Weierstraß-Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint ISSN 0946 – 8633

Sensitivities for Bermudan options by regression methods

Denis Belomestny, [1] Grigori N. Milstein [2] and John Schoenmakers [3]

submitted: 26th July 2007

[1] Weierstrass Institute
for Applied Analysis and Stochastics,
Mohrenstr. 39, 10117
Berlin, Germany
E-Mail: belomest@wias-berlin.de

[2] Department of Mathematics,
Ural State University,
Lenin Str. 51, 620083
Ekaterinburg, Russia
E-Mail: Grigori.Milstein@usu.ru

[3] Weierstrass Institute
for Applied Analysis and Stochastics,
Mohrenstr. 39, 10117
Berlin, Germany
E-Mail: schoenma@wias-berlin.de

No. 1247
Berlin 2007

2000 *Mathematics Subject Classification.* 60H30, 65C05, 91B28 .

Key words and phrases. American and Bermudan options, Optimal stopping times, Monte Carlo simulation, Deltas, Conditional probabilistic representations, Regression methods .

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Mohrenstraße 39
10117 Berlin
Germany

Fax: + 49 30 2044975
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

ABSTRACT. In this article we propose several pathwise and finite difference based methods for calculating sensitivities of Bermudan options using regression methods and Monte Carlo simulation. These methods rely on conditional probabilistic representations which allow, in combination with a regression approach, for efficient simultaneous computation of sensitivities at many initial positions. Assuming that the price of a Bermudan option can be evaluated sufficiently accurate, we develop a method for constructing deltas based on least squares. We finally propose a testing procedure for assessing the performance of the developed methods.

1. INTRODUCTION

Valuation of high-dimensional American/Bermudan options is one of the most difficult numerical problems in financial engineering. Besides its practical relevance, investigations in this field are of great theoretical importance because pricing of such options is an archetype for high-dimensional optimal stopping problems. Recently several approaches have been proposed for pricing of American options using Monte Carlo simulation technique and regression methods (see [1]-[12], [15], [20], [22], [25], [30], [31], [34] and references therein). The problem of efficient evaluation of price sensitivities however is considerably more involved. Even for European options this problem remains to be of great interest both in theory and practice. The simplest methods of evaluating price sensitivities for European options are based on finite difference approximations using Monte Carlo simulation. A shortcoming of these methods is a rather poor accuracy. In [27], finite difference based methods are essentially improved regarding accuracy and efficiency. Many articles are devoted to pathwise methods of estimating Greeks (see [15] and references therein, see [26] as well). In comparison with finite difference based methods, these methods are more accurate but less universal and they require rather large computational expenses. In contrast to European options there are only a few articles devoted to Greeks for American/Bermudan derivatives ([29], [13], [19], and references therein).

In this paper we propose a number of pathwise and finite difference based methods for computing sensitivities of Bermudan options via regression methods and Monte Carlo simulation. We basically assume that we have a pricing method for a Bermudan option which is accurate enough, and then construct a convenient least squares based method for computing its sensitivities (deltas). Our main goal is the development of methods which allow for evaluating sensitivities at any point in time using a single set of Monte Carlo simulated trajectories. The presented approach is related to [28] where conditional probabilistic representations are introduced for sensitivities of diffusion processes. There the sensitivities are estimated via a regression approach and used for variance reduced Monte Carlo simulation of diffusions. In [2] a regression based martingale estimator for dual upper bounds of Bermudan products is developed. The martingale estimator in [2], which essentially relies on a Clarc-Ocone type formula, implicitly allows for estimating sensitivities. So, the methods relying on conditional probabilistic representations in connection with regression play as such a decisive role in the problem of efficient estimating sensitivities.

Computing prices and sensitivities of Bermudan options involves a number of estimation errors of different nature which are hard to treat on a theoretical base. As an alternative we propose some kind of testing procedure for the application of the developed methods. Extensive numerical investigations of the here introduced approaches will be considered in a separate work.

In Section 2 we introduce our modelling framework and recall some known facts on pricing of American/Bermudan derivatives in a form suitable for our purposes. Section 3 is devoted to Monte Carlo estimation of Bermudan sensitivities using unconditional probabilistic representations and in Section 4 we study the estimation of sensitivities via regression methods using conditional probabilistic representations. In Section 5 an efficient least squares method for constructing deltas is proposed in the case where at all (time-space) positions a sufficiently accurate price of the Bermudan option under consideration is available. A method for testing the proposed procedures in an application is outlined in Section 6. In Section 7 it is proposed to implement this test procedure for regularisation of calibration routines.

