ and $\mathbf{n} = (n_0, n_1, \dots, n_L)$ be two sequences of natural numbers, satisfying $k_0 < k_1 < \dots < k_L$ and $n_0 > n_1 > \dots > n_L$. Define

$$V_0^{\mathbf{n},\mathbf{k}} = \frac{1}{n_0}\sum_{r=1}^{n_0} g_{\tau_{k_0}^{(r)}}\left(Z_{\tau_{k_0}^{(r)}}^{(r)}\right) + \sum_{l=1}^{L}\frac{1}{n_l}\sum_{r=1}^{n_l}\left[g_{\tau_{k_l}^{(r)}}\left(Z_{\tau_{k_l}^{(r)}}^{(r)}\right) - g_{\tau_{k_{l-1}}^{(r)}}\left(Z_{\tau_{k_{l-1}}^{(r)}}^{(r)}\right)\right]$$

with

$$\tau_k^{(r)} = \inf\left\{0 \le j \le \mathscr{J} : g_j(Z_j^{(r)}) \ge C_{k,j}(Z_j^{(r)})\right\}, \quad k \in \mathbb{N},$$

where for any $l = 1, \dots, L$, both estimates $C_{k_l,j}$ and $C_{k_{l-1},j}$ are based on one set of k_l training trajectories. Let us analyse the properties of the estimate $V_0^{\mathbf{n},\mathbf{k}}$. First note that its bias coincides with the bias of $g_{\tau_{k_L}}(Z_{\tau_{k_L}})$ corresponding to the finest approximation level. As to the variance of $V_0^{\mathbf{n},\mathbf{k}}$, it can be significantly reduced due the use of "good" continuation value estimates $C_{k_{l-1},j}$ and $C_{k_l,j}$ (that are both close to C_j^*) on the same set of testing trajectories in each level. In this way a "coupling" effect is achieved. The following theorem quantifies the above heuristics.

Theorem 7. *Let (AP), (AQ) and (AM) hold with some $\alpha > 0$, then the estimate $V_0^{\mathbf{n},\mathbf{k}}$ has the bias of order $\gamma_{k_L}^{(1+\alpha)/2}$ and the variance of order*

$$\frac{\mathrm{Var}[g(X_{\tau_{k_0}})]}{n_0} + \sum_{l=1}^{L}\frac{\gamma_{k_{l-1}}^{\alpha/2}}{n_l}.$$

Furthermore, under assumption (AC) the cost of $V_0^{\mathbf{n},\mathbf{k}}$ is bounded from above by a multiple of

$$\sum_{l=0}^{L}(k_l^{\varkappa_1+1} + n_l \cdot k_l^{\varkappa_2})$$

Finally, the complexity of $V_0^{\mathbf{n},\mathbf{k}}$ is given by the following theorem.

Theorem 8. *Let assumptions (AP), (AC), (AQ) and (AM) hold with*

$$\gamma_{k_l} = k_l^{-\mu}, \quad k_l \in \mathbb{N}$$

for some $\mu > 0$. Then under the choice $k_l^* = k_0 \cdot \theta^l$, $l = 0, 1, \ldots, L$, with $\theta > 1$,

$$L = \left\lceil \frac{2}{\mu(1+\alpha)} \log_\theta \left(\varepsilon^{-1} \cdot k_0^{-\mu(1+\alpha)/2} \right) \right\rceil$$

and

$$n_l^* = \varepsilon^{-2} \left(\sum_{i=1}^{L} \sqrt{k_i^{(\varkappa_2 - \mu\alpha/2)}} \right) \cdot \sqrt{k_l^{(-\varkappa_2 - \mu\alpha/2)}}$$

the complexity of the estimate (2.3) *is bounded, up to a constant, from above by*

$$\mathscr{C}_{\mathbf{n}^*,\mathbf{k}^*}(\varepsilon) \lesssim \begin{cases} \varepsilon^{-2\cdot\max\left(\frac{\varkappa_1+1}{\mu(1+\alpha)},1\right)}, & 2\cdot\varkappa_2 < \mu\alpha \\ \varepsilon^{-2\cdot\frac{\varkappa_1+1}{\mu(1+\alpha)}}, & 2\cdot\varkappa_2 = \mu\alpha \text{ and } \frac{\varkappa_1+1}{\mu(1+\alpha)} > 1 \\ \varepsilon^{-2}\cdot\left(\log\varepsilon\right)^2, & 2\cdot\varkappa_2 = \mu\alpha \text{ and } \frac{\varkappa_1+1}{\mu(1+\alpha)} \le 1 \\ \varepsilon^{-2\cdot\max\left(\frac{\varkappa_1+1}{\mu(1+\alpha)},1+\frac{\varkappa_2-\mu\alpha/2}{\mu(1+\alpha)}\right)}, & 2\cdot\varkappa_2 > \mu\alpha \end{cases} \tag{4.1}$$

