Optimal dual martingales, their analysis and application to new algorithms for Bermudan products[1]

John Schoenmakers and Junbo Huang

April 29, 2011

Abstract

In this paper we introduce and study the concept of optimal and surely optimal dual martingales in the context of dual valuation of Bermudan options, and outline the development of new algorithms in this context. We provide a characterization theorem, a theorem which gives conditions for a martingale to be surely optimal, and a stability theorem concerning martingales which are near to be surely optimal in a sense. Guided by these results we develop a framework of backward algorithms for constructing such a martingale. In turn this martingale may then be utilized for computing an upper bound of the Bermudan product. The methodology is pure dual in the sense that it doesn't require certain (input) approximations to the Snell envelope.

In an Itô-Lévy environment we outline a particular regression based backward algorithm which allows for computing dual upper bounds without nested Monte Carlo simulation. Moreover, as a by-product this algorithm also provides approximations to the continuation values of the product, which in turn determine a stopping policy. Hence, we may obtain lower bounds at the same time.

In a first numerical study we demonstrate a backward dual regression algorithm in a Wiener environment that is easy to implement and is regarding accuracy comparable with the method of Belomestny et. al. (2009).

Keywords: Bermudan options, duality, Monte Carlo simulation, linear regression.

MSC: 62L15, 65C05, 91B28

[1]This paper is an updated version of the previous preprint Schoenmakers and Huang (2010). Supported by DFG Research Center MATHEON "Mathematics for Key Technologies" in Berlin. Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstr. 39, 10117 Berlin, {schoenma,huang}@wias-berlin.de .

1 Introduction

It is well-known that evaluation of Bermudan callable derivatives comes down to solving an optimal stopping problem. For many callable exotic products, for example interest products, the underlying state space is high-dimensional however. As such these products are usually very hard to solve with deterministic (PDE) methods and therefore simulation based (Monte Carlo) methods are called for. The first developments in this respect concentrated on the construction of a "good" exercise policy. We mention, among others, regression based methods by Carriere (1996), Longstaff and Schwartz (2001), and Tsistsiklis and Van Roy (1999), the stochastic mesh method of Broadie and Glasserman (2004), and quantization algorithms by Bally and Pages (2003). Especially for very high dimension, Kolodko and Schoenmakers (2004) developed a policy improvement approach which can be effectively combined with Longstaff and Schwartz (2001) for example (see Bender et al. (2008) and Bender et al. (2006)).

As a common feature, the aforementioned simulation methods provide lower biased estimates for the Bermudan product under consideration. As a new breakthrough, Rogers (2001), and Haugh and Kogan (2004) introduced a dual approach, which comes down to minimization over a set of martingales rather than maximization over a family of stopping times. By its very nature the dual approach gives upper biased estimates for the Bermudan product and after its appearance several numerical algorithms for computing dual upper bounds are proposed. Probably the most popular one is the method of Andersen and Broadie (2004), although this method requires nested Monte Carlo simulation (see also Kolodko and Schoenmakers (2004) and Schoenmakers (2005)). In a Wiener environment, Belomestny et. al. (2009) provide a fast generic method for computing dual upper bounds by non-nested simulation.

The algorithms for computing dual upper bounds so far have in common that they start with some given "good enough" approximation of the Snell envelope and then construct the Doob-martingale due to this approximation. In a recent paper Rogers (2010), points out how to construct a particular 'good' martingale via a sequence of martingales which are constant on an ever bigger time interval. In this construction no input approximation to the Snell envelope is used. The methods proposed in this paper have some flavor of the method of Rogers (2010), in the sense that no approximation to the Snell envelope is involved either. In a recent paper Desai et. al. (2010) treat the dual problem by methods from convex optimization theory.

Starting with a short resume of well-known facts on Bermudan derivatives in Section 2, we analyze in Section 3 the almost sure property of the dual representation in detail. We there introduce the concept of a surely optimal martingale, which is loosely speaking, a martingale that minimizes the dual representation with a particular almost sure property. In this respect we will point out that a martingale which minimizes the dual representation is not necessarily surely optimal, and on the other hand, a surely optimal martingale is generally not unique.

In Section 4 we present, as one of the main contributions in this paper, a characterizing theorem for surely optimal martingales (Theorem 6), and an adaptiveness criterion that guarantees that a martingale is surely optimal (Theorem 10 and Corollary 11).

In the application of the Andersen and Broadie (2004) algorithm one generally observes that, the better the constructed dual martingale, the lower the variance of the upper bound estimator. Actually this observation was not well studied from a mathematical point of view so far. In Section 5 we study this phenomenon and, as a next main contribution, give an explanation of it by a convergence or stability Theorem 12, related to surely optimal martingales.

Guided by Theorem 6, Theorem 10, and Theorem 12, we outline the development of backward algorithms in Section 6 for constructing martingales which are in a sense near to be surely optimal, and which subsequently may be used for evaluation of dual upper bounds. In this context we give a short recap of estimating conditional variances with kernel and regression based methods. In an Itô-Lévy environment it is shown how a particular backward algorithm may be designed as a regression procedure which yields dual martingales that allow for computing upper bounds without nested Monte Carlo (like in Belomestny et. al. (2009)). In this environment we obtain moreover, as a by-product, estimations of continuation values. Thus, as a result, we end up with a procedure which computes upper bounds as well as lower bounds at the same time by non-nested simulation. In a Wiener environment this procedure is quite easy to implement and may be considered as an interesting alternative to the non-nested method of Belomestny et. al. (2009), where a dual martingale is obtained by constructing a discretized Clark-Ocone derivative of some (input) approximation to the Snell envelope via regression.

In a first numerical study (Section 7) we illustrate at two multi-dimensional benchmark products (the same products as considered in Belomestny et. al. (2009)) a backward regression algorithm that is of the same quality as the one in Belomestny et. al. (2009) regarding speed and accuracy of upper bounds, and moreover produces very fast good lower bounds.

Finally, we underline that the focus in this paper is on the introduction and rigorous mathematical treatment of surely optimal martingales, and the way it leads to new dual algorithms for Bermudan products. An in depth analysis of convergence and performance of these algorithms is an interesting subject for further study, but considered beyond the scope of this paper.

2 Bermudan derivatives and optimal stopping

Let $(Z_i : i = 0, 1, \ldots, T)$[1] be a non-negative stochastic process in discrete time on a filtered probability space $(\Omega, \mathcal{F}, P)$, adapted to a filtration $\mathbb{F} := (\mathcal{F}_i : 0 \le i \le T)$ which satisfies $E|Z_i| < \infty$, for $0 \le i \le T$. The measure P may be considered as a pricing measure and the process Z may be seen as a (discounted)

[1] For notational convenience we have chosen for this stylized time set. The reader may reformulate all statements and results in this paper for a general discrete time set $\{T_0, T_1, \ldots, T_J\}$ in a trivial way.

cash-flow, which an investor may exercise once in the time set $\{0, ..., T\}$. As such he is faced with a Bermudan derivative. As a well known fact, a fair price of such a derivative is the value of the Snell envelope process

$$Y_i^* = \sup_{\tau \in \{i,...,T\},} E_i Z_\tau, \quad 0 \leq i \leq T, \tag{1}$$

at time $i = 0$. In (1), τ denotes a stopping time, $E_i := E_{\mathcal{F}_i}$ denotes conditional expectation with respect to the σ-algebra $\mathcal{F}_i$, and sup (inf) is to be understood as *essential supremum* (*infimum*) if it ranges over an uncountable family of random variables. Let us recall some well-known facts (e.g. see Neveu (1975)).

1. The Snell envelope Y^* of Z is the smallest super-martingale that dominates Z.
2. A family of optimal stopping times is given by

   $$\tau_i^* = \inf\{j : j \geq i, \quad Z_j \geq Y_j^*\}, \quad 0 \leq i \leq T.$$

   In particular,

   $$Y_i^* = E_i Z_{\tau_i^*}, \quad 0 \leq i \leq T,$$

   and the above family is the family of first optimal stopping times if several optimal stopping families exist.

The optimal stopping problem (1) has a natural interpretation in the point of view of the option holder: He seeks for an optimal exercise strategy which optimizes his expected pay-off. On the other hand, the seller of the option rather seeks for the minimal cash amount (smallest super-martingale) he has to have at hand in any case the holder of the option exercises.

3 Duality and surely optimal martingales

We briefly recall the dual approach proposed by Rogers (2001) and, independently, Haugh and Kogan (2004). The dual approach is based on the following observation. For any martingale (M_j) with $M_0 = 0$ we have,

$$Y_0^* = \sup_{\tau \in \{0,...,T\}} E_0 Z_\tau \leq \sup_{\tau \in \{0,...,T\}} E_0 (Z_\tau - M_\tau) \leq E_0 \max_{0 \leq j \leq T} (Z_j - M_j), \tag{2}$$

hence the right-hand side provides a (dual) upper bound for Y_0^*. Rogers (2001), and independently Haugh and Kogan (2004), showed that (2) holds with equality for the martingale part of the Doob decomposition of Y^*, i.e. $Y_j^* = Y_0^* + M_j^* - A_j^*$, where M^* is a martingale with $M_0^* = 0$, and A^* is predictable with $A_0^* = 0$. I.e.,

$$M_j^* = \sum_{l=1}^{j} (Y_l^* - E_{l-1} Y_l^*), \quad A_j^* = \sum_{l=1}^{j} \left(Y_{l-1}^* - E_{l-1} Y_l^*\right), \tag{3}$$

where A^* is non-decreasing due to the Bellman principle. In addition, they showed that

$$Y_0^* = \max_{0 \le j \le T} \left(Z_j - M_j^* \right) \quad \text{a.s.} \tag{4}$$

The next lemma, by Kolodko and Schoenmakers (2006), provides a somewhat more general class of super-martingales, which turn (2) into an equality such that moreover (4) holds.

Lemma 1 *Let S be a super-martingale with $S_0 = 0$, and such that $Z_j - Y_0^* \le S_j$, $1 \le j \le T$. It then holds,*

$$Y_0^* = \max_{0 \le j \le T} (Z_j - S_j) \quad a.s. \tag{5}$$

For the proof see Kolodko and Schoenmakers (2006).

Examples 2 *Obviously, by taking for S the Doob martingale (3), Lemma 1 applies. But, this is not the only one. For example, in the case $Z > 0$ a.s. we may also take*

$$S_j = (N_j^* - 1) Y_0^*,$$

where N^* is the multiplicative Doob part of the Snell envelope. I.e., $Y_j^* = Y_0^* N_j^* B_j^*$ for a martingale N^* with $N_0^* = 1$ and predictable B^* with $B_0^* = 1$. Hence

$$N_j^* = \prod_{l=1}^{j} \frac{Y_l^*}{E_{l-1} Y_l^*}, \quad B_j^* = \prod_{l=1}^{j} \frac{E_{l-1} Y_l^*}{Y_{l-1}^*}. \tag{6}$$

Indeed, since B^* is non-increasing due to the Bellman principle, we have

$$S_j = Y_0^* \left(\frac{Y_j^*}{Y_0^* B_j^*} - 1 \right) \ge Y_0^* \left(\frac{Y_j^*}{Y_0^*} - 1 \right) = Y_j^* - Y_0^* \ge Z_j - Y_0^*,$$

and so Lemma 1 applies again.