2. PRELIMINARIES

2.1. **Modelling framework.** We consider a price system consisting of a saving account B (riskless asset) and price processes X^i, $i=1,\ldots,d$, of risky assets, which satisfies the stochastic differential equations (SDEs)

$$\frac{dX^i}{X^i} = r(t,X)dt + \sum_{j=1}^{d} \sigma^{ij}(t,X)dW^j, \frac{dB}{B} = r(t,X)dt, \tag{2.1}$$

in the (risk-neutral) measure P. In (2.1) $W=(W^1,...,W^d)^\top$ is a d-dimensional standard Wiener process on a probability space $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\in[t_0,T]},P)$, where the P-augmentation of the filtration generated by W is denoted by $(\mathcal{F}_t)$. It is assumed that the interest rate $r(t,X)$ and the matrix $\sigma(t,x)=\{\sigma^{ij}(t,x)\}, t\in[t_0,T], x\in\mathbf{R}^d_+ := \{x : x^1>0,\ldots,x^d>0\}$ are such that for all $x\in\mathbf{R}^d_+$ and $s\in[t_0,T]$ there exists a unique solution $(X_t,B_t)=(X_t^{s,x},B_t^{s,x,b})\in\mathbf{R}^d_+\times\mathbf{R}_+$ of (2.1) for $s\le t\le T$ with $(X_s^{s,x},B_s^{s,x,b})=(x,b)$ (note that $B_t^{s,x,b}=bB_t^{s,x,1}$ for any $b>0$), which is smooth in x, and that all X_t/B_t are (true) martingales on $[t_0,T]$ under the risk-neutral measure P. For example, it is sufficient to assume that the functions

$$a^i(t,x) := x^i r(t,x), c^{ij}(t,x) := x^i\sigma^{ij}(t,x) \tag{2.2}$$

are continuous in t, and have bounded derivatives with respect to x up to some order. In addition, we assume that the volatility matrix $\sigma(t,x)$ has full rank for every $(t,x), t\in[t_0,T], x\in\mathbf{R}^d_+$. Under these assumptions the price system (X,B) constitutes a complete market [21]. In this article we consider exact solutions of (2.1) although, of course, in reality we have only approximations usually. For instance, one may integrate (2.1) numerically and so construct an accurate enough solution of (2.1) using, for example, a standard Euler scheme with sufficiently small time step Δt.

Let us now consider a Bermudan contract with exercise dates $t_0=T_0<T_1<...<T_I=T$ and corresponding pay-off functions $f_i(x), 0\le i\le I$. The holder of this contract has the right to call (once) a cash-flow $f_i(X_{T_i})$ at an exercise date T_i of his choice. Henceforth we will write, where convenient, subindex i instead of subindex T_i. For instance $B_i := B_{T_i}$, $X_i := X_{T_i}$, and so on. The process X_i is a Markov chain with respect to the discrete filtration $(\mathcal{F}_i)_{0\le i\le I} := (\mathcal{F}_{T_i})_{0\le i\le I}$. For notational convenience we sometimes use the cash-flow discounted from time t to s :

$$Z_t^{s,x} := \frac{f_t(X_t^{s,x})}{B_t^{s,x,1}}, \quad s\le t. \tag{2.3}$$

It is well known (e.g. [11]) that the fair price of a Bermudan contract with remaining exercise dates $\{T_i, T_{i+1}, .., T_I\}$, at a time $T_{i-1} < s \leq T_i$, is given by

$$u_s(x) := \sup_{\tau \in \mathcal{T}_{i,I}} EZ_\tau^{s,x} = EZ_{\tau^{s,x}}^{s,x}, \quad T_{i-1} < s \leq T_i, \tag{2.4}$$

where $\mathcal{T}_{i,I}$ the set of stopping times τ taking values in $\{T_i, T_{i+1}, ..., T_I\}$, and $\tau^{s,x}$ is the minimal optimal stopping index. The value function $u_s(x)$ is determined by induction as follows:

$$\begin{aligned} u_I(x) &= f_I(x), \\ u_i(x) &= \max\left\{ f_i(x), E\left(\frac{u_{i+1}(X_{i+1}^{i,x})}{B_{i+1}^{i,x,1}} \right) \right\}, \\ u_s(x) &= E\left(\frac{u_{i+1}(X_{i+1}^{s,x})}{B_{i+1}^{s,x,1}} \right), \ T_i < s \leq T_{i+1}, \ i = I-1, ..., 0. \end{aligned} \tag{2.5}$$

We see that, in principle, the problem of evaluating $u_0(x_0)$, i.e. the price of the Bermudan option at the initial position (t_0, x_0), is easily solved using the backward recursive procedure (2.5), also called backward dynamic program. However, if X is high dimensional and I is large, this procedure is practically not feasible.