Discussion Let us compare the complexities of the estimates $V_0^{n^*,k^*}$ and $V_0^{\mathbf{n}^*,\mathbf{k}^*}$. For the sake of clarity we will assume that $\varkappa_1 = \varkappa_2 = \varkappa$ as in the mesh or local regression methods. Then (4.1) versus (3.2) can be written as

$$\begin{cases} \varepsilon^{-2\cdot\max\left(\frac{\varkappa+1}{\mu(1+\alpha)},1\right)}, & 2\cdot\varkappa < \mu\alpha \\ \varepsilon^{-2\cdot\frac{\varkappa+1}{\mu(1+\alpha)}}, & 2\cdot\varkappa = \mu\alpha \text{ and } \frac{\varkappa+1}{\mu(1+\alpha)} > 1 \\ \varepsilon^{-2}\cdot\left(\log\varepsilon\right)^2, & 2\cdot\varkappa = \mu\alpha \text{ and } \frac{\varkappa+1}{\mu(1+\alpha)} \le 1 \\ \varepsilon^{-2\cdot\max\left(\frac{\varkappa+1}{\mu(1+\alpha)},1+\frac{\varkappa-\mu\alpha/2}{\mu(1+\alpha)}\right)}, & 2\cdot\varkappa > \mu\alpha \end{cases} \vee \varepsilon^{-2\cdot\max\left(\frac{\varkappa+1}{\mu(1+\alpha)},\,1+\frac{\varkappa}{\mu(1+\alpha)}\right)}$$

Now it is clear that multilevel algorithm will not be superior to the standard Monte Carlo algorithm in the case $\mu(1+\alpha) \le 1$. In the case $\mu(1+\alpha) > 1$, the computational gain, up to a logarithmic factor, is given by

$$\begin{cases} \varepsilon^{-2\cdot\min\left(\frac{\varkappa}{\mu(1+\alpha)},1-\frac{1}{\mu(1+\alpha)}\right)}, & 2\cdot\varkappa < \mu\alpha \\ \varepsilon^{-2\cdot\left(1-\frac{1}{\mu(1+\alpha)}\right)}, & 2\cdot\varkappa = \mu\alpha \text{ and } \frac{\varkappa+1}{\mu(1+\alpha)} > 1 \\ \varepsilon^{-2\cdot\frac{\varkappa}{\mu(1+\alpha)}}, & 2\cdot\varkappa = \mu\alpha \text{ and } \frac{\varkappa+1}{\mu(1+\alpha)} \le 1 \\ \varepsilon^{-2\cdot\min\left(1-\frac{1}{\mu(1+\alpha)},\frac{\mu\alpha/2}{\mu(1+\alpha)}\right)}, & 2\cdot\varkappa > \mu\alpha \end{cases}$$

Taking into account the fact that $\alpha = 1$ in the usual situation, we conclude that it is advantageous to use MLMC as long as $\mu > 1/2$.

5 Numerical example: Bermudan max calls on multiple assets

Suppose that the price of the underlying asset $X = (X^1, \dots, X^d)$ follows a Geometric Brownian motion (GBM) under the risk-neutral measure, i.e.,

$$dX_t^i = (r - \delta)X_t^i dt + \sigma X_t^i dB_t^i, \tag{5.1}$$

where r is the risk-free interest rate, δ the dividend rate, σ the volatility, and $B_t = (B_t^1, \dots, B_t^d)$ is a vector of d independent standard Brownian motions. At any time $t \in \{t_0, ..., t_{\mathscr{J}}\}$ the holder of the option may exercise it and receive the payoff

$$h(X_t) = e^{-rt}(\max(X_t^1, ..., X_t^d) - \kappa)^+.$$

We consider a benchmark example (see, e.g. Broadie and Glasserman (2004), p. 462) when $d = 5$, $\sigma = 0.2$, $r = 0.05$, $\delta = 0.1$, $\kappa = 100$ $t_j = jT/\mathscr{J}$, $j = 0, \dots, \mathscr{J}$, with $T = 3$ and $\mathscr{J} = 3$.