The multiplicative Doob decomposition in (6) is used by Jamshidian (2007) for constructing a multiplicative dual representation. In a comparative study, Chen and Glasserman (2007) pointed out however, that from a numerical point of view additive dual algorithms perform better due to the nice almost sure property (4).

Remark 3 *It is **not** true that for any martingale M which turns (2) into equality the almost sure statement (4) holds. As a simple counterexample consider $T = 1$, $Z_0 = 0$, $Z_1 = 2$, $M_0 = 0$, and $M_1 = \pm 1$ each with probability $1/2$. Indeed, we see that $Y_0^* = 2 = E_0(2 - M_1) = E_0 \max(0, 2 - M_1)$, but, $Y_0^* \ne \max(0, 2 - M_1)$ a.s.*

In order to have a unified dual representation for the Snell envelope Y_i^* at any i, it is convenient to drop the assumption that martingales start at zero. We then may restate the dual theorem as

$$Y_i^* = \inf_{M \in \mathcal{M}} E_i \max_{i \le j \le T} (Z_j - M_j + M_i) \tag{7}$$

$$= \max_{i \le j \le T} \left(Z_j - M_j^* + M_i^*\right) \quad \text{a.s.}, \tag{8}$$

for any i, $0 \le i \le T$, where $\mathcal{M}$ is the set of all martingales and M^* is the Doob martingale part of Y^*.

In view of Remark 3 and Examples 2, a martingale for which the infimum (7) is attained must not necessarily satisfy the almost sure property (8), and, martingales which do satisfy (8) are generally not unique.

Definition 4 *We say that a martingale M is **surely optimal** for the Snell envelope Y^* at a time i, $0 \le i \le T$, if (8) holds.*

Remark 5 *Obviously, the Doob martingale of Y^* is surely optimal at each i, $0 \le i \le T$, and any martingale M is trivially surely optimal at $i = T$. However, it is **not** true that sure optimality for some i with $i < T$ implies sure optimality at $i + 1$. As a counterexample let us consider $T = 2$, and $Z_0 = 4$, $Z_1 = 0$, $Z_2 = 2$. Take as martingale $M_0 = 0$, $M_1 = \pm 1$, each with probability $1/2$, and $M_2 = M_1 \pm 1$, each with probability $1/2$ conditional on M_1. Then $\max_{0 \le j \le 2} (Z_j - M_j + M_0) = 4$ a.s. Since we have trivially $Y_0^* = 4$, M is surely optimal at $i = 0$. But, $\max_{1 \le j \le 2} (Z_j - M_j + M_1) = 2 - M_2 + M_1 \notin \mathcal{F}_1$, so M is not surely optimal for Y^* at $i = 1$.*

4 Characterization of surely optimal martingales

In this section we give a unique characterization of martingales that are surely optimal for all $i = 0, \ldots, T$.

Theorem 6 *A martingale M with $M_0 = 0$ is surely optimal for $i = 0, \ldots, T$, if and only if there exists a sequence of adapted random variables $(\zeta_i)_{0 \le i \le T}$, such that $E_{i-1}\zeta_i = 1$, and $\zeta_i \ge 0$ for all $0 < i \le T$, and*

$$M_i = M_i^* - A_i^* + \sum_{l=1}^{i} \left(A_l^* - A_{l-1}^*\right) \zeta_i, \tag{9}$$

where, respectively, M^* is the Doob-martingale part and A_i^* the predictable part of the Snell envelope Y^*, given in (3).

Proof. i) Let us assume that M is surely optimal as stated. Then by (8) it holds for any $0 < i \le T$,

$$\begin{aligned} Y_{i-1}^* &= \max_{i-1 \le j \le T} (Z_j - M_j + M_{i-1}) \\ &= \max(Z_{i-1}, M_{i-1} - M_i + \max_{i \le j \le T} (Z_j - M_j + M_i)) \\ &= \max(Z_{i-1}, M_{i-1} - M_i + Y_i^*). \end{aligned} \tag{10}$$

Since always $Z_{i-1} \le Y^*_{i-1}$, and $Z_{i-1} < Y^*_{i-1}$ implies $A^*_{i-1} = A^*_i$, we obtain from (10),

$$\begin{aligned} Y^*_{i-1} - Z_{i-1} &= (M_{i-1} - M_i + Y^*_i - Z_{i-1})_+ \\ &= \left(M_{i-1} - M_i + M^*_i - M^*_{i-1} - A^*_i + A^*_{i-1} + Y^*_{i-1} - Z_{i-1}\right)_+ \\ &= 1_{Z_{i-1}<Y^*_{i-1}} \left(M_{i-1} - M_i + M^*_i - M^*_{i-1} + Y^*_{i-1} - Z_{i-1}\right)_+ \\ &+ 1_{Z_{i-1}=Y^*_{i-1}} \left(M_{i-1} - M_i + M^*_i - M^*_{i-1} - A^*_i + A^*_{i-1}\right)_+ . \end{aligned}$$

We so must have

$$\begin{gathered} 1_{Z_{i-1}<Y^*_{i-1}} \left(Y^*_{i-1} - Z_{i-1}\right) = \\ 1_{Z_{i-1}<Y^*_{i-1}} \left(M_{i-1} - M_i + M^*_i - M^*_{i-1} + Y^*_{i-1} - Z_{i-1}\right), \quad \text{and} \\ 1_{Z_{i-1}=Y^*_{i-1}} \left(M_{i-1} - M_i + M^*_i - M^*_{i-1} - A^*_i + A^*_{i-1}\right)_+ = 0, \end{gathered}$$

respectively. Hence we get

$$1_{Z_{i-1}<Y^*_{i-1}} \left(M_{i-1} - M_i + M^*_i - M^*_{i-1}\right) = 0, \quad \text{and} \tag{11}$$

$$1_{Z_{i-1}=Y^*_{i-1}} \left(M_{i-1} - M_i + M^*_i - M^*_{i-1} - A^*_i + A^*_{i-1}\right) = -1_{Z_{i-1}=Y^*_{i-1}}\mu_i, \tag{12}$$

for some non-negative $\mathcal{F}_i$-measurable random variable μ_i. W.l.o.g. we assume that $\mu_i \equiv 0$ on the set $\{Z_{i-1} < Y^*_{i-1}\}$. By taking $\mathcal{F}_{i-1}$ conditional expectations on both sides of (12), and using the martingale property of both M and M^*, and the predictability of A^*, it then follows that

$$E_{i-1}\mu_i = 1_{Z_{i-1}=Y^*_{i-1}} E_{i-1}\mu_i = 1_{Z_{i-1}=Y^*_{i-1}} \left(A^*_i - A^*_{i-1}\right). \tag{13}$$

In particular, since $\mu_i \ge 0$ almost surely, it follows from (13) that $\mu_i = 0$ on the set $\{A^*_i = A^*_{i-1}\}$ ($\supseteq \{Z_{i-1} < Y^*_{i-1}\}$). By next defining

$$\zeta_i := \begin{cases} \left(A^*_i - A^*_{i-1}\right)^{-1} \mu_i \quad \text{if} \quad A^*_i > A^*_{i-1}, \\ 1 \quad \text{elsewhere} \end{cases}, \tag{14}$$

we have $\mu_i = \left(A^*_i - A^*_{i-1}\right)\zeta_i$ a.s., and by (13) we have (using the convention $0 \cdot \infty = 0$)

$$\begin{aligned} E_{i-1}\zeta_i &= 1_{A^*_i > A^*_{i-1}} E_{i-1} \left(A^*_i - A^*_{i-1}\right)^{-1} \mu_i + 1_{A^*_i = A^*_{i-1}} \\ &= 1_{A^*_i > A^*_{i-1}} 1_{Z_{i-1}=Y^*_{i-1}} + 1_{A^*_i > A^*_{i-1}} 1_{Z_{i-1}<Y^*_{i-1}} + 1_{A^*_i = A^*_{i-1}} = 1, \end{aligned}$$

since the middle term is trivially zero. We thus obtain from (11) and (12),

$$M_{i-1} - M_i + M^*_i - M^*_{i-1} - A^*_i + A^*_{i-1} = -\left(A^*_i - A^*_{i-1}\right)\zeta_i,$$

from which (9) follows.
ii) Conversely, if a martingale M satisfies (9), we have for any $0 \leq i \leq T$,

$$\max_{i\leq j\leq T}\left(Z_j - M_j + M_i\right) = \max_{i\leq j\leq T}\Bigg(Z_j - M_j^* + A_j^* - \sum_{l=1}^{j}\left(A_l^* - A_{l-1}^*\right)\zeta_l$$
$$+M_i^* - A_i^* + \sum_{l=1}^{i}\left(A_l^* - A_{l-1}^*\right)\zeta_l\Bigg)$$
$$= Y_i^* + \max_{i\leq j\leq T}\left(Z_j - Y_j^* - \sum_{l=i+1}^{j}\left(A_l^* - A_{l-1}^*\right)\zeta_l\right) \leq Y_i^*,$$

and then by (7) the almost sure optimality follows. ■

By Theorem 6 we have immediately the next alternative characterization of almost sure martingales.

Corollary 7 *A martingale M with $M_0 = 0$ is surely optimal for $i = 0, \ldots, T$, if and only if there exists an non-decreasing adapted process N (which is not necessarily predictable!) with $N_0 = 0$, such that*

$$Y_i^* = Y_0^* + M_i - N_i.$$

Proof. If M is surely optimal as stated, we have by the "if" part of Theorem 6 (see (9)),

$$Y_i^* - Y_0^* - M_i = -\sum_{l=1}^{i}\left(A_l^* - A_{l-1}^*\right)\zeta_i =: -N_i, \tag{15}$$

with N being adapted, non-decreasing and $N_0 = 0$. Conversely, if

$$Y_i^* = Y_0^* + M_i - N_i$$

for some martingale M, $M_0 = 0$, and non-decreasing adapted N, $N_0 = 0$, we consider for each i, $0 \leq i \leq T$,

$$\max_{i\leq j\leq T}\left(Z_j - M_j + M_i\right) = \max_{i\leq j\leq T}\left(Z_j - Y_j^* - N_j + Y_i^* + N_i\right) \leq Y_i,^*$$

and then apply (7) again. ■

The following corollary is important in the context of the algorithms developed in Section 6.

Corollary 8 *Let the martingale M with $M_0 = 0$ be surely optimal for $i = 0, \ldots, T$. For the non-decreasing process N defined by (15) it then holds*

$$Y_i^* - M_i + M_{i-1} - Z_{i-1} = Y_{i-1}^* - N_i + N_{i-1} - Z_{i-1} =: U_i,$$

and since by (15), $N_i - N_{i-1} = \left(A_i^* - A_{i-1}^*\right)\zeta_i$ we thus have

$$\left(U_i\right)_+ = Y_{i-1}^* - Z_{i-1} \quad a.s.$$

So, in particular we have that $\left(U_i\right)_+$ is $\mathcal{F}_{i-1}$-measurable while U_i itself is generally ***not****, except for the case where $M = M^*$.*

From Theorem 6 it is clear that there exist infinitely many martingales which are surely optimal for all $i = 0, \dots, T$. In the following example we construct a (one-)parametric family of such martingales which includes the Doob martingale of the Snell envelope.