For the solution $(X_t, B_t) = (X_t^{t_0,x}, B_t^{t_0,x,1})$ of (2.1) on the interval $[t_0, T]$, the discrete discounted value process $Y_i := u_i(X_i)/B_i$, called the Snell envelope, is a supermartingale due to (2.5), and in particular Y_i is the smallest supermartingale which dominates the discounted cash-flow process $f_i(X_i)/B_i$, e.g. [32].

2.2. **Continuation values and optimal stopping times.** For the considered Bermudan option, we introduce the continuation value

$$C_i(x) = E\left(\frac{u_{i+1}(X_{i+1}^{i,x})}{B_{i+1}^{i,x,1}} \right), \ i = 0, ..., I-1; \ C_I(x) = f_I(x), \tag{2.6}$$

the continuation region $\mathcal{C}$, and the exercise (stopping) region $\mathcal{E}$:

$$\begin{aligned} \mathcal{C} &= \{(t_i, x) : f_i(x) < C_i(x), \ i = 0, ..., I\}, \\ \mathcal{E} &= \{(t_i, x) : f_i(x) \geq C_i(x), \ i = 0, ..., I\}. \end{aligned} \tag{2.7}$$

Clearly, $(t_I, x) \in \mathcal{E}$ for any x. Due to (2.5)

$$u_i(x) = \max\{f_i(x), C_i(x)\}, \ i = 0, ..., I. \tag{2.8}$$

It is not difficult to see that $C_i(x)$ has also the probabilistic representation

$$C_i(x) = E\left(\frac{f_\tau(X_\tau^{i,x})}{B_\tau^{i,x,1}} \right), \ \tau = \tau^{t_{i+1}, X_{i+1}}, \ X_{i+1} = X_{i+1}^{i,x}, \tag{2.9}$$

where the optimal stopping time $\tau^{i,x} = \tau^{t_i,x}$ is defined recurrently by the dynamic programming principle in the following way. We set

$$\begin{aligned}
\tau^{I,x} &= \tau^{T,x} = T, \\
\tau^{i,x} &= T_i \chi_{\{C_i(x) \le f_i(x)\}} + \tau^{i+1, X_{i+1}^{i,x}} \chi_{\{C_i(x) > f_i(x)\}} \\
&= T_i \chi_{\{u_i(x) = f_i(x)\}} + \tau^{i+1, X_{i+1}^{i,x}} \chi_{\{u_i(x) > f_i(x)\}}, \\
i &= I-1, \ldots, 0.
\end{aligned} \tag{2.10}$$

Thus, for any position (T_i, x), the optimal stopping time $\tau^{i,x}$ is either equal to $T_i :\ \tau^{i,x} = T_i$, or $\tau^{i,x} > T_i$. It is also clear that (T_i, x) is a stopping point (i.e., $\tau^{i,x} = T_i$) iff $(T_i, x) \in \mathcal{E}$ (i.e., (T_i, x) belongs to the exercise region). The instant $\tau^{i,x}$ is either the first time that the trajectory $(T_j, X_j^{i,x})$ enters the region $\mathcal{E}$ during $i \le j \le I-1$, or $\tau^{i,x} = T_I$. So, $(\tau^{i,x}, X_{\tau^{i,x}}^{i,x}) \in \mathcal{E}$ (see (2.7)).

2.3. **General schemes for the regression method.** Let us consider a sample $({}_mX, {}_mV)$, $m = 1$, ... $, M$, from a generic random pair (X, V). We are interested in the estimation of the regression function

$$c(x) = E(V | X = x). \tag{2.11}$$

There are different regression methods for estimating $c(x)$ in fact. We here recall a general scheme for the linear regression method (see, e.g., [18]). Let $\{\psi_r(x)\}_{r=1}^K$ be a set of basis functions each mapping $\mathbf{R}_+^d$ into $\mathbf{R}$. An estimate $\hat{c}(x)$ of $c(x)$ is obtained as a function of the form $\sum_{k=1}^K \alpha_k \psi_k(x)$ which minimizes the empirical risk

$$\hat{\alpha} = \arg \min_{\alpha \in \mathbf{R}^K} \frac{1}{M} \sum_{m=1}^M ({}_mV - \sum_{k=1}^K \alpha_k \psi_k({}_mX))^2. \tag{2.12}$$

So

$$\hat{c}(x) = \sum_{k=1}^K \hat{\alpha}_k \psi_k(x). \tag{2.13}$$

Thus, the usual base material for the linear regression procedure is a sample $({}_mX, {}_mV)$, $m = 1, \ldots, M$, from a generic random pair (X, V), where X is a d-dimensional random vector, V is a one-dimensional random variable, and a set of basis functions $\{\psi_r\}$.