5.1 Mesh method

First note that for the mesh method the conditions of Theorem 6 and Theorem 8 are fulfilled with $\gamma_k = 1/k$ in (AQ) and $\kappa_1 = \kappa_2 = 1$ in (AC). Moreover, for the problem at hand, the assumption (AB) holds with $\alpha = 1$. Consider the standard MC mesh approach. For any $\varepsilon > 0$ we set

$$k = (\varepsilon/2.4)^{-1}, \quad n = (\varepsilon/2.4)^{-2}$$

and simulate independently k training paths of the process Z using the exact formula

$$Z_j^{(i)} = Z_{j-1}^{(i)} \exp\left(\left[r - \delta - \frac{1}{2}\sigma^2\right](t_j - t_{j-1}) + \sigma\sqrt{(t_j - t_{j-1})} \cdot \xi_j^i\right),$$

where ξ_j^i, $i = 1, \dots, k$, are i. i. d. standard normal random variables. The conditional density of Z_j given Z_{j-1} is given by

$$p_j(x, y) = \prod_{i=1}^{d} p_j(x_i, y_i), \quad x = (x_1, \dots, x_d), \quad y = (y_1, \dots, y_d),$$

where

$$\begin{aligned} p_j(x_i, y_i) &= \frac{x_i}{y_i \sigma \sqrt{2\pi(t_j - t_{j-1})}} \times \\ &\quad \times \exp\left(\frac{-\left(\log\left(\frac{y_i}{x_i}\right) - \left(r - \delta - \frac{1}{2}\sigma^2\right)(t_j - t_{j-1})\right)^2}{2\sigma^2(t_j - t_{j-1})}\right). \end{aligned}$$

Using the above paths we construct the sequence of the estimates (training phase)

$$C_{k,0}(x), \ldots, C_{k,\mathscr{J}}(x)$$

as described in Example 3 and then in testing phase compute the estimate $V_0^{n,k}$ via (2.3). Note that for the variance reduction we use inner and outer control variates based on the analytical formula for the European max-call option (see Broadie and Glasserman (2004))

$$\mathscr{E}(x,t,T) := E\left[e^{-rT} \max_{k=1,\ldots,d} \left(X_T^k - \kappa\right)^+ \middle| X_t = x\right].$$

Finally we approximate the mean square error (MSE) of the estimate $V_0^{n,k}$ based on 100 repetitions of the training and testing phases. The plot of the estimated quotient $\sqrt{\text{MSE}}/\varepsilon$ is shown on the l.h.s. of Figure 5.1. Turn now to the ML approach. Here we take $\mathbf{n} = (n_0, n_1, \ldots, n_l)$ and $\mathbf{k} = (k_0, k_1, \ldots, k_l)$ with $k_0 = 5$

$$k_l = k_0 \cdot 2^l, \quad n_l = \frac{1}{(\varepsilon/8)^2} \left(\sum_{i=1}^{L} \sqrt{k_i^{1/2}}\right) \sqrt{k_l^{-3/2}}, \quad l = 0, \ldots, L,$$

and

$$L = \left\lceil \log_\theta \left(8 \cdot k_0/\varepsilon\right) \right\rceil.$$

The grid for ε on the r.h.s. of Figure 5.1 is chosen in such a way that $L(\varepsilon)$ runs through the set $\{1, 2, \ldots, 7\}$. The plot of the estimated quotient $\sqrt{\text{MSE}}/\varepsilon$ is shown on the r.h.s. of Figure 5.1. Figure 5.1 suggests that the rates given in Theorem 5 and Theorem 8 do hold. Next we compare the computational cost

$$\sum_{l=0}^{L} (k_l^2 + n_l \cdot k_l)$$

where k_l, n_l and L are defined above to the theoretical complexity given by $\varepsilon^{-2.5}$. In Figure 5.2 we present the corresponding log-plots of complexities and the gains as functions of $\log(1/\varepsilon)$.

5.2 Local regression

We use the local constant regression with the simplest kernels of the form:

$$K(z) = 1(|z| \leq 1)$$

and define

$$C_{k,j}(z) = \sum_{i=1}^{k} \zeta_{k,j+1}(Z_{j+1}^{(i)}) \cdot w_{ij}^k(z), \tag{5.2}$$