Example 9 *Let us assume $Z > 0$ a.s. (if Z is strictly bounded from below by a constant $-K$, we may consider the equivalent stopping problem due to $Z + K$). Then $Y^* > 0$ a.s., and for any α, $0 \le \alpha \le 1$, we consider*

$$\zeta_i := 1 - \alpha + \alpha \frac{Y_i^*}{E_{i-1} Y_i^*} = 1 - \alpha + \alpha \frac{N_l^*}{N_{l-1}^*},$$

where N^* is the martingale part of the multiplicative decomposition $Y_i^* = Y_0^* N_i^* B_i^*$ of the Snell envelope (see Examples 2). Obviously, it holds $E_{i-1}\zeta_i = 1$ and $\zeta_i \ge 0$, and so by Theorem 6 we obtain for each $0 \le \alpha \le 1$ a martingale

$$\begin{aligned} M_i &= M_i^* - A_i^* + \sum_{l=1}^{i} \left(A_l^* - A_{l-1}^* \right) \left(1 - \alpha + \alpha \frac{N_l^*}{N_{l-1}^*} \right) \\ &= M_i^* - \alpha A_i^* + \alpha \sum_{l=1}^{i} \left(A_l^* - A_{l-1}^* \right) \frac{N_l^*}{N_{l-1}^*}, \end{aligned}$$

which is surely optimal for $i = 0, ..., T$. Thus, for $\alpha = 0$ (i.e. $\zeta_i \equiv 1$) we retrieve the standard Doob martingale of the Snell envelope, and for $\alpha = 1$ we obtain

$$\begin{aligned} M_i &= Y_i^* - Y_0^* + \sum_{l=1}^{i} \left(A_l^* - A_{l-1}^* \right) \frac{N_l^*}{N_{l-1}^*} \\ &= \sum_{l=1}^{i} \left(Y_l^* - Y_{l-1}^* + Y_{l-1}^* \left(1 - \frac{B_l^*}{B_{l-1}^*} \right) \frac{N_l^*}{N_{l-1}^*} \right) \\ &= Y_0^* \sum_{l=1}^{i} \left(N_l^* B_l^* - N_{l-1}^* B_{l-1}^* + B_{l-1}^* \left(1 - \frac{B_l^*}{B_{l-1}^*} \right) N_l^* \right) \\ &= Y_0^* \sum_{l=1}^{i} B_{l-1}^* \left(N_l^* - N_{l-1}^* \right). \end{aligned} \tag{16}$$

Note that this martingale differs from the martingale $Y_0^* (N_i^* - 1)$ in Examples 2 (they would coincide after dropping the factors B_{l-1}^*). It is easy to show (using Theorem 6 again!) that the latter martingale is generally only optimal at $i = 0$, while the martingale (16) is surely optimal for all $i = 0, ..., T$, by construction.

The next theorem provides a key criterion for identifying surely optimal martingales.

Theorem 10 *Let Y^* be the Snell envelope of the cash-flow Z and let M be any martingale. Then, for every $i \in \{0, ..., T\}$ the following statement holds:*

For any set $\mathcal{A}_i \in \mathcal{F}_i$ we have

$$1_{\mathcal{A}_i} \max_{i \le j \le T} (Z_j - M_j + M_i) \in \mathcal{F}_i \quad \Longrightarrow \quad 1_{\mathcal{A}_i} \max_{i \le j \le T} (Z_j - M_j + M_i) = 1_{\mathcal{A}_i} Y_i^*.$$

Proof. We use backward induction on the number i. If $i = T$ the statement reads, for any $\mathcal{A}_T \in \mathcal{F}_T$ we have $1_{\mathcal{A}_T} Z_T \in \mathcal{F}_T \quad \Longrightarrow \quad 1_{\mathcal{A}_T} Z_T = 1_{\mathcal{A}_T} Y_T^*$, which is trivially true. Suppose the statement holds for some $i+1$ with $0 \le i < T$, and assume for an arbitrary but fixed set $\mathcal{A}_i \in \mathcal{F}_i$ that

$$1_{\mathcal{A}_i} \vartheta_i := 1_{\mathcal{A}_i} \max_{i \le j \le T} (Z_j - M_j + M_i) \in \mathcal{F}_i.$$

We then have

$$\begin{aligned} 1_{\mathcal{A}_i} \vartheta_i &= 1_{\mathcal{A}_i} \max(Z_i,\ M_i - M_{i+1} + \max_{i+1 \le j \le T} (Z_j - M_j + M_{i+1})) \\ &= 1_{\mathcal{A}_i} \max(Z_i,\ M_i - M_{i+1} + \vartheta_{i+1}) \in \mathcal{F}_i. \end{aligned}$$

We next consider the following $\mathcal{F}_i$ measurable events:

$$\begin{aligned} &\mathcal{A}_i \cap A^{\mathrm{i})} \ \text{ with } A^{\mathrm{i})} := \{\vartheta_i = Z_i\} \cap \{M_i - M_{i+1} + \vartheta_{i+1} \le Z_i\}, \quad \text{and} \\ &\mathcal{A}_i \cap A^{\mathrm{ii})} \ \text{ with } A^{\mathrm{ii})} = \Omega \backslash A^{\mathrm{i})} \\ &\qquad := \{\vartheta_i = M_i - M_{i+1} + \vartheta_{i+1}\} \cap \{Z_i < M_i - M_{i+1} + \vartheta_{i+1}\}. \end{aligned}$$

By taking $\mathcal{F}_i$-conditional expectations it follows that

$$1_{\mathcal{A}_i \cap A^{\mathrm{i})}} Z_i = E_i\ 1_{\mathcal{A}_i \cap A^{\mathrm{i})}} Z_i \ge E_i\ (M_i - M_{i+1} + \vartheta_{i+1})\, 1_{\mathcal{A}_i \cap A^{\mathrm{i})}} = 1_{\mathcal{A}_i \cap A^{\mathrm{i})}} E_i\ \vartheta_{i+1}. \tag{17}$$

Since M is a martingale, $E_{i+1}\ \vartheta_{i+1}$ is an upper bound for Y_{i+1}^* by (7), and we thus have,

$$E_i\ \vartheta_{i+1} = E_i E_{i+1}\ \vartheta_{i+1} \ge E_i Y_{i+1}^*,$$

which yields combined with (17),

$$1_{\mathcal{A}_i \cap A^{\mathrm{i})}} E_i Y_{i+1}^* \le 1_{\mathcal{A}_i \cap A^{\mathrm{i})}} Z_i,$$

and so

$$1_{\mathcal{A}_i \cap A^{\mathrm{i})}} \vartheta_i = 1_{\mathcal{A}_i \cap A^{\mathrm{i})}} Z_i = 1_{\mathcal{A}_i \cap A^{\mathrm{i})}} Y_i^*. \tag{18}$$

On the other hand, we notice that $1_{\mathcal{A}_i \cap A^{\mathrm{ii})}} \vartheta_{i+1} \in \mathcal{F}_{i+1}$, so we have by induction that $1_{\mathcal{A}_i \cap A^{\mathrm{ii})}} \vartheta_{i+1} = 1_{\mathcal{A}_i \cap A^{\mathrm{ii})}} Y_{i+1}^*$. It then follows that

$$\begin{aligned} 1_{\mathcal{A}_i \cap A^{\mathrm{ii})}} \vartheta_i &= E_i\, 1_{\mathcal{A}_i \cap A^{\mathrm{ii})}} \vartheta_i = E_i\, 1_{\mathcal{A}_i \cap A^{\mathrm{ii})}} (M_i - M_{i+1} + \vartheta_{i+1}) = 1_{\mathcal{A}_i \cap A^{\mathrm{ii})}} E_i \vartheta_{i+1} \\ &= E_i\, 1_{\mathcal{A}_i \cap A^{\mathrm{ii})}} \vartheta_{i+1} = E_i\, 1_{\mathcal{A}_i \cap A^{\mathrm{ii})}} Y_{i+1}^* = 1_{\mathcal{A}_i \cap A^{\mathrm{ii})}} E_i Y_{i+1}^*. \end{aligned} \tag{19}$$

Combining (18) and (19) finally yields $1_{\mathcal{A}_i} \vartheta_i = 1_{\mathcal{A}_i} Y_i^*$. ∎

As an immediate consequence of Theorem 10 we obtain:

Corollary 11 *Let Y^*, Z, and M be any martingale, as in Theorem 10. For $i \in \{0, ..., T\}$ it holds,*

$$\max_{i \le j \le T} (Z_j - M_j + M_i) \in \mathcal{F}_i \quad \Longrightarrow \quad M \text{ is surely optimal at } i.$$

5 Stability of surely optimal martingales

In equivalent terms, Corollary 11 states that, if a martingale M is such that the conditional variance of

$$\vartheta_i^{(M)} := \max_{i\le j\le T}\left(Z_j - M_j + M_i\right), \quad i = 0, ..., T,$$

is zero, i.e.

$$Var_i\left(\vartheta_i^{(M)}\right) := E_i\left(\vartheta_i - E_i\vartheta_i\right)^2 = 0, \text{ a.s.}, \ \ i = 0, ..., T,$$

then $\vartheta_i^{(M)} = Y_i^*$, for $i = 0, ..., T$. Hence the martingale M is surely optimal for $i = 0, ..., T$. In this section we present a stability result for martingales M which are, loosely speaking, near to be surely optimal in the sense that each i, $\mathrm{Var}_i\left(\vartheta_i^{(M)}\right)$ is small. More specifically, for a sequence of martingales $(M^{(n)})_{n\ge1}$ we provide mild conditions which guarantee that the corresponding upper bounds converge to the Snell envelope (although the sequence of martingales $(M^{(n)})$ does not need to converge itself). We have the following theorem.