2.4. **Recursively estimating continuation values and optimal stopping times.** Suppose we have a sample $(t,\ {}_mX_t,\ {}_mB_t)$ of M independent trajectories all starting from the position $(t_0, X_0, 1)$. There are different regression methods (and even different generic members and samples) for estimating u_i by backward recursion. For example, one may estimate the continuation value function $C_i(x)$ backwardly by regression via (2.12)-(2.13) by considering the $(d+1)$-dimensional sample

$$({}_mX_i, {}_mV_i) := ({}_mX_i, \frac{{}_mB_i}{{}_mB_{i+1}} \hat{u} \tag{2.14}$$

from $(X_i, V_i) := (X_i, \frac{B_i}{B_{i+1}} \hat{u}_{i+1}(X_{i+1}))$ where $\hat{u}_{i+1}(X_{i+1})$ is an estimate of $u_{i+1}(X_{i+1})$ obtained by (2.8) using an approximation of $C_{i+1}(x)$ in (2.6). This procedure is due to [34], see also [15].

In the method of [25] one considers backwardly the sample

$$(2.15)\quad ({}_mX_i,\ \frac{{}_mB_i}{{}_mB_\tau} f_\tau({}_mX_\tau^{t_{i+1},\ {}_mX_{i+1}})) = ({}_mX_i,\ \frac{{}_mB_i}{{}_mB_\tau} f_\tau({}_mX_\tau^{t_i,\ {}_mX_i})),\ \tau = \tau^{t_{i+1},\ {}_mX_{i+1}},\ m = 1,...,M,$$

from $(X_i,\ \frac{B_i}{B_\tau} f_\tau(X_\tau^{t_{i+1},X_{i+1}})) = (X_i,\ \frac{B_i}{B_\tau} f_\tau(X_\tau^{t_i,X_i}))$, where $\tau = \tau^{t_{i+1},X_{i+1}}$, and then constructs an approximation of $C_i(x)$ due (2.9) by regression via (2.12)-(2.13).

We here briefly recap the regression method from [25] which is analyzed rigorously in [10]. Let ψ_r, $r = 1,...,K$, be a system of real valued base functions on the state space $\mathbf{R}_+^d$. Set $\hat{\tau}^{I,x} = T_I$, and $\hat{C}_I({}_mX_I) = -1,\ \ m = 1,\dots,M$. For $i < I$ we recursively construct $\hat{\tau}^{i,\ {}_mX_i}$, $\hat{C}_i({}_mX_i)$ from $\hat{\tau}^{i+1,\ {}_mX_{i+1}}$, $\hat{C}_{i+1}({}_mX_{i+1})$, $m = 1,\dots,M$, as follows. Via standard least squares minimization we compute a system of regression coefficients $(c_{ir})_{1\le r\le K}$,

$$(2.16)\qquad (c_{ir})_{1\le r\le K} := \underset{(c_r)_{1\le r\le K}}{\operatorname{argmin}} \sum_{m=1}^{M} \left(\sum_{r=1}^{K} c_r \psi_r({}_mX_i) - {}_mZ^{i,{}_mX_i}_{\hat{\tau}^{i+1,{}_mX_{i+1}}} \right)^2,$$

and set

$$(2.17)\qquad \hat{C}_i({}_mX_i) := \sum_{r=1}^{K} c_{ir}\psi_r({}_mX_i), \quad m = 1,\dots,M.$$

We then define

$$\hat{\tau}^{i,\ {}_mX_i} = T_i \quad \text{if} \quad \hat{C}_i({}_mX_i) \le f_i({}_mX_i), \quad \text{else} \quad \hat{\tau}^{i,\ {}_mX_i} = \hat{\tau}^{i+1,\ {}_mX_{i+1}}.$$

As a result all the positions $(t,\ {}_mX_t)$, $T_{i-1} < t \le T_i$, are equipped with stopping times $\hat{\tau}^{t_i,\ {}_mX_i}$. This gives us the following estimate for $u_0(X_0)$:

$$(2.18)\qquad \hat{u}_0(X_0) = \frac{B_0}{M} \sum_{m=1}^{M} \frac{1}{B_{{}_m\tau}} f_{{}_m\tau}({}_mX_{{}_m\tau}^{t_0,\ X_0}),$$

where either ${}_m\tau = t_0$ (i.e., (t_0, X_0) belongs to the exercise region and $\hat{u}_0(X_0) = f_0(X_0)$) or ${}_m\tau > t_0$ (i.e., ${}_m\tau = \hat{\tau}^{t_1,\ {}_mX_1}$).