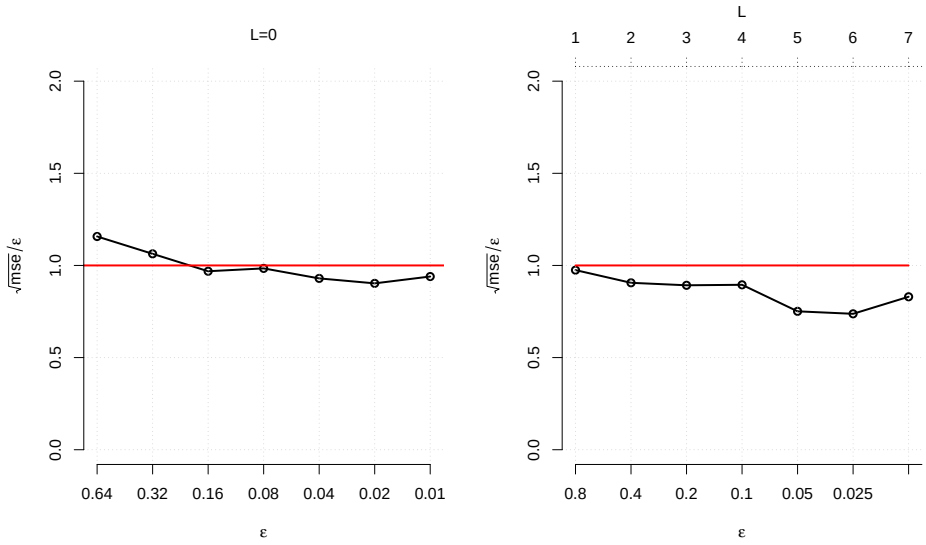

Figure 5.1: Mesh method: mean square errors of the standard Monte Carlo estimate $V_0^{n,k}$ (left) and the multilevel estimate $V_0^{\mathbf{n},\mathbf{k}}$ (right) in the units of the expected error ε.

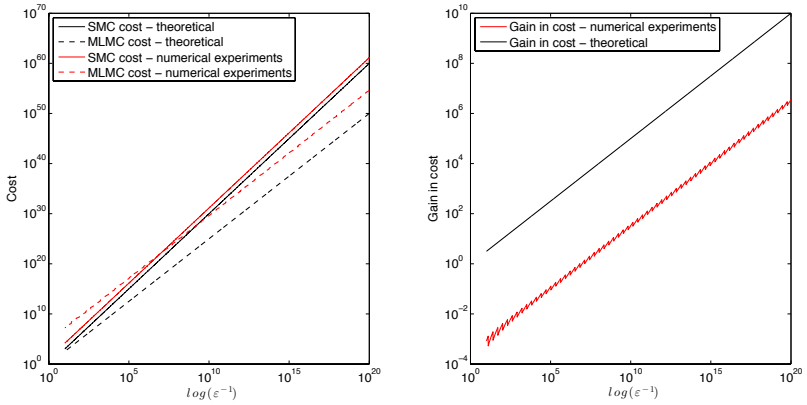

Figure 5.2: Comparison of theoretical and numerical costs (left) and theoretical and numerical gains (right) as functions of $\log(\varepsilon^{-1})$

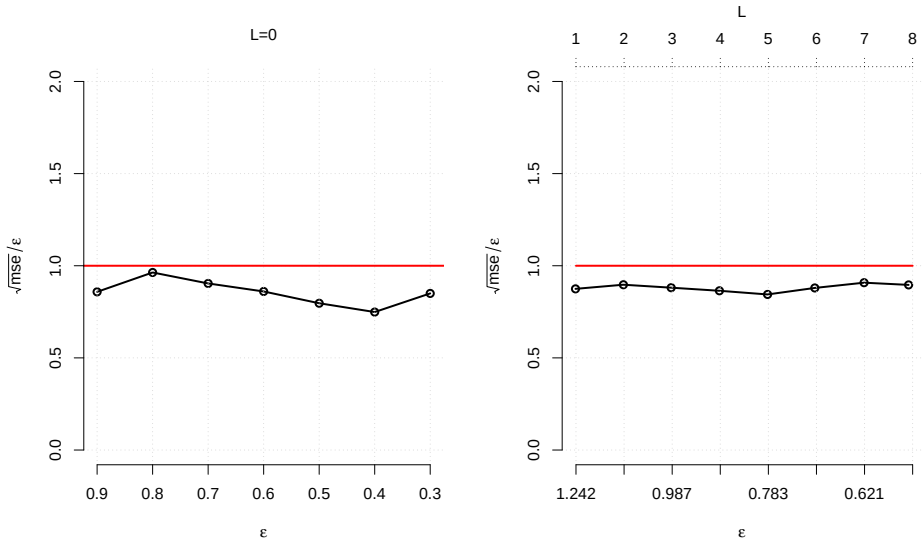

Figure 5.3: Local regression: mean square errors of the standard Monte Carlo estimate $V_0^{n,k}$ (left) and the multilevel estimate $V_0^{\mathbf{n},\mathbf{k}}$ (right) in the units of the expected error ε.