Theorem 12 *Suppose that for each i, $0 \le i \le T$, $Var_i(\vartheta_i^{(n)}) \stackrel{P}{\to} 0$, if $n \to \infty$, where $\vartheta_i^{(n)} := \max_{i\le j\le T}\left(Z_j - M_j^{(n)} + M_i^{(n)}\right)$. In addition, suppose that for each i, the sequence of martingales $\left(M_i^{(n)}\right)_{n\ge1}$ is uniformly integrable for $n \ge 1$. It then holds $\left(\vartheta_i^{(n)}\right)_{n\ge1}$ is uniformly integrable for each i, and*

$$E_i\ \vartheta_i^{(n)} \stackrel{L_1}{\to} Y_i^*, \qquad i = 0, ..., T.$$

Proof. We will prove the theorem by backward induction on i. For $i = T$ there is nothing to prove. Suppose the theorem is proved for $i+1$, $i < T$. Let us consider

$$\begin{aligned}\vartheta_i^{(n)} - Z_i &:= \max_{i\le j\le T}\left(Z_j - M_j^{(n)} + M_i^{(n)}\right) - Z_i \\ &= \left(M_i^{(n)} - M_{i+1}^{(n)} + \max_{i+1\le j\le T}\left(Z_j - Z_i - M_j^{(n)} + M_{i+1}^{(n)}\right)\right)_+ \\ &= \left(\vartheta_{i+1}^{(n)} - Z_i + M_i^{(n)} - M_{i+1}^{(n)}\right)_+, \qquad (20)\end{aligned}$$

and define

$$\psi_i^{(n)} := \vartheta_{i+1}^{(n)} - Z_i + M_i^{(n)} - M_{i+1}^{(n)}. \qquad (21)$$

Due to the induction hypothesis $E_{i+1}\vartheta_{i+1}^{(n)} \stackrel{L_1}{\to} Y_{i+1}^*$ with $\vartheta_{i+1}^{(n)}$ being uniformly integrable. Thus, $\psi_i^{(n)}$ and $\vartheta_i^{(n)}$ are uniformly integrability due to (20), (21), and the uniform integrability of the martingales. So it holds,

$$E_i\,\psi_i^{(n)} = E_i\,E_{i+1}\vartheta_{i+1}^{(n)} - Z_i \stackrel{L_1}{\to} E_iY_{i+1}^* - Z_i. \qquad (22)$$

We will then show that

$$E_i\left(\psi_i^{(n)}\right)_+ \stackrel{L_1}{\to} \left(E_i Y_{i+1}^* - Z_i\right)_+, \tag{23}$$

which in turn implies by (20),

$$E_i\,\vartheta_i^{(n)} \stackrel{L_1}{\to} Z_i + \left(E_i Y_{i+1}^* - Z_i\right)_+ = \max\left(Z_i, E_i Y_{i+1}^*\right) = Y_i^*.$$

Since the family $E_i\left(\psi_i^{(n)}\right)_+$ is uniformly integrable too, it is enough to show that

$$E_i\left(\psi_i^{(n)}\right)_+ \stackrel{P}{\to} \left(E_i Y_{i+1}^* - Z_i\right)_+. \tag{24}$$

From (22) it follows

$$\left(E_i\,\psi_i^{(n)}\right)_+ \stackrel{P}{\to} \left(E_i Y_{i+1}^* - Z_i\right)_+,$$

so it is sufficient to prove that

$$E_i\left(\psi_i^{(n)}\right)_+ - \left(E_i\,\psi_i^{(n)}\right)_+ \stackrel{P}{\to} 0.$$

On the one side, by Jensen's inequality, we have

$$E_i\left(\psi_i^{(n)}\right)_+ \geq \left(E_i \psi_i^{(n)}\right)_+ \quad \text{a.s.}$$

Now take an arbitrary $\epsilon > 0$ and consider the inequality

$$\begin{aligned} &1_{\left\{E_i\left(\psi_i^{(n)}\right)_+ > \left(E_i\,\psi_i^{(n)}\right)_+ + \epsilon\right\}} 1_{\left\{\left(\psi_i^{(n)}\right)_+ \leq \left(E_i\,\psi_i^{(n)}\right)_+\right\}} \\ &\leq 1_{\left\{E_i\left(\psi_i^{(n)}\right)_+ > \left(E_i\,\psi_i^{(n)}\right)_+ + \epsilon\right\}} 1_{\left\{\left(\psi_i^{(n)}\right)_+ \leq E_i\left(\psi_i^{(n)}\right)_+ - \epsilon\right\}}. \end{aligned} \tag{25}$$

A conditional version of Chebyschev's inequality implies that

$$\begin{aligned} I_n &:= 1_{\left\{E_i\left(\psi_i^{(n)}\right)_+ > \left(E_i\,\psi_i^{(n)}\right)_+ + \epsilon\right\}} E_i\, 1_{\left\{\left(\psi_i^{(n)}\right)_+ \leq E_i\left(\psi_i^{(n)}\right)_+ - \epsilon\right\}} \\ &\leq 1_{\left\{E_i\left(\psi_i^{(n)}\right)_+ > \left(E_i\,\psi_i^{(n)}\right)_+ + \epsilon\right\}} \frac{\mathrm{Var}_i\left(\left(\psi_i^{(n)}\right)_+\right)}{\epsilon^2} \\ &= 1_{\left\{E_i\left(\psi_i^{(n)}\right)_+ > \left(E_i\,\psi_i^{(n)}\right)_+ + \epsilon\right\}} \frac{\mathrm{Var}_i\left(\vartheta_i^{(n)}\right)}{\epsilon^2} \stackrel{P}{\to} 0, \end{aligned}$$

by the assumptions of the Theorem. Since obviously $0 \leq I_n \leq 1$, this implies that $I_n \stackrel{L_1}{\to} 0$. Then note that (see (25))

$$\begin{aligned} &1_{\left\{E_i\left(\psi_i^{(n)}\right)_+ > \left(E_i\,\psi_i^{(n)}\right)_+ + \epsilon\right\}} 1_{\left\{\left(\psi_i^{(n)}\right)_- > 0\right\}} \\ &\leq 1_{\left\{E_i\,\psi_i^{(n)} > \left(E_i\,\psi_i^{(n)}\right)_+ + \epsilon\right\}} 1_{\left\{\left(\psi_i^{(n)}\right)_+ \leq \left(E_i\,\psi_i^{(n)}\right)_+\right\}}. \end{aligned}$$

As a consequence it follows that

$$0 \le E_i 1_{\left\{E_i\left(\psi_i^{(n)}\right)_+ > \left(E_i\,\psi_i^{(n)}\right)_+ + \epsilon\right\}} 1_{\left\{\left(\psi_i^{(n)}\right)_- > 0\right\}} \le I_n \xrightarrow{L_1} 0. \tag{26}$$

Now since the family $\left(\psi_i^{(n)}\right)_-$ is uniformly integrable also, it is not difficult to see that (26) implies

$$1_{\left\{E_i\left(\psi_i^{(n)}\right)_+ > \left(E_i\,\psi_i^{(n)}\right)_+ + \epsilon\right\}} E_i\left(\psi_i^{(n)}\right)_- \xrightarrow{L_1} 0. \tag{27}$$

Next we consider

$$\begin{aligned}
&1_{\left\{E_i\left(\psi_i^{(n)}\right)_+ > \left(E_i\,\psi_i^{(n)}\right)_+ + \epsilon\right\}} E_i\left(\psi_i^{(n)}\right)_- \\
&= 1_{\left\{E_i\left(\psi_i^{(n)}\right)_+ > \left(E_i\,\psi_i^{(n)}\right)_+ + \epsilon\right\}} \left[E_i\left(\psi_i^{(n)}\right)_+ - E_i\,\psi_i^{(n)}\right] \\
&\ge 1_{\left\{E_i\left(\psi_i^{(n)}\right)_+ > \left(E_i\,\psi_i^{(n)}\right)_+ + \epsilon\right\}} \left[\left(E_i\,\psi_i^{(n)}\right)_+ - E_i\,\psi_i^{(n)} + \epsilon\right] \\
&= 1_{\left\{E_i\left(\psi_i^{(n)}\right)_+ > \left(E_i\,\psi_i^{(n)}\right)_+ + \epsilon\right\}} \left[\left(E_i\,\psi_i^{(n)}\right)_- + \epsilon\right] \\
&\ge \epsilon 1_{\left\{E_i\left(\psi_i^{(n)}\right)_+ > \left(E_i\,\psi_i^{(n)}\right)_+ + \epsilon\right\}} \ge 0,
\end{aligned}$$

and so by (27),

$$P\left(E_i\left(\psi_i^{(n)}\right)_+ > \left(E_i\,\psi_i^{(n)}\right)_+ + \epsilon\right) \to 0.$$

■

The following simple example illustrates that Theorem 12 would not be true when the uniform integrability condition is dropped.

Example 13 *Take $T = 1$, $Z_0 = Z_1 = 0$, $M_0^{(n)} = 0$, $M_1^{(n)} =: -\xi_n$ with $E_0\xi_n = 0$, $n = 1, 2, \ldots$ Then obviously $Y_0^* = 0$, and we have*

$$\vartheta_0^{(n)} = \max(Z_0 - M_0^{(n)}, Z_1 - M_1^{(n)}) = \max(0, \xi^{(n)}) = \xi_+^{(n)}.$$

Now take

$$\xi^{(n)} = \begin{cases} 1 \textit{ with Prob. } \frac{n-1}{n} \\ 1-n \textit{ with Prob. } \frac{1}{n} \end{cases}$$

(hence $E_0\xi^{(n)} = 0$). Then, for $n \to \infty$ we have $Var_0(\vartheta_0^{(n)}) = E_0(\xi_+^{(n)})^2 - \left(E_0\xi_+^{(n)}\right)^2 = \frac{n-1}{n} - \left(\frac{n-1}{n}\right)^2 = \frac{n-1}{n^2} \to 0$, whereas $E_0\vartheta_0^{(n)} = E_0\xi_n^+ = \frac{n-1}{n} \to 1$. Clearly, for each $K > 1$, $E_0\left|M_1^{(n)}\right| 1_{\left\{\left|M_1^{(n)}\right| > K\right\}} \ge \frac{n-1}{n} 1_{\{n-1>K\}} \to 1$ as $n \to \infty$, hence the $\left(M_1^{(n)}\right)$ are not uniformly integrable.

Remark 14 *Theorem 12 is important in practical situations, for instance, if there exists some underlying (multi-dimensional) Markovian structure with respect to some (multi-dimensional) Wiener filtration. In this environment we may consider the following class of uniformly integrable martingales.*

Let X be a D-dimensional Markov process adapted to a filtration generated by an m-dimensional Brownian motion W and let the function $c(\cdot,\cdot):\mathbb{R}_{\geq 0}\times\mathbb{R}^D\to\mathbb{R}_{\geq 0}$ be such that $E_0\int_0^T c^2(s,X_s)ds<\infty$. Then the class $\mathcal{M}^{UI}$ of martingales defined by

$$M\in\mathcal{M}^{UI}:\Longleftrightarrow M_t=\int_0^t b^{\mathsf{T}}(s,X_s)dW_s,\qquad 0\leq t\leq T,\quad \text{for } b \text{ with } \quad |b|\leq c$$

is uniformly integrable due to the criterion of de la Vallée Poussin, since

$$\sup_{M\in\mathcal{M}^{UI}} E\left|M_t\right|^2\leq\int_0^T E_0c^2(s,X_s)ds<\infty,\quad 0\leq t\leq T.$$

Remark 15 *Consider any class of uniformly integrable martingales $\mathcal{M}^{UI}$, for example the one considered in Remark 14. As the topology of convergence in probability is metrizable by the Ky Fan Metric (e.g., see Dudley (2002))*

$$d_P(X,Y):=\inf\{\epsilon>0:\ P(|X-Y|>\epsilon)\leq\epsilon\},$$

one may restate Theorem 12 as follows. For any $\epsilon>0$ there exist a $\delta>0$ such that

$$\left[M\in\mathcal{M}^{UI}\quad\wedge\quad\max_{0\leq i\leq T} d_P(Var_i(\vartheta_i^{(M)}),0)<\delta\right]\Longrightarrow\max_{0\leq i\leq T}\left\|\vartheta_i^{(M)}-Y_i^*\right\|_{L_1}<\epsilon.$$

In view of Remark 15, Theorem 12 may be considered as a stability theorem related to the statement of Corollary 11.