Moreover, via (2.17) we get an approximate sequence of continuation value functions

$$(2.19)\qquad \hat{C}_i(x) := \sum_{r=1}^{K} c_{ir}\psi_r(x), \quad 0 \le i < I, \quad \hat{C}_I = -1,$$

and obtain a lower biased approximation

$$(2.20)\qquad \hat{u}_s(x) = EZ^{s,x}_{\hat{\tau}^{s,x}}, \quad T_{i-1} < s \le T_i,$$

for the true price of a Bermudan option with exercise possibilities $\{T_i, T_{i+1}, .., T_I\}$, via the stopping rule

$$(2.21)\qquad \hat{\tau}^{s,x} = \inf\{T_j \ge T_i : \hat{C}_j(X_{T_j}^{s,x}) \le f_j(X_{T_j}^{s,x})\}.$$

It should be noted that for a lower biased estimate of (2.20) by Monte Carlo simulation, one should use a new set of trajectories all starting from (s, x).

Another estimate for $\hat{u}_i(x) = \hat{u}_{T_i}(x)$ can be obtained by (see (2.8) and (2.19))

$$(2.22)\qquad \hat{u}_i(x) = \max\left\{f_i(x), \hat{C}_i(x)\right\},\ i = 0,..., I.$$

This estimate is less expensive but apparently less accurate than (2.20). Having $\hat{u}_i(x)$, we can estimate $u_s(x)$ for $T_{i-1} < s < T_i$ by pricing the European claim at the maturity time T_i specified by the payoff function $\hat{u}_i(x)$.

3. Monte Carlo estimation of Bermudan sensitivities at a fixed position

In this section we concentrate on the evaluation of deltas

$$\partial_s^k(x) := \frac{\partial u_s(x)}{\partial x^k},\ k = 1,...,d,\ t_0 \le s \le T,$$

at the starting position (t_0, X_0). Let us introduce the notation

$$\partial_s(x) = (\partial_s^1(x), ..., \partial_s^d(x))^\intercal = \left(\frac{\partial u_s(x)}{\partial x^1}, ..., \frac{\partial u_s(x)}{\partial x^d}\right)^\intercal.$$

In order to simplify the presentation we now assume $r(t,x) \equiv: r$ in (2.1) to be constant.

3.1. **Pathwise approach.** We note that $\tau^{s,x}$ in (2.4) depends on s, x and on an elementary random event $\omega: \ \tau^{s,x} = \tau^{s,x}(\omega)$. Due to finiteness of $\mathcal{T}_{i,I}$, we have $\tau^{s,x+\Delta x}(\omega) = \tau^{s,x}(\omega)$ almost surely for all Δx small enough if (s,x) does not belong to the boundary of the exercise region. Therefore we obtain from (2.4):

$$\partial_s^k(x) = B_s E\left(\frac{1}{B_{\tau^{s,x}}}\sum_{i=1}^{d}\frac{\partial f_{\tau^{s,x}}(X_{\tau^{s,x}}^{s,x})}{\partial x^i}\delta^k(X_{\tau^{s,x}}^{s,x})^i\right), \tag{3.1}$$

where

$$\delta^k X_t^i := \delta^k(X_t^{s,x})^i = \frac{\partial (X_t^{s,x})^i}{\partial x^k},\ s \le t \le T,\ k,\ i = 1,...,d,$$

satisfies the system of variational equations associated with (2.1) and (2.2):

$$d\delta^k X = \sum_{i=1}^{d}\delta^k X^i \frac{\partial a(t,X)}{\partial x^i}dt + \sum_{i=1}^{d}\delta^k X^i \frac{\partial c(t,X)}{\partial x^i}\,dW(t), \tag{3.2}$$
$$\delta^k X^i(s) = 0,\ \text{if } i \ne k, \text{ and } \delta^k X^k(s) = 1.$$

We note that if the position (s,x) belongs to the exercise region (hence $\tau^{s,x} = s = T_k$ for some $k = 0, ..., I$), then formula (3.1) evidently gives

$$\partial_s^k(x) = \frac{\partial f_s(x)}{\partial x^k}.$$

Let $\tau^{t_0,x} > t_0$. Based on the independent trajectories $(t,\ {}_mX_t)$, $m = 1,...,M$, all starting from the point (t_0, X_0), we obtain from (3.1) with $(s,x) = (t_0, X_0)$ the following Monte Carlo estimates for the deltas:

$$\hat{\partial}_{t_0}^k(X_0) = \frac{B_0}{M}\sum_{m=1}^{M}\frac{1}{B_{{}_m\tau}}\sum_{i=1}^{d}\frac{\partial f_{{}_m\tau}({}_mX_{{}_m\tau}^{t_0,X_0})}{\partial x^i}\delta^k({}_mX_{{}_m\tau}^{t_0,X_0})^i,\ {}_m\tau = \tau^{t_1,\ {}_mX_1}. \tag{3.3}$$

For a corresponding estimator in the case where the interest rate $r = r(t,x)$ is non-constant we get somewhat more complicated expressions (see e.g. [26]).