where

$$w_{ij}^k(z) = \frac{1\big(|z - Z_j^{(i)}| \le \delta_k\big)}{\sum_{l=1}^k 1\big(|z - Z_j^{(l)}| \le \delta_k\big)}, \quad i = 1, \dots, k,$$

with $\delta_k = 100 \cdot k^{-1/(d+2)}$ (see Example 3). For any $\varepsilon > 0$ we set

$$k = (\varepsilon/1.2)^{-6}, \quad n = (\varepsilon/1.2)^{-2}$$

corresponding to the choice $\gamma_k = k^{-1/6}$ ($\mu = 1/6$) in Theorem 6 and approximate the mean square error (MSE) of the MC estimate $V_0^{n,k}$ based on 100 repetitions of the training and testing phases. In the case of the MLMC algorithm we take $k_0 = 100$,

$$L = \left\lceil 6 \cdot \log_\theta \left(\frac{3}{\varepsilon \cdot k_0^{1/6}} \right) \right\rceil,$$

and

$$k_l = k_0 2^l, \quad n_l = \frac{10}{(\varepsilon/3)^2} \left(\sum_{i=1}^L \sqrt{(k_i)^{11/12}} \right) \sqrt{(k_l)^{-13/12}}, \quad l = 0, \dots, L.$$

The results in form of the quotients $\sqrt{MSE}/\varepsilon$ are shown in Figure 5.2.

6 Proofs

6.1 Proof of Theorem 5

A family of stopping times $\left(\tau_j\right)_{j=0,\dots,\mathscr{J}}$ w.r.t. the filtration $(\mathscr{F}_j)_{j=0,\dots,\mathscr{J}}$ is called consistent if

$$j \le \tau_j \le \mathscr{J}, \quad \tau_{\mathscr{J}} = \mathscr{J}$$

and

$$\tau_j > j \quad \Longrightarrow \quad \tau_j = \tau_{j+1}.$$

Lemma 9. *Let $(Y_j)_{j=0,\dots,\mathscr{J}}$ be a process adapted to the filtration $(\mathscr{F}_j)_{j=0,\dots,\mathscr{J}}$ and let $\left(\tau_j^1\right)$ and $\left(\tau_j^2\right)$ be two consistent families of stopping times. Then*

$$\mathrm{E}^{\mathscr{F}_j}\left[Y_{\tau_j^1} - Y_{\tau_j^2}\right] = \mathrm{E}^{\mathscr{F}_j}\left\{\sum_{l=j}^{\mathscr{J}-1}\left(Y_l - \mathrm{E}^{\mathscr{F}_l}\left[Y_{\tau_{l+1}^1}\right]\right)\left(1_{\{\tau_l^1=l,\tau_l^2>l\}} - 1_{\{\tau_l^1>l,\tau_l^2=l\}}\right)1_{\{\tau_l^2>l\}}\right\}$$

for any $j = 0,\dots,\mathscr{J}-1$.

Proof. We have

$$\begin{aligned} Y_{\tau_j^1} - Y_{\tau_j^2} &= \left[Y_j - Y_{\tau_j^2}\right]1_{\{\tau_j^1=j,\tau_j^2>j\}} + \left[Y_{\tau_j^1} - Y_j\right]1_{\{\tau_j^1>j,\tau_j^2=j\}} \\ &\quad + \left[Y_{\tau_j^1} - Y_{\tau_j^2}\right]1_{\{\tau_j^1>j,\tau_j^2>j\}} \\ &= \left[Y_j - Y_{\tau_{j+1}^1}\right]1_{\{\tau_j^1=j,\tau_j^2>j\}} + \left[Y_{\tau_{j+1}^1} - Y_j\right]1_{\{\tau_j^1>j,\tau_j^2=j\}} \\ &\quad + \left[Y_{\tau_{j+1}^1} - Y_{\tau_{j+1}^2}\right]1_{\{\tau_j^1=j,\tau_j^2>j\}} + \left[Y_{\tau_{j+1}^1} - Y_{\tau_{j+1}^2}\right]1_{\{\tau_j^1>j,\tau_j^2>j\}}. \end{aligned}$$

Therefore it holds for $\Delta_j = \mathrm{E}^{\mathscr{F}_j}\left[Y_{\tau_j^1} - Y_{\tau_j^2}\right]$

$$\Delta_j = \left[Y_j - \mathrm{E}^{\mathscr{F}_j}\left[Y_{\tau_{j+1}^1}\right]\right]\left(1_{\{\tau_j^1=j,\tau_j^2>j\}} - 1_{\{\tau_j^1>j,\tau_j^2=j\}}\right) + \mathrm{E}^{\mathscr{F}_j}\left\{\Delta_{j+1}1_{\{\tau_j^2>j\}}\right\}$$

with $\Delta_{\mathscr{J}} = 0$ and

$$\Delta_j = \mathrm{E}^{\mathscr{F}_j}\left\{\sum_{l=j}^{\mathscr{J}-1}\left(Y_l - \mathrm{E}^{\mathscr{F}_l}\left[Y_{\tau_{l+1}^1}\right]\right)\left(1_{\{\tau_l^1=l,\tau_l^2>l\}} - 1_{\{\tau_l^1>l,\tau_l^2=l\}}\right)1_{\{\tau_l^2>l\}}\right\}.$$