6 Towards new dual algorithms for pricing of Bermudan derivatives

In this section we outline the design of new dual algorithms for solving multiple stopping problems, hence pricing bermudan products, which rely on the trust of Corollary11 and Theorem 12: For constructing a dual martingale $\widehat{M}$ that yields a tight upper bound it is sufficient to establish that the expected conditional variances

$$E\,\mathrm{Var}_i\left(\vartheta_i^{\widehat{M}}\right):=E\,\mathrm{Var}_i\left(\max_{i\leq j\leq T}(Z_j-\widehat{M}_j+\widehat{M}_i)\right),\quad i=0,...,T,$$

are close enough to zero. Note for Theorem 12 that $E\,\mathrm{Var}_i\left(\vartheta_i^{\widehat{M}}\right)\to 0$ implies $\mathrm{Var}_i\left(\vartheta_i^{\widehat{M}}\right)\overset{P}{\to}0$. Henceforth we assume an environment where there exists an underlying D-dimensional Markovian process X adapted to the filtration $\mathbb{F}$, such that the cash-flow process is of the form

$$Z_i = Z_i(X_i), \quad 0 \le i \le T.$$

As an immediate consequence, the Snell envelope is also of this form, i.e. $Y_i^* = Y_i^*(X_i), \quad 0 \le i \le T$. For a generic description of our approach that covers a variety of applications, we introduce a parametric system of adapted random variables in the following way. Let for each i, $1 \le i \le T$, I_i be a given (deterministic) index set, and let $\mathcal{I}_i := \otimes_{l=i}^T I_l$ be the formal (tensor) product set consisting of products of the form $\mathbf{a}_i := \otimes_{l=i}^T a_l \in \mathcal{I}_i$ with $a_l \in I_l$, $i \le l \le T$. We then consider a system of adapted random variables,

$$\xi_i^{\mathbf{a}_i} \in \mathcal{F}_i, \quad \mathbf{a}_i \in \mathcal{I}_i, \quad 0 \le i \le T, \quad \text{such that} \quad E_i \xi_{i+1}^{\mathbf{a}_i} = 0, \quad \text{if} \quad 0 \le i < T. \tag{28}$$

We moreover require that for any i, $1 \le i \le T$, we have that for any bounded measurable function $f_i : \mathbb{R}^{T-i+1} \times \mathbb{R}^{D\times(T-i+1)} \to \mathbb{R}$, and any sequence $(a_l)_{i\le l\le T}$ with $a_l \in I_l$, it holds

$$E_{i-1} f_i(\xi_i^{\mathbf{a}_i}, ..., \xi_T^{\mathbf{a}_T}, X_i, ..., X_T) = E_{X_{i-1}} f_i(\xi_i^{\mathbf{a}_i}, ..., \xi_T^{\mathbf{a}_T}, X_i, ..., X_T) \tag{29}$$

with $\mathbf{a}_j = \otimes_{l=j}^T a_l$, $i \le j \le T$. Next, for any sequence $(a_l)_{1\le l\le T}$ with $a_l \in I_l$, we consider the martingale

$$M_i^{\mathbf{a}} := \sum_{j=1}^{i} \xi_j^{\mathbf{a}_j}, \tag{30}$$

with $M_0^{\mathbf{a}} := 0$ and $\mathbf{a}_j = \otimes_{l=j}^T a_l$, $1 \le j \le T$. By requirement (29), a kind of Markov property, the simultaneous distribution of $X_i, ..., X_T$, and any martingale increment $M_j^{\mathbf{a}} - M_{i-1}^{\mathbf{a}}$, $j \ge i > 0$, given $\mathcal{F}_{i-1}$, is already determined by X_{i-1}. Obviously, any surely optimal martingale, for which the characterizing random sequence (ζ_i) in Theorem 6 fulfills (29), in particular the Doob martingale of the Snell envelope, has this property. It should also be noted that by construction (30), martingale increments

$$M_l^{\mathbf{a}} - M_{i-1}^{\mathbf{a}}, \ l \ge i > 0,$$

are already determined by the tail product $\mathbf{a}_i = \otimes_{l=i}^T a_l$. For our applications it is anticipated that, in a Monte Carlo simulation of X, the outcomes of the random variables (28) may be obtained in a tractable way (for example as closed form expressions in the underlying state variable) for any parameter choice, on each trajectory of X.

Canonical cases

Case I. Let us take $I_i = \mathbb{R}$ for each i, $1 \le i \le T$ and consider

$$\xi_i^{\mathbf{a}_i} := \sum_{k=i}^{T} a_k (E_i Z_k - E_{i-1} Z_k), \quad \text{where } \mathbf{a}_i := \otimes_{l=i}^T a_l.$$

We then obtain via (30)

$$M_i^{\mathbf{a}} = \sum_{k=1}^{T} a_k (E_i Z_k - E_0 Z_k), \quad 1 \le i \le T, \tag{31}$$

hence a linear combination of martingales induced by European options corresponding to the cash-flow process. In fact, this case is considered in Joshi and Theis (2002), in the context of Bermudan swaptions. The corresponding dual optimization problem

$$\widehat{\mathbf{a}} := \arg\inf_{\mathbf{a}\in\mathcal{I}_1} E \max_{0\le j\le T} \left(Z_j - M_j^{\mathbf{a}}\right),$$

with linearly structured $M^{\mathbf{a}}$ according to (31) is there solved by minimizing the usual Monte Carlo estimator for $E \max_{0\le j\le T}\left(Z_j - M_j^{\mathbf{a}}\right)$ on a fixed set of Monte Carlo trajectories, using a multi-dimensional search procedure for determining an optimal vector parameter $\widehat{\mathbf{a}} = \otimes_{l=1}^{T} \widehat{a}_l$. However, such multi-dimensional minimization problems are not always easy to solve, and in general it may be more effective to have an algorithm as developed in this section later on, where the sequence $(\widehat{a}_l)$ is constructed in a (backward) recursive way.

Case II. Let us now take $I_i = \mathbb{R}^{T-i+1}$ and consider

$$\xi_i^{\mathbf{a}_i} := \xi_i^{a_i} := \sum_{k=i}^{T} a_{ik}(E_i Z_k - E_{i-1} Z_k), \quad a_i = (a_{ii}, ..., a_{iT}) \in I_i, \qquad \mathbf{a}_i := a_i \otimes \mathbf{a}_{i+1},$$

hence, in this case $\xi_i^{\mathbf{a}_i}$ doesn't depend on $\mathbf{a}_{i+1}$. We then obtain via (30) the structure

$$M_i^{\mathbf{a}} = \sum_{k=1}^{T} \sum_{l=1}^{i\wedge k} a_{lk}(E_l Z_k - E_{l-1} Z_k),$$

which boils down to (31) under the restriction $a_{1k} = ... = a_{kk} =: a_k$, $k = 1, ..., T$.

Case III. In many environments one may construct a large class of sequences of "elementary" martingale increments $(\xi_i)_{1\le i\le T}$. We here consider an example environment which applies in many situations. Suppose the dynamics of X_t are given in continuous time for $t \in [0, T]$, by the unique strong solution of an Itô-Lévy SDE,

$$X_t = X_0 + \int_0^t \mathfrak{a}(s, X_s)ds + \int_0^t \mathfrak{b}(s, X_s)dW_s + \int_0^t \mathfrak{c}(s-, X_{s-})dH_s, \tag{32}$$

where the D-dimensional vector valued function $\mathfrak{a}$, the $D \times m$ matrix valued function $\mathfrak{b}$, and the $D \times q$ matrix valued function $\mathfrak{c}$, satisfy sufficient regularity conditions, W is a m-dimensional standard Wiener process, and

$$H_t := \int_0^t \int_{\mathbb{R}^q} u(\mu(ds, du) - F_s(du)ds), \quad 0 \le t \le T,$$

is a q-dimensional jump Lévy-martingale with Lévy measure $F_s(du)ds$ on $\mathbb{R}^q \times \mathbb{R}_+$, satisfying (the for notational convenience a little stronger than usual) integrability condition $\int_0^T \int_{\mathbb{R}^q} \left(\|u\| \wedge \|u\|^2 \right) F_s(du)ds < \infty$ (see for example Øksendal and Sulem (2004)). Calling upon the Brownian martingale representation theorem, and the martingale representation theorem of Kunita and Watanabe for Lévy processes (e.g. see Applebaum (2004) p- 266 and the references therein), we may construct "elementary" martingale increments of two types. Firstly, due to the Wiener part of (32) we may consider (as in Belomestny et. al. (2009)) for any given "elementary" (possibly time independent) basis function $\varphi(t,x) : \mathbb{R}_+ \times \mathbb{R}^D \to \mathbb{R}^m$,

$$\xi_i^W := \int_i^{i+1} \varphi^\top(s, X_s) dW_s. \tag{33}$$

Secondly, for any "elementary" (possibly time independent) basis function $\varphi^J(t,x,u) : \mathbb{R}_+ \times \mathbb{R}^D \times \mathbb{R}^q \to \mathbb{R}$, we may consider

$$\xi_i^J := \int_i^{i+1} \int_{\mathbb{R}^q} \varphi^J(s-, X_{s-}, u)(\mu(ds,du) - F_s(du)ds). \tag{34}$$

In order to guarantee that (33) and (34) define martingale increments indeed, the basis functions are assumed to satisfy sufficient integrability (e.g. Novikov) conditions. Now, for given sets of basis functions

$$\left(\varphi_k(t,x)\right)_{1 \le k \le K}, \quad \text{and} \quad \left(\varphi_k^J(t,x,u)\right)_{1 \le k \le K},$$

as in (33) and (34), respectively, we may take $I_i := \mathbb{R}^K \times \mathbb{R}^K$ for each $1 \le i \le T$, and consider

$$\xi_i^{\mathbf{a}_i} := \xi_i^{a_i} := \sum_{k=1}^K \beta_{i,k}^W \int_i^{i+1} \varphi_k^\top(s, X_s) dW_s \tag{35}$$
$$+ \sum_{k=1}^K \beta_{i,k}^J \int_i^{i+1} \varphi_k^J(s-, X_{s-}, u)(\mu(ds,du) - F_s(du)ds),$$

for any $a_i := (\beta_{i,1}^W, ..., \beta_{i,K}^W, \beta_{i,1}^J, ..., \beta_{i,K}^J) \in I_i$. As in case (II), $\xi_i^{\mathbf{a}_i}$ only depends on $a_i \in I_i$. On any trajectory $X_s = X_s(\omega)$, induced by a trajectory of $W_s = W_s(\omega)$ and $H_s = H_s(\omega)$ via (32), the integrals in (35) may be approximated by suitable numerical schemes. For instance the Wiener integrals my be computed by the standard Euler scheme

$$\int_i^{i+1} \varphi_k^\mathsf{T}(s, X_s) dW_s \approx \sum_{l=0}^{\delta^{-1}-1} \varphi_k^\top(i + l\delta, X_{i+l\delta}) \left(W_{i+(l+1)\delta} - W_{i+l\delta}\right), \tag{36}$$

for a small enough $\delta > 0$. For more sophisticated approximations of the Wiener integrals see Kloeden and Platen (1992), and for approximations of integrals involving Levy processes see Platen and Bruti-Liberati (2010) (we omit the details). The thus constructed family of martingale increments given by (35) clearly satisfies (28) and (29).