3.2. **Finite difference based approach.** Formula (3.1) requires computation of $\delta^k X_t^i$, i.e., to evaluate deltas by the pathwise approach one has to integrate not only the d-dimensional system for X from (2.1) but also d additional systems, each of dimension d. This can cause severe computational difficulties in practice. However, in [16] an adjoint simulation approach is presented to speed up this procedure. As an alternative, we may compute (approximately) the derivatives in (3.1) by finite differences as studied in [27] for the European case. This method is based on the finite difference formula (we put $X_0 = x$)

$$(3.4) \qquad \frac{\partial u_t(x)}{\partial x^k} = \frac{u_t(x^1,\ldots,x^k+\Delta x^k,\ldots,x^d) - u_t(x^1,\ldots,x^k-\Delta x^k,\ldots,x^d)}{2\Delta x^k} + O\left(\left(\Delta x^k\right)^2\right).$$

In typical situations, we need to use approximations $\hat{u}_t(x^1,\ldots,x^k \pm \Delta x^k,\ldots,x^d)$ for $u_t(x^1,\ldots,x^k \pm \Delta x^k,\ldots,x^d)$, for example (see (2.18)):

$$(3.5) \qquad \hat{u}_{t_0}(x^1,\ldots,x^k \pm \Delta x^k,\ldots,x^d) = \frac{B_0}{M}\sum_{m=1}^{M}\frac{1}{B_{m\tau}} f_{m\tau}({}_mX_{m\tau}^{t_0,x^1,\ldots,x^k\pm\Delta x^k,\ldots,x^d}).$$

In (3.5), ${}_mX_t$, $t_0 \le t \le {}_m\tau$, is an approximate solution of (2.1) obtained by a numerical integration scheme. M triples of approximate trajectories are simulated, each triple consists of a trajectory starting from x and a pair of trajectories starting from $x \pm \Delta x^k e^k := (x^1,\ldots,x^k \pm \Delta x^k,\ldots,x^d)$ at the moment t_0. The triples are independent, but the three trajectories of the same triple are dependent: they correspond to the same realization of the Wiener process. We therefore take the same stopping time ${}_m\tau = \hat{\tau}^{t_1,\,{}_mX_1}$, ${}_mX_1 = {}_mX_{T_1}^{t_0,x}$ for each trajectory of the m-th triple. For a particular triple this is correct when Δx^k is small enough. In the finite difference approach one takes one fixed Δx^k for all trajectories however, so for some trajectories this may fail to be correct. But, with decreasing Δx^k the number of such trajectories decreases as well. As an estimator for $\partial_{t_0}^k(X_0)$ we thus obtain

$$(3.6) \qquad \hat{\partial}_{t_0}^k(X_0) = \frac{1}{2\Delta x^i}\frac{B_0}{M}\sum_{m=1}^{M}\frac{1}{B_{m\tau}}\left[f_{m\tau}({}_mX_{m\tau}^{t_0,x+\Delta x^k e^k}) - f_{m\tau}({}_mX_{m\tau}^{t_0,x-\Delta x^k e^k})\right].$$

For European options it is proved in [27] that the delta-estimator (3.6) has the same convergence rate as the estimator for the option value. Using the fact that Bermudan option values are continuous in x at the exercise boundary, and good properties of the estimator (3.6) in the European case, a similar result can be shown for (3.6) as well.

3.3. **The case of accurately known $\hat{u}_{T_i}(x)$.** Proceeding with the problem of evaluating deltas, we may naturally presuppose that $\hat{u}_{T_i}(x)$ is a sufficiently accurate estimate of $u_{T_i}(x)$ for all x and i. In this case one can use results obtained for European options (see, e.g., [27] and references therein) for evaluation of $\partial_t^k(x)$ at a position (t,x), where $T_{i-1} \le t < T_i$. (Note that if $t = T_{i-1}$, one needs to take the max-operator into account, (2.5)). In particular, the theory developed for European options enables us to apply weak methods of numerical integration for SDEs and variance reduction techniques. In this context we recall an approach in [27] which is based on a Clark-Ocone type formula. Let $\xi = (\xi^1,\ldots,\xi^d)^\intercal$ be a d-dimensional random vector with i.i.d. components distributed by $P(\xi^k = \pm 1) = 1/2$, and h be a positive number. With (2.2) and the matrix $c(t,x) := \{c^{ij}(t,x)\}$ we denote by $\bar{X}$ the one-step approximation

$$\bar{X} := x + ha(t,x) + h^{1/2}c(t,x)\xi.$$

Then,

$$c^{\top}(t,x)\partial_t(x) = \frac{1}{\sqrt{h}} E[u_{t+h}(\bar{X})\xi] + O(h).$$