□

Taking into account that

$$C_l^*(Z_l) = \mathrm{E}^{\mathscr{F}_l}\left[g_{\tau_{l+1}^*}(Z_{\tau_{l+1}^*})\right] \le g_l(Z_l)$$

on $\{\tau_l^* = l\}$ and

$$C_l^*(Z_l) < g_l(Z_l)$$

on $\{\tau_l^* > l\}$, we get from Lemma 9 for $R = V_0^{n,k} - V_0^*$

$$\begin{aligned} |R| &= \left|\mathrm{E}\left[g_{\tau_k^*}(Z_{\tau_k^*}) - g_{\tau_k}(Z_{\tau_k})\right]\right| \\ &\le \mathrm{E}\left[\sum_{l=0}^{\mathscr{J}-1}\left|C_l^*(Z_l) - g_l(Z_l)\right|\left(1_{\{\tau_{k,l}^*=l,\tau_{k,l}>l\}} + 1_{\{\tau_{k,l}^*>l,\tau_{k,l}=l\}}\right)\right]. \end{aligned}$$

Introduce

$$\begin{aligned} \mathscr{E}_{k,j} &= \{g_j(Z_j) > C_{k,j}^*(Z_j),\ g_j(Z_j) \le C_{k,j}(Z_j)\} \\ &\quad \cup \{g_j(Z_j) \le C_k^*(Z_j),\ g_j(Z_j) > C_{k,j}(Z_j)\}, \\ \mathscr{A}_{k,j,0} &= \left\{0 < \left|g_j(Z_j) - C_j^*(Z_j)\right| \le \gamma_k^{-1/2}\right\}, \\ \mathscr{A}_{k,j,i} &= \left\{2^{i-1}\gamma_k^{-1/2} < \left|g_j(Z_j) - C_j^*(Z_j)\right| \le 2^i\gamma_k^{-1/2}\right\} \end{aligned}$$

for $j = 0, \ldots, \mathscr{J}-1$ and $i > 0$. It holds

$$\begin{aligned} |R| &\le \mathrm{E}\left[\sum_{l=0}^{\mathscr{J}-1}\left|C_l^*(Z_l) - g_l(Z_l)\right| 1_{\{\mathscr{E}_{k,l}\}}\right] \\ &= \mathrm{E}\left[\sum_{i=0}^{\infty}\sum_{l=0}^{\mathscr{J}-1}\left|C_l^*(Z_l) - g_l(Z_l)\right| 1_{\{\mathscr{E}_{k,l}\cap\mathscr{A}_{k,l,i}\}}\right] \\ &= \gamma_k^{-1/2}\sum_{l=0}^{\mathscr{J}-1}\mathrm{P}\left(\left|g_l(Z_l) - C_l^*(Z_l)\right| \le \gamma_k^{-1/2}\right) \\ &\quad + \mathrm{E}\left[\sum_{i=1}^{\infty}\sum_{l=0}^{\mathscr{J}-1}\left|C_l^*(Z_l) - g_l(Z_l)\right| 1_{\{\mathscr{E}_{k,l}\cap\mathscr{A}_{k,l,i}\}}\right]. \end{aligned}$$

Using the fact that $\left|g_l(Z_l) - C_l^*(Z_l)\right| \le \left|C_l(Z_l) - C_l^*(Z_l)\right|$ on $\mathscr{E}_{k,l}$, we derive

$$\begin{aligned} |R| &\le \gamma_k^{-1/2}\sum_{l=0}^{\mathscr{J}-1}\mathrm{P}\left(\left|g_l(Z_l) - C_l^*(Z_l)\right| \le \gamma_k^{-1/2}\right) \\ &\quad + \sum_{i=1}^{\infty}2^i\gamma_k^{-1/2}\mathrm{E}\left[\sum_{l=0}^{\mathscr{J}-1}1_{\left\{\left|g_j(Z_j) - C_j^*(Z_j)\right| \le 2^i\gamma_k^{-1/2}\right\}}\mathrm{P}^k\left(\left|C_{k,l}(Z_l) - C_l^*(Z_l)\right| > 2^{i-1}\gamma_k^{-1/2}\right)\right] \\ &\le A\mathscr{J}\gamma_k^{-\alpha/2} + A\mathscr{J}\gamma_k^{-\alpha/2}\sum_{i=1}^{\infty}2^iB_1\exp(-B_2 2^{i-1}). \end{aligned}$$

6.2 Proof of Theorem 6

Based on (3.1) we have the optimization problem

$$k^{\varkappa_1+1} + n \cdot k^{\varkappa_2} \to \min$$
$$\gamma_k^{(1+\alpha)/2} \le \varepsilon$$
$$n \ge \varepsilon^{-2}$$

It is clear that

$$\gamma_k^{(1+\alpha)/2} = k^{-\mu(1+\alpha)/2} \Rightarrow k \ge \varepsilon^{-\frac{2}{\mu(1+\alpha)}},$$

which immediately leads to the statement.