In view of Theorem 12, we are now going to design a generic dual backward algorithm for constructing an optimal (product) parameter $\widehat{\mathbf{a}}$ for the martingale $M^{\mathbf{a}}$ where $\mathbf{a} = \otimes_{j=1}^{T} a_j \in \mathcal{I}_1$, via minimizing backwardly, i.e. starting from $i = T$ down to $i = 0$, the expected conditional variances

$$\begin{aligned} E\operatorname{Var}_i\left(\vartheta_i^{\mathbf{a}_{i+1}}\right) &:= E\operatorname{Var}_i\left(\max_{i\leq j\leq T}(Z_j - M_j^{\mathbf{a}} + M_i^{\mathbf{a}})\right) \\ &= E\operatorname{Var}_{X_i}\left(\max_{i\leq j\leq T}(Z_j - \sum_{l=i+1}^{j}\xi_l^{\mathbf{a}_l})\right), \quad \mathbf{a}_l := \otimes_{r=l}^{T} a_r \end{aligned} \tag{37}$$

with $\mathbf{a}_{T+1}$ being defined as an empty product. Note that due to requirement (29), in (37) variance conditional on $\mathcal{F}_i$ coincides with variance conditional on X_i. For minimizing (37) we estimate the expected conditional variances on one fixed Monte Carlo sample of trajectories $(X_i^{(m)},\ 0 \leq i \leq T)_{m=1,\ldots,M}$, using a suitable estimation procedure, for instance as described in Section 6.1 below. Finally, the martingale $M^{\widehat{\mathbf{a}}}$ obtained at last may be used for estimating a price upper bound

$$\widetilde{Y}_0^{up} := \frac{1}{\widetilde{M}}\sum_{m=1}^{\widetilde{M}}\max_{0\leq j\leq T}\left(Z_j(\widetilde{X}_j^{(m)}) - \widetilde{M}_j^{\widehat{\mathbf{a}},(m)}\right) \tag{38}$$

using a second Monte Carlo simulation that generates $(\widetilde{X}_i^{(m)}, 0 \leq i \leq T)_{m=1,\ldots,\widetilde{M}}$, and where $\widetilde{M}_j^{\widehat{\mathbf{a}},(m)}$ is evaluated along each trajectory of this simulation.

The potential efficiency of the whole procedure is supported by the following remark.

Remark 16 *If the parametric family of martingales $(M^{\mathbf{a}})$ contains a martingale $M^{\widehat{\mathbf{a}}}$ that is "close to" to be surely optimal, the expected conditional variances (37) to be estimated due to the parameter choice $\widehat{\mathbf{a}}$ in the corresponding backward algorithm are "low". In general, estimating the variance of a random variable which actual variance is close to zero by Monte Carlo is most efficient in the sense that it requires only a relatively small sample size.*

Main lines of the algorithm

In a pseudo-algorithmic language, the procedure is as follows.

Step A1) At the initialization time $i = T$ we simply have $\vartheta_T = \widehat{\vartheta}_T = Z_T(X_T)$, hence $\operatorname{Var}_T\left(\widehat{\vartheta}_T\right) = 0$ almost surely, and formally $\widehat{\mathbf{a}}_{T+1}$ is initialized as an empty product.

Step A2) Now assume that for a fixed i, $0 < i \leq T$, an estimation of $\widehat{\mathbf{a}}_{i+1} = \otimes_{l=i+1}^{T} \widehat{a}_l$ is constructed, and that for each trajectory m, $m = 1, ..., M$, the pathwise maximum

$$\vartheta_i^{\widehat{\mathbf{a}}_{i+1},(m)} := \max_{i \leq j \leq T} (Z_j^{(m)} - \sum_{l=i+1}^{j} \xi_l^{\widehat{\mathbf{a}}_l,(m)})$$

(with $Z_j^{(m)} := Z_j(X_j^{(m)})$) is constructed and handed over. We are then going to estimate a solution of the minimization problem

$$\begin{aligned} a_i &:= \arg\inf_{a \in I_i} E \operatorname{Var}_{X_{i-1}} \left(\vartheta_{i-1}^{a \otimes \widehat{\mathbf{a}}_{i+1}} \right) \\ &= \arg\inf_{a \in I_i} E \operatorname{Var}_{X_{i-1}} \left(\max(Z_{i-1}, \vartheta_i^{\widehat{\mathbf{a}}_{i+1}} - \xi_i^{a \otimes \widehat{\mathbf{a}}_{i+1}} \right) \\ &= \arg\inf_{a \in I_i} E \operatorname{Var}_{X_{i-1}} \left(\vartheta_i^{\widehat{\mathbf{a}}_{i+1}} - \xi_i^{a \otimes \widehat{\mathbf{a}}_{i+1}} - Z_{i-1} \right)_+ \\ &=: \arg\inf_{a \in I_i} E \operatorname{Var}_{X_{i-1}} (U_i^a)_+ \end{aligned}$$

(as Z_{i-1} is X_{i-1}-measurable). Since for any random variable U it holds

$$\operatorname{Var}_{X_{i-1}} U_+ \leq \operatorname{Var}_{X_{i-1}} U, \tag{39}$$

we here have two options:

A2a) Estimate a solution of the minimization problem

$$\widehat{a}_i = \arg\inf_{a \in I_i} E \operatorname{Var}_{X_{i-1}} (U_i^a)_+ , \tag{40}$$

A2b) or estimate a solution of the minimization problem

$$\widehat{a}_i =: \arg\inf_{a \in I_i} E \operatorname{Var}_{X_{i-1}} (U_i^a) , \tag{41}$$

where, for example, estimation procedures for expected conditional variances described in Section 6.1 below may be used.

- ad A2a) Because of (39) the attainable minimum for the expected conditional variance in (41) is obviously always above (or equal to) the attainable minimum in (40). Let us imagine that we have at hand $\vartheta_i^{\widehat{\mathbf{a}}_{i+1}}$ exactly, i.e. $Y_i^* = \vartheta_i^{\widehat{\mathbf{a}}_{i+1}}$. From Corollary 8 we then see that, if our pool of martingales $(M^{\mathbf{a}})$ contains an almost sure one, but not the Doob-martingale of the Snell envelope M^*, it may happen that, while for a certain $\widehat{a} \in I_i$, in (40) $E \operatorname{Var}_{X_{i-1}} \left(U_i^{\widehat{a}} \right)_+$ is zero or "close" to zero, $E \operatorname{Var}_{X_{i-1}} (U_i^a)$ in (41) is not "close" to zero for any $a \in I_i$. In this situation we therefore just minimize (40) and proceed.

- ad A2b) If, on the other hand, the underlying class of martingales $(M^{\mathbf{a}})$ is rich enough in the sense that it contains martingales which are close to M^* (Doob martingale of the Snell envelope) in some sense, we may obtain a minimum in (41) that is "close" to zero for some $\widehat{a} \in I_i$, i.e. U_i^a is "near to be" X_{i-1}-measurable. So, as we see again from Corollary 8, the random variable

$$U_i^{\widehat{a}} + Z_{i-1} = \vartheta_i^{\widehat{\mathbf{a}}_{i+1}} - \xi_i^{\widehat{a} \otimes \widehat{\mathbf{a}}_{i+1}} = \max_{i \le j \le T} (Z_j - \sum_{l=i+1}^{j} \xi_l^{\widehat{\mathbf{a}}_l}) - \xi_i^{\widehat{a} \otimes \widehat{\mathbf{a}}_{i+1}}, \quad (42)$$

  can then be interpreted as an approximation to the continuation value $E_{X_{i-1}} Y_i^*$. If moreover the $\xi_i^{\widehat{a} \otimes \widehat{\mathbf{a}}_{i+1}}$ are linearly structured as in (35) for example, we show in Section 6.1 below how to include the determination of $\widehat{a}$ in a simultaneous regression procedure for the minimization problem (41), where the expected conditional variance is computed and minimized at the same time in fact. Then, this procedure also delivers as a "by-product" an estimation $C_{i-1}^{\widehat{a} \otimes \widehat{\mathbf{a}}_{i+1}}(x)$ (say) to the continuation functions $E_{X_{i-1}} Y_i^*$. This function can then be used afterwards for defining an exercise policy, hence constructing a lower bound (see (43)).

Step A3) Having estimated $\widehat{a}_i$ in the previous step we next set $\widehat{\mathbf{a}}_i = \widehat{a}_i \otimes \widehat{\mathbf{a}}_{i+1} = \otimes_{l=i}^{T} \widehat{a}_l$, and update

$$\vartheta_i^{\widehat{\mathbf{a}}_i,(m)} = \vartheta_{i-1}^{\widehat{a}_i \otimes \widehat{\mathbf{a}}_{i+1},(m)} = \max(Z_{i-1}^{(m)}, \vartheta_i^{\widehat{\mathbf{a}}_{i+1},(m)} - \xi_i^{\widehat{a}_i \otimes \widehat{\mathbf{a}}_{i+1}(m)})$$

for each trajectory $m = 1, ..., M$.

Step A4) After working all the way back we thus end up with an estimation $\widehat{\mathbf{a}} = \otimes_{l=1}^{T} \widehat{a}_l$, and a martingale $M^{\widehat{\mathbf{a}}}$, and finally compute an upper biased estimate of a price upper bound via (38) using a second simulation of X.

Remark 17 *Note that in canonical case (I), with the structure (31), we may set $\widehat{\mathbf{a}}_T = \widehat{a}_T = 1$ in advance, and each backward step (A2) involves the estimation of only one scalar quantity $\widehat{a}_i$.*

In case the above algorithm is carried out along the path A2b) with linearly structured $\xi_i^{\widehat{a} \otimes \widehat{\mathbf{a}}_{i+1}}$, using the regression procedure (47) outlined below, we may define an exercise policy

$$\tau_0 := \inf\{i : 0 \le i \le T,\ Z_i(X_i) \ge C_i^{\widehat{\mathbf{a}}_i}(X)\},$$

and simulate a lower biased price estimate,

$$Y_0^{low} \approx \frac{1}{\widetilde{M}} \sum_{\widetilde{m}=1}^{\widetilde{M}} Z_{\tau_0^{(\widetilde{m})}}(\widetilde{X}_{\tau_0^{(\widetilde{m})}}^{(\widetilde{m})}), \quad \text{where} \quad (43)$$
$$\tau_0^{(\widetilde{m})} = \inf\{i : 0 \le i \le T, \quad Z_i(\widetilde{X}_i^{(\widetilde{m})}) \ge C_i^{\widehat{\mathbf{a}}_i}(\widetilde{X}_i^{(\widetilde{m})})\}, \quad \widetilde{m} = 1, ..., \widetilde{M},$$

where the simulation sample in (38) may be used again.