This formula can be used for effective estimation of $\hat{\partial}_t(x),\ T_{i-1} \le t < t+h \le T_i$ via Monte Carlo simulation, provided $\hat{u}_{T_i}(x)$ is close to $u_{T_i}(x)$ (see [27]).

4. Regression methods for Bermudan sensitivities at arbitrary positions

The pathwise Monte Carlo estimator (3.3) and finite difference Monte Carlo estimator (3.6) can be used for estimating sensitivities at a given starting position of the process X, say $(t_0, X_{t_0}) = (t_0, x_0)$, using a sample set of trajectories starting at (t_0, x_0). For estimating sensitivities at another position, let us say (s,x) with $t_0 < s \le T_1$, one can of course simulate a new set of trajectories and use (3.3) or (3.6) again. As a more efficient alternative, we propose to infer the sensitivities from the old trajectories via the well-known method of linear regression.

4.1. **Pathwise methods with linear regression.** We here propose to use the old trajectories (together with the corresponding stopping times) starting at (t_0, x_0), for evaluation of the price and deltas for the new position (s,x), $t_0 < s \le t_1$. The key tools are the following expressions (see (2.4) and (3.1)),

$$u_s(x) = B_s E\left(\frac{f_\tau(X_\tau^{s,x})}{B_\tau}\right) \text{ with } \tau = \tau^{s,x} \tag{4.1}$$

$$= B_s E\left(\frac{1}{B_\tau} f_\tau(X_\tau^{s,X}) \text{ with } X := X_s^{t_0,x_0} \text{ and } \tau = \tau^{t_0,x_0} \mid X_s^{t_0,x_0} = x\right)$$

and

$$\partial_s^k(x) = B_s E\left(\frac{1}{B_\tau}\sum_{i=1}^d \frac{\partial f_\tau(X_\tau^{s,x})}{\partial x^i}\delta^k(X_\tau^{s,x})^i\right) \text{ with } \tau = \tau^{s,x} \tag{4.2}$$

$$= B_s E\left(\frac{1}{B_\tau}\sum_{i=1}^d \frac{\partial f_\tau(X_\tau^{s,X})}{\partial x^i}\delta^k(X_\tau^{s,X})^i \mid X_s^{t_0,x_0} = x\right) \text{ with } X := X_s^{t_0,x_0} \text{ and } \tau = \tau^{t_0,x_0}.$$

Thus, we have two different probabilistic representations both for $u_s(x)$ and $\partial_s^k(x)$: the first one is in the form of an unconditional expectation and the second one is in the form of a conditional one. The first forms can be naturally computed by direct Monte Carlo and the second one by a regression method. Using regression based on one and the same set of trajectories $(t, {}_mX_t^{t_0,x_0})$ and stopping times ${}_m\tau = {}_m\tau^{t_0,x_0}$, $m = 1,\dots,M$, for many positions (s,x) is of course much more efficient than simulating new from each desired position (s,x) (though generally less accurate of course)

A proof of (4.1) and (4.2) relies on the following assertion: if ζ is $\tilde{\mathcal{F}}$-measurable, $f(x,\omega)$ is independent of $\tilde{\mathcal{F}}$, and $Ef(x,\omega) = \phi(x)$, then $E(f(\zeta,\omega)|\tilde{\mathcal{F}}) = \phi(\zeta)$ (see, e.g., [23]). From this assertion, for any measurable g it holds (with $\zeta = X_s^{t_0,x_0}$, $\tilde{\mathcal{F}} = \sigma(X_s^{t_0,x_0})$, $f(x,\omega) = g(X_\tau^{s,x}(\omega))$):

$$E\left(g(X_\tau^{s,X}) \mid X_s^{t_0,x_0} = x\right) = Eg(X_\tau^{s,x}) \text{ with } X := X_s^{t_0,x_0} \text{ and } \tau \ge s,$$

hence (4.1) and (4.2).