6.3 Proof of Theorem 7

The formula for the variance follows from the estimate

$$\begin{aligned}
\mathrm{E}\left[g_{\tau_{k_l}}\left(Z_{\tau_{k_l}}\right) - g_{\tau_{k_{l-1}}}\left(Z_{\tau_{k_{l-1}}}\right)\right]^2 &\le \mathrm{E}\left[g_{\tau^*}\left(Z_{\tau^*}\right) - g_{\tau_{k_{l-1}}}\left(Z_{\tau_{k_{l-1}}}\right)\right]^2 \\
&\quad + \mathrm{E}\left[g_{\tau_{k_l}}\left(Z_{\tau_{k_l}}\right) - g_{\tau^*}\left(Z_{\tau^*}\right)\right]^2 \\
&\le 2^{\mathscr{J}} G^2 \sum_{l=0}^{\mathscr{J}-1}\left[\mathrm{P}(\mathscr{E}_{k_{l-1},l}) + \mathrm{P}(\mathscr{E}_{k_l,l})\right],
\end{aligned}$$

where for any k

$$\begin{aligned}
\sum_{l=0}^{\mathscr{J}-1} \mathrm{P}(\mathscr{E}_{k,l}) &\le \sum_{l=0}^{\mathscr{J}-1} \mathrm{P}\left(\left|g_l(Z_l) - C_l^*(Z_l)\right| \le \gamma_k^{-1/2}\right) \\
&\quad + \sum_{i=1}^{\infty} 2^i \mathrm{E}\left[\sum_{l=0}^{\mathscr{J}-1} 1_{\left\{\left|g_j(Z_j) - C_j^*(Z_j)\right| \le 2^i \gamma_k^{-1/2}\right\}} \mathrm{P}^k\left(\left|C_{k,l}(Z_l) - C_l^*(Z_l)\right| > 2^{i-1}\gamma_k^{-1/2}\right)\right] \\
&\le A\mathscr{J}\gamma_k^{-\alpha/2} + A\mathscr{J}\gamma_k^{-\alpha/2} \sum_{i=1}^{\infty} 2^i B_1 \exp(-B_2 2^{i-1}).
\end{aligned}$$

6.4 Proof of Theorem 8

Due to the monotone structure of the functional, we can consider the following optimization problem:

$$\sum_{l=0}^{L} k_l^{\varkappa_1+1} + n_l \cdot k_l^{\varkappa_2} \to \min \tag{6.1}$$

$$\gamma_{k_L}^{(1+\alpha)/2} = k_L^{-\mu(1+\alpha)/2} = \left(k_0 \cdot \theta^L\right)^{-\mu(1+\alpha)/2} \le \varepsilon \tag{6.2}$$

$$\frac{1}{n_0} + \sum_{l=1}^{L} \frac{\gamma_{k_{l-1}}^{\alpha/2}}{n_l} \asymp k_0^{-\mu\alpha/2} \cdot \sum_{l=0}^{L} \frac{\theta^{-l\mu\alpha/2}}{n_l} = \varepsilon^2 \tag{6.3}$$

Now the Lagrange multiplier method with respect to n_l gives us

$$k_l^{x_2} = -\lambda \frac{k_l^{-\mu\alpha/2}}{n_l^2} \Rightarrow n_l = \sqrt{(-\lambda) \cdot k_l^{(-x_2-\mu\alpha/2)}}.$$

Now one can put the value of n_l in (6.3):

$$\sum_{l=1}^{L} \frac{\gamma_{k_{l-1}}^{\alpha/2}}{n_l} \asymp \sum_{l=1}^{L} \frac{k_l^{-\mu\alpha/2}}{\sqrt{(-\lambda) \cdot k_l^{(-x_2-\mu\alpha/2)}}} \asymp \varepsilon^2$$

$$\Downarrow$$

$$\sqrt{(-\lambda)} = \varepsilon^{-2} \cdot \sum_{l=1}^{L} \sqrt{k_l^{(x_2-\mu\alpha/2)}}$$

$$\Downarrow$$

$$n_l = \varepsilon^{-2} \left(\sum_{i=1}^{L} \sqrt{k_i^{(x_2-\mu\alpha/2)}} \right) \cdot \sqrt{k_l^{(-x_2-\mu\alpha/2)}}.$$

For total number of level we have from (6.2):

$$\left(k_0 \cdot \theta^L\right)^{-\mu(1+\alpha)/2} \leq \varepsilon \Rightarrow L \geq \frac{2}{\mu(1+\alpha)} \log_\theta \left(\varepsilon^{-1} \cdot k_0^{-\mu(1+\alpha)/2}\right).$$

Now we can rewrite (6.1) as

$$\sum_{l=0}^{L} k_l^{x_1+1} + n_l \cdot k_l^{x_2} \asymp k_L^{x_1+1} + \varepsilon^{-2} \cdot \left(\sum_{l=1}^{L} \sqrt{k_l^{(x_2-\mu\alpha/2)}} \right)^2,$$

so we will have three cases.