6.1 Estimators for expected conditional variance

In this subsection we propose some usual procedures for estimating the expected conditional variances which appear in the minimization algorithm described above. In fact, the object to be estimated (for a fixed parameter) is

$$E\,\mathrm{Var}_X(U) := E\left[E_X U - (E_X U)^2\right] = E\,U^2 - E\,(E_X U)^2\,, \tag{44}$$

for a generic random vector $(X,U) \in \mathbb{R}^D \times \mathbb{R}$, where the estimation is going to be based on a Monte Carlo sample

$$(X^{(m)}, U^{(m)})_{m=1,\ldots,M}.$$

By (44) the only non-trivial issue is estimation of the term $E\,(E_X U)^2$.

Kernel estimation. One popular way to estimate conditional expectation is the use of a kernel estimator. We here give only a sketch based on a simply structured kernel, for a detailed treatment of kernel estimators we refer to the literature (e.g. Liero (1989)). For a given kernel $\Phi(x)$ on $\mathbb{R}^D$ with compact support, $\Phi(x) = \Phi(-x)$, $\Phi \geq 0$, and $\int \Phi(x)dx = 1$, and a suitably chosen band-with parameter $h > 0$, we may consider the estimator

$$E_{X=x} U \approx \frac{\sum_{m=1}^M U^{(m)} \Phi(\frac{X^{(m)}-x}{h})}{\sum_{m=1}^M \Phi(\frac{X^{(m)}-x}{h})} = \overline{U}(x),$$

and then estimate

$$E\,\mathrm{Var}_X(U) \approx \widehat{s}^{(M)}_{E\,\mathrm{Var}_X(U)} := \frac{1}{M}\sum_{m=1}^M \left(U^{(m)}\right)^2 - \frac{1}{M}\sum_{m=1}^M \left(\overline{U}(X^{(m)})\right)^2. \tag{45}$$

It is known that, due to the fact that the kernel has compact support, computation of (45) requires computation time at most proportional to $M \log^D M$ (see e.g. Greengard and Strain (1991) for details). This computation time must be considered par time slice i and par parameter trial a for the minimization problem

$$\widehat{a} =: \arg\inf_{a\in I} \widehat{s}^{(M)}_{E\,\mathrm{Var}_X(U_a)},$$

where, for example, I may stand for I_i, and U_a for $(U_i^a)_+$ or (U_i^a) in (40) or (41) respectively. The kernel estimation method is usually effective when the index set I_i is one-dimensional. Then typically only a few number of trials will be needed for a good estimate $\widehat{a}_i$. Moreover, the objects $\Phi(\frac{X^{(m)}-X^{(m')}}{h})$ may be stored in advance of the search procedure, as they are independent of the trial parameters.

Global linear regression. As another (maybe even more) popular approach we recap the global regression method. In this method one represents an estimation $\overline{U}(x)$ for $E_{X=x}U$ by a linear combination

$$\overline{U}(x) =: \sum_{k=1}^{K} \overline{\gamma}_k \psi_k(x)$$

of a family of (real valued) basis functions $x \in \mathbb{R}^D \to \psi_k(x)$, $k = 1, ..., K$, where $\overline{\gamma}$ solves the standard linear regression problem

$$\overline{\gamma} := \underset{\gamma \in \mathbb{R}^K}{\arg\inf} \sum_{m=1}^{M} \left(U^{(m)} - \sum_{k=1}^{K} \gamma_k \psi_k(X^{(m)}) \right)^2 .$$

The expected conditional variance is then estimated by

$$E \operatorname{Var}_X(U) \approx \sum_{m=1}^{M} \left(U^{(m)} - \sum_{k=1}^{K} \overline{\gamma}_k \psi_k(X^{(m)}) \right)^2 . \tag{46}$$

The regression method is most powerful when in addition U_a is linearly structured in the form

$$U_a = U^{(0)} + \sum_{l=1}^{L} a_l U^{(l)}, \quad a \in I := \mathbb{R}^L,$$

and $U^{(l)}$ are certain given "elementary"random variables (as in (42)). In this case, the estimation of the expected conditional variance (46), and its minimization over the parameter set can be done in one and the same linear regression procedure(!),

$$(\widehat{a}, \overline{\gamma}) = \underset{a \in \mathbb{R}^L,\, \gamma \in \mathbb{R}^K}{\arg\inf} \sum_{m=1}^{M} \left(U^{(0),(m)} + \sum_{l=1}^{L} a_l U^{(l),(m)} - \sum_{k=1}^{K} \gamma_k \psi_k(X^{(m)}) \right)^2 . \tag{47}$$

When we are in an environment as described in the important Canonical Case (III), the regression (47) may be effectively applied to the problem (41), where U_a stands for $\vartheta_i^{\widehat{\mathbf{a}}_{i+1}} - \xi_i^{a_i} - Z_{i-1}$, and $\xi_i^{a_i} = \xi_i^{\widehat{a}_i \otimes \widehat{\mathbf{a}}_{i+1}}$ is of the form (35). Moreover, the functions

$$C_i^{\widehat{\mathbf{a}}_i}(x) = \sum_{k=1}^{K} \overline{\gamma}_{i,k} \psi_k(x),$$

so obtained on the flight, can be used to construct a lower bound, see (43). In the next section we will present a numerical example in this context.

Remark 18 *We expect that the convergence of the kernel based and regression based algorithms presented in this section can be proved in the spirit of Belomestny et. al. (2010), where related methods are studied in the context of optimal control, see also Clement et al. (2002).*

7 Numerical examples

In this section we consider an algorithm for a Wiener environment (see case III) that is based on step A2b) and the global regression method described in the previous section. It is tested at the two benchmark examples, Bermudan basket-put on 5 assets and Bermudan max-call on 2 and 5 assets, which are also considered in Bender et al. (2006a) and Belomestny et. al. (2009), respectively. In both examples, the risk-neutral dynamic of each asset is governed by

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

where D is the number of assets, W_t^d, $d = 1, ..., D$, are independent one-dimensional Brownian motions (hence $m = D$ in case III), and r, δ and σ are constants. Exercise opportunities are equally spaced at times $T_j = \frac{jT}{J}$, $j = 0, ..., J$. The discounted payoff from exercise at time t is given by

$$Z_t(X_t) = e^{-rt}(K - \frac{X_t^1 + \ldots + X_t^D}{D})^+ \quad \text{for the Bermudan basket-put,}$$

and

$$Z_t(X_t) = e^{-rt}(\max(X_t^1, \ldots, X_t^D) - K)^+ \quad \text{for the Bermudan max-call,}$$

where $X_t = (X_t^1, \ldots, X_t^D)$. For each numerical implementation, the time interval $[T_j, T_{j+1}]$, $j = 0, \ldots, J-1$, is partitioned into L equal subintervals of width $\Delta t = \frac{T}{N}$ with $N = J \times L$. The numerical procedure can be described briefly as follows. We first simulate M independent samples of Brownian increments $\{(\Delta W_i^{1,(m)}, \ldots, \Delta W_i^{D,(m)}),\ i = 1, \ldots, N\}$, $m = 1, \ldots, M$. Then the trajectories of $X_i^{(m)} = (X_i^{1,(m)}, \ldots, X_i^{D,(m)})$, $i = 1, \ldots, N$, $m = 1, \ldots, M$ are given by

$$X_i^{d,(m)} = X_{i-1}^{d,(m)} \exp\{(r - \delta - \frac{1}{2}\sigma^2)\Delta t + \sigma \Delta W_i^{d,(m)}\}, \tag{48}$$

for $d = 1, \ldots, D$ and initial data X_0. We next carry out the algorithm from Section 6, along step A2b). In the spirit of (47) we solve the regression problem

$$(\widehat{a}_i, \overline{\gamma}_i) := \underset{(\beta,\gamma)}{\arg\min} \sum_{m=1}^{M} \left[\vartheta_i^{\widehat{\mathbf{a}}_{i+1}(m)} - \sum_{k=1}^{K} \beta_k \int_i^{i+1} \varphi_k^{\top}(u, X_u^{(m)}) dW_u^{(m)} \right. \\ \left. - \sum_{k=1}^{K_1} \gamma_k \psi_k(i, X_i^{(m)}) \right]^2, \tag{49}$$

for basis functions $(\varphi_k) = \left(\varphi_k^{(d)}\right)$ with $\varphi_k^{(d)} = \varphi_k^{(1)}$ for any d, and (ψ_k), chosen as explained below. In (49) the Wiener integrals are approximated by the standard Euler scheme (see (36)), using the same Brownian increments as in (48).

As one may expect, the choice of basis functions is crucial to obtain tight upper and lower bounds. In this respect, special information on the pricing problem may help us finding suitable basis functions. Suppose that $E_t\left(Z_T(X_T)\right) = f(t, X_t)$ for $0 \leq t \leq T = T_J$. Then, by Itô's formula and the fact that $E_t(Z_T)$ is a martingale we have

$$Z_T(X_T) - E_{T_{J-1}}\left(Z_T(X_T)\right) = \sum_{d=1}^{D} \sigma \int_{T_{J-1}}^{T} f_{x^d}(t, X_t) X_t^d dW_t^d.$$

Recall that $\widehat{\vartheta}_T = Z_T$ and $E_{T_{J-1}}\left(Z_T(X_T)\right)$ can be expressed in the following form

$$E_{T_{J-1}}\left(Z_T(X_T)\right) = e^{-rT_{J-1}} EP(T_{J-1}, X_{T_{J-1}}; T),$$

where $EP(t, x; T)$ is the price of the corresponding European option with maturity T at time t. Thus, it is natural to choose from time T to time T_{J-1} European option values for the basis $(\psi_k(t, x))$ and the corresponding European deltas multiplied by the value of the underlying asset for the basis $(\varphi_k(t, x))$. Although for the following steps ($t < T_{J-1}$) there is no easy way to predict optimal choices of (ψ_k) and (φ_k), the above analysis suggests us to always include the still-alive European options into the basis (ψ_k) and include the information on the European deltas into the basis (φ_k). In fact, based on similar arguments, these choices of basis functions were already proposed in Belomestny et. al. (2009).

We next carry out Step A3) and work all the way back, to obtain estimations of the coefficients $(\widetilde{a}_i, \overline{\gamma}_i)$, $i = 1, ..., T$. Finally by a new independent simulation we estimate an upper bound Y_0^{up} as described in step A4) and a lower bounds Y_0^{low} via (43).

7.1 Bermudan basket-put

In this example, we take the following parameter values,

$$r = 0.05, \quad \delta = 0, \quad \sigma = 0.2, \quad D = 5, \quad T = 3,$$

and

$$X_0^1 = \ldots = X_0^D = x_0, \quad K = 100.$$

For $T_j \leq t < T_{j+1}$, $j = 0, \ldots, J-1$, we choose the set $\{1, Pol_3(X_t), Pol_3(EP(t, X_t; T_{j+1})), Pol_3(EP(t, X_t; T_J))\}$ as basis functions (ψ_k), where $Pol_n(y)$ denotes the set of polynomials of degree up to n in the components of a vector y and $EP(t, X; T)$ denotes the (approximated) value of a European basket-put with maturity T at time t. Further we use $\{1, X_t^d \frac{\partial EP(t, X_t; T_{j+1})}{\partial X_t^d}, X_t^d \frac{\partial EP(t, X_t; T_J)}{\partial X_t^d}, d = 1, \ldots, D\}$ as basis (φ_k). Since there is no closed-form formula for the still-alive European basket-put, we use the moment-matching method to approximate their values (e.g., see Brigo et al. (2004), and Lord (2006)). Let $S_t = \dfrac{X_t^1 + \ldots + X_t^D}{D}$, and consider another asset G_t whose risk-neutral dynamic follows

$$dG_t = rG_t dt + \tilde{\sigma} G_t dW_t^1,$$

where $\tilde{\sigma}$ is a constant. The value of the European put on this asset can be easily computed by the well-known Black-Scholes formula, that is,

$$E[e^{-rT}(K - G_T)^+] = BS(G_0, r, \tilde{\sigma}, K, T). \tag{50}$$

If S_T and G_T have the same moments up to two, then the Black-Scholes price in (50) can be regarded as a good approximation for the value of the European basket-put $E\left(e^{-rT}(K - S_T)^+\right)$. Since

$$E(S_T) = \frac{1}{D}\sum_{d=1}^{D} X_0^d e^{rT},$$

$$E(S_T^2) = \frac{1}{D^2} e^{2rT}\left(\sum_{i,j=1}^{D} X_0^i X_0^j \exp(1_{i=j}\sigma^2 T)\right)$$

and

$$E(G_T) = G_0 e^{rT}, \qquad E(G_T^2) = G_0^2 e^{2rT+\tilde{\sigma}^2 T},$$

we can simply set

$$G_0 = \frac{1}{D}\sum_{d=1}^{D} X_0^d$$

and

$$\tilde{\sigma}^2 = \frac{1}{T}\ln\left(\frac{1}{(\sum_{d=1}^{D} X_0^d)^2}\sum_{i,j=1}^{D} X_0^i X_0^j \exp(1_{i=j}\sigma^2 T)\right).$$

The European deltas can be approximated by

$$\frac{\partial BS}{\partial G_0}\frac{\partial G_0}{\partial X_0^d} = -\mathcal{N}(-d_1)\frac{1}{D}, \quad d = 1, \ldots, D,$$

where $d_1 = \dfrac{\ln(\frac{G_0}{K}) + (r + \frac{\sigma^2}{2})T}{\tilde{\sigma}\sqrt{T}}$ and $\mathcal{N}$ denotes the cumulative standard normal distribution function.

The numerical results are shown in Table 1. 100000 simulations are used for the regression procedure and 100000 simulations for computing the upper and lower bounds. Values in parentheses are the standard errors. The last column in the table shows the lower and upper bounds obtained in Bender et al. (2006a).

7.2 Bermudan max-call

We use the same parameter values as in Section 7.1 except $\delta = 0.1$ and $D = 2$ or 5. As in the previous example we use European (call) options in the basis (ψ_k) and the corresponding deltas in the basis (φ_k). The value of the European

Table 1: Lower and upper bounds for Bermudan basket-put on 5 assets with parameters $r = 0.05$, $\delta = 0$, $\sigma = 0.2$, $K = 100$, $T = 3$ and different J and x_0

J	x_0	Lower Bound (SD)	Upper Bound (SD)	BKS Price Interval
3	90	10.0000 (0.0000)	10.0000 (0.0000)	[10.000,10.004]
	100	2.1649 (0.0119)	2.1817 (0.0015)	[2.154,2.164]
	110	0.5357 (0.0060)	0.5584 (0.0008)	[0.535,0.540]
6	90	10.0000 (0.0000)	10.0007 (0.0001)	[10.000,10.000]
	100	2.3986 (0.0112)	2.4336 (0.0013)	[2.359,2.412]
	110	0.5870 (0.0061)	0.5994 (0.0007)	[0.569,0.580]
9	90	10.0000 (0.0000)	10.0024 (0.0001)	[10.000,10.005]
	100	2.4862 (0.0109)	2.5197 (0.0012)	[2.385,2.502]
	110	0.6006 (0.0060)	0.6164 (0.0006)	[0.577,0.600]

max-call option is computed by the following formula (Johnson (1987)),

$$\sum_{l=1}^{D} X_0^l \frac{e^{-\delta T}}{\sqrt{2\pi}} \int_{(-\infty,d_+^l]} \exp[-\frac{1}{2}z^2] \prod_{\substack{l'=1\\ l'\neq l}}^{D} \mathcal{N}\left(\frac{\ln \frac{X_0^l}{X_0^{l'}}}{\sigma\sqrt{T}} - z + \sigma\sqrt{T}\right) dz$$

$$- Ke^{-rT} + Ke^{-rT}\prod_{l=1}^{D}\left(1 - \mathcal{N}\left(d_-^l\right)\right),$$

where

$$d_-^l := \frac{\ln \frac{X_0^l}{K} + (r - \delta - \frac{\sigma^2}{2})T}{\sigma\sqrt{T}}, \quad d_+^l = d_-^l + \sigma\sqrt{T}.$$

We use the central difference quote $\dfrac{f(x+\frac{h}{2}) - f(x-\frac{h}{2})}{h}$ (or the forward difference quote $\dfrac{f(x+h) - f(x)}{h}$) to approximate the European deltas. Note that the rounding errors resulting from too small values of h may give totally different basis function. So special attentions need to be paid to the choice of a suitable h. In our numerical experiment, we choose $h = 0.1$.

The numerical results are shown in Table 2. They are based on 1000 simulations for the regression procedure, 1000 simulations for computing the upper bound, 100000 simulations for computing the lower bound. The price intervals in the last column are quoted from Andersen and Broadie (2004).

Concluding remark

The numerical results presented in Tables 1,2 due to our new algorithm may be considered satisfactory given the required low computation times, which are in the order of minutes (in a C++ compiled implementation). In this respect it should be noted that computing upper bounds (in a rather generic way) in the order of minutes is quite fast compared to Bender et al. (2006a), Table 1,

Table 2: Lower and upper bounds for Bermudan max-call with parameters $r = 0.05$, $\delta = 0.1$, $\sigma = 0.2$, $K = 100$, $T = 3$ and different D and x_0

D	x_0	Lower Bound (SD)	Upper Bound (SD)	A&B Price Interval
	90	8.1027 (0.0380)	8.1369 (0.0379)	[8.053, 8.082]
2	100	13.8049 (0.0475)	14.0501 (0.0467)	[13.892, 13.934]
	110	21.3208 (0.0556)	21.4860 (0.0531)	[21.316, 21.359]
	90	16.6026 (0.0516)	16.7735 (0.0609)	[16.602,16.655]
5	100	26.0786 (0.0613)	26.4854 (0.0763)	[26.109,26.292]
	110	36.6561 (0.0693)	37.0723 (0.0844)	[36.704,36.832]

and Andersen and Broadie (2004), Table 2, which are computed with nested Monte Carlo simulation requiring much more computation time. Moreover, the algorithm delivers very fast and surprisingly good lower bounds while the upper bounds are comparable with the ones obtained with the algorithm in Belomestny et. al. (2009). Needless to say that, as for the method of Belomestny et. al. (2009), the performance of the here presented algorithm will highly depend on the choice of the basis functions. An in depth treatment of this issue is considered beyond scope however.

Acknowledgements

J.S. is grateful to Denis Belomestny for interesting discussions on this topic, and to Christian Bender for an inspiring remark, which led to counterexample Example 13.

References

L. Andersen, M. Broadie (2004). A primal-dual simulation algorithm for pricing multidimensional American options. Management Sciences, **50**, No. 9, 1222-1234.

D. Applebaum (2004). Lévy processes and stochastic calculus. Cambridge University Press.

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

D. Belomestny, A. Kolodko, and J. Schoenmakers (2010). Regression Methods for Stochastic Control Problems and their Convergence Analysis. SIAM J. Control Optim., Vol. 48, No. 5, pp. 3562–3588.

D. Belomestny, C. Bender, and J. Schoenmakers (2009). True upper bounds for Bermudam products via non-nested Monte Carlo. Mathematical Finance, **19**, No. 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.

C. Bender, A. Kolodko, J. Schoenmakers (2006). Iterating cancellable snowballs and related exotics. RISK, September 2006 pp. 126–130.

C. Bender, A. Kolodko, J. Schoenmakers (2006a). Policy iteration for american options: overview. Monte Carlo Methods and Appl., **12**, No. 5-6, pp. 347–362.

D. Brigo, F. Mercurio, F. Rapisarda, and R. Scotti (2004). Approximated moment-matching dynamics for basket-options pricing. Quantitative Finance, 4, 1–16.

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.

E. Clément, D. Lamberton, P. Protter (2002). An analysis of a least squares regression algorithm for American option pricing. Finance and Stochastics, **6**, 449-471.

V. Desai, V. Farias, C. Moallemi (2010). Pathwise Optimization for Optimal Stopping Problems. Working paper.

R.M. Dudley (2002). Real analysis and probability. Cambridge University Press.

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

L. Greengard, J. Strain (1991). The fast Gauss transform. SIAM J. Sci. Stat. Comp., **1**2, No. 1, 79-94.

M. Haugh, L. Kogan (2004). Pricing American options: a duality approach. Operations Research, **52**, No. 2, 258–270.

F. Jamshidian (2007). The duality of optimal exercise and domineering claims: A Doob-Meyer decomposition approach to the Snell envelope. Stochastics, **79**, No. 1–2, 27–60.

H. Johnson (1987). Options on the maximum or the minimum of several assets. Journal of Financial and Quantitative Analysis 22, 227–83.

M.S. Joshi, J. Theis (2002). Bounding Bermudan swaptions in a swap-rate market model. Quantitative Finance, **2**(5), 370-377.

P.E. Kloeden, E. Platen (1992). Numerical solution of stochastic differential equations. Springer Verlag Berlin.

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.

H. Liero (1989). Strong Uniform Consistency of Nonparametric Regression Function Estimates. Probab. Th. Rel. Fields, **82**, 587–614.

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

R. Lord (2005). Partially exact and bounded approximations for arithmetic Asian options. Journal of Computational Finance, **10**, no. 2, pp. 1-52.

J. Neveu (1975). Discrete Parameter Martingales. North-Holland, Amsterdam.

B. Øksendal and A. Sulem (2004). Applied Stochastic Control of Jump Diffusions. Springer.

E. Platen, N. Bruti-Liberati (2010). Numerical Solution of Stochastic Differential Equations with Jumps in Finance, Springer, Germany.

L.C.G. Rogers (2001). Monte Carlo valuation of American options. Mathematical Finance, **12**, 271–286.

L.C.G. Rogers (2010). Dual Valuation and Hedging of Bermudan Options. SIAM J. Financial Math., **1**, 604–608.

J. Schoenmakers (2005). Robust Libor Modelling and Pricing of Derivative Products. Chapman & Hall/CRC.

J. Schoenmakers, J. Huang (2010). Optimal dual martingales and their stability; fast evaluation of Bermudan products via dual backward regression. WIAS Preprint 1574.

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