Case 1. $2 \cdot x_2 = \mu\alpha$.

$$\begin{aligned} k_L^{x_1+1} + \varepsilon^{-2} \cdot \left(\sum_{l=1}^{L} \sqrt{k_l^{(x_2-\mu\alpha/2)}} \right)^2 &\gtrsim k_L^{x_1+1} + \varepsilon^{-2} \cdot L^2 \\ &\geq \varepsilon^{-\frac{2\cdot(x_1+1)}{\mu(1+\alpha)}} + \varepsilon^{-2} \cdot L^2 \end{aligned}$$

Case 2. $2 \cdot x_2 < \mu\alpha$.

$$\begin{aligned} k_L^{x_1+1} + \varepsilon^{-2} \cdot \left(\sum_{l=1}^{L} \sqrt{k_l^{(x_2-\mu\alpha/2)}} \right)^2 &\gtrsim k_L^{x_1+1} + \varepsilon^{-2} \\ &\geq \varepsilon^{-\frac{2\cdot(x_1+1)}{\mu(1+\alpha)}} + \varepsilon^{-2} \end{aligned}$$

Case 3. $2 \cdot \varkappa_2 > \mu\alpha$.

$$k_L^{\varkappa_1+1} + \varepsilon^{-2} \cdot \left(\sum_{l=1}^{L} \sqrt{k_l^{(\varkappa_2 - \mu\alpha/2)}} \right)^2 \gtrsim k_L^{\varkappa_1+1} + \varepsilon^{-2} \cdot k_L^{\varkappa_2 - \mu\alpha/2}$$
$$\geq \varepsilon^{-\frac{2\cdot(\varkappa_1+1)}{\mu(1+\alpha)}} + \varepsilon^{-2-\frac{2\varkappa_2-\mu\alpha}{\mu(1+\alpha)}}$$

Combining all three cases one will get (4.1).

REFERENCES

A. Agarwal and S. Juneja (2013). Comparing optimal convergence rate of stochastic mesh and least squares method for bermudan option pricing. Proceedings of the 2013 Winter Simulation Conference.

L. Andersen: A simple approach to the pricing of Bermudan swaptions in the multi-factor Libor Market Model. *J. Computat. Financ.*, **3**, 5–32 (2000).

D. Belomestny: Pricing Bermudan options using nonparametric regression: optimal rates of convergence for lower estimates. *Finance and Stochastics*, **15**(4), 655–683 (2011).

D. Belomestny, J. Schoenmakers and F. Dickmann: Multilevel dual approach for pricing American style derivatives, to appear in *Finance and Stochastics* (2013).

M. Broadie and P. Glasserman: Pricing American-style securities using simulation. *J. Econ. Dyn. Con.*, **21**, 1323–1352 (1997).

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

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

D. Egloff: Monte Carlo algorithms for optimal stopping and statistical learning. *Ann. Appl. Probab.*, **15**, 1396–1432 (2005).

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

E. Giné and A. Guillou: A law of the iterated logarithm for kernel density estimators in the presence of censoring. *Ann. I. H. Poincaré*, **37**, 503–522 (2001)

P. Glasserman: Monte Carlo Methods in Financial Engineering. In: Springer (2004).

P. Glasserman and B. Yu: Number of Paths Versus Number of Basis Functions in American Option Pricing. *Ann. Appl. Probab.*, **14**, 2090–2119 (2004)

P. Kloeden and E. Platen: Numerical solution of stochastic differential equations. Applications of Mathematics (New York), 23. Springer-Verlag, Berlin (1992).

F. Longstaff and E. Schwartz: Valuing American options by simulation: a simple least-squares approach. *Rev. Financ. Stud.*, **14**, 113–147 (2001).

J. Tsitsiklis and B. Van Roy: Regression methods for pricing complex American style options. *IEEE Trans. Neural. Net.*, **12**, 694–703 (1999).

D. Zanger: Quantitative error estimates for a least-squares Monte Carlo algorithms for American option pricing. *Finance and Stochastics*, **17**, 503–534 (2013).