For evaluating $u_s(x)$ by regression, the pairs (X,V) and $({}_mX,{}_mV)$ (see Subsection 2.3) have the form

$$(X,V) \sim \left(X_s^{t_0,x_0}, \frac{B_s}{B_\tau} f_\tau(X_\tau^{s,X}) \right), \tag{4.3}$$

$$({}_mX,{}_mV) \sim \left({}_mX_s^{t_0,x_0}, \frac{B_s}{B_{{}_m\tau}} f_{{}_m\tau}({}_mX_{{}_m\tau}^{s,{}_mX}) \right),$$

and for evaluating $\partial_s^k(x)$ they have the form

$$(X,V) \sim \left(X_s^{t_0,x_0}, \frac{B_s}{B_\tau} \sum_{i=1}^{d} \frac{\partial f_\tau(X_\tau^{s,X})}{\partial x^i} \delta^k (X_\tau^{s,X})^i \right), \tag{4.4}$$

$$({}_mX,{}_mV) \sim \left({}_mX_s^{t_0,x_0}, \frac{B_s}{B_{{}_m\tau}} \sum_{i=1}^{d} \frac{\partial f_{{}_m\tau}({}_mX_{{}_m\tau}^{s,{}_mX})}{\partial x^i} {}_m\delta^k ({}_mX_{{}_m\tau}^{s,{}_mX})^i \right).$$

Having the sample (4.3), one can obtain the estimate $\hat{u}_s(x)$ by linear regression as

$$\hat{u}_s(x) = \sum_{l=1}^{L} \hat{\alpha}_l \varphi_l(x).$$

In principle one can differentiate this expression to obtain an estimate for $\partial_s^k(x)$. However, in general this is not a suitable method. It is better to apply linear regression directly to the sample (4.4) to get

$$\hat{\partial}_s^k(x) = \sum_{l=1}^{L_1} \hat{\beta}_l^k \psi_l(x).$$

Of course, the choice of basis functions $\{\varphi_l\}$ and $\{\psi_l\}$ is a problem in it's own right.

4.2. **Finite difference based methods via linear regression.** We have

$$\partial_s^k(x) \simeq \frac{1}{2\Delta x^k} [u_s(x+\Delta x^k) - u_s(x-\Delta x^k)]. \tag{4.5}$$

Similar to (4.1) and (4.2), the right part of (4.5) has the following probabilistic representations:

$$\frac{1}{2\Delta x^k} [u_s(x+\Delta x^k) - u_s(x-\Delta x^k)] \tag{4.6}$$

$$= \frac{1}{2\Delta x^k} E \left(\frac{B_s}{B_\tau} \left[f_\tau(X_\tau^{s,x+\Delta x^k e^k}) - f_\tau(X_\tau^{s,x-\Delta x^k e^k}) \right] \text{ with } \tau = \tau^{s,x} \right)$$

$$= \frac{1}{2\Delta x^k} E \left(\frac{B_s}{B_\tau} \left[f_\tau(X_\tau^{s,X+\Delta x^k e^k}) - f_\tau(X_\tau^{s,X-\Delta x^k e^k}) \right] \text{ with } \tau = \tau^{t_0,x_0} | X = X_s^{t_0,x_0} = x \right).$$

The corresponding generic member (X,V) and the sample $({}_mX,{}_mV)$ have the form

$$(X,V) \sim \left(X_s^{t_0,x_0}, \frac{B_s}{2\Delta x^k B_\tau} \left[f_\tau(X_\tau^{s,X+\Delta x^k e^k}) - f_\tau(X_\tau^{s,X-\Delta x^k e^k}) \right] \right), \tag{4.7}$$

$$({}_mX,{}_mV) \sim \left({}_mX_s^{t_0,x_0}, \frac{B_s}{2\Delta x^k B_{{}_m\tau}} \left[f_{{}_m\tau}({}_mX_{{}_m\tau}^{s,{}_mX+\Delta x^k e^k}) - f_{{}_m\tau}({}_mX_{{}_m\tau}^{s,{}_mX-\Delta x^k e^k}) \right] \right).$$

For any m we have to consider three trajectories defined on the interval $s \leq t \leq {}_m\tau$. One of them is $(t, {}_mX_t^{t_0,x_0}) = (t, X_t^{s,{}_mX})$ which is constructed previously. Two others $(t, X_t^{s,{}_mX \pm \Delta x^k e^k})$ start from points $X \pm \Delta x^k e^k$ at time s, and they have to be constructed. We recall that these three trajectories are dependent: two last trajectories have the same realization of the Wiener process as the first (basic) trajectory. Besides, let us note that Δx^k is the same for all m.

4.3. **Greeks by local polynomial regression.** In the case that we only need $\hat{u}_s(x)$ and $\hat{\partial}_s^k(x)$ at a particular point x at time s it is advantageous to use local regression (see e.g. [14]). In the local regression approach the regression function

$$c(x) = E(V|X = x) \tag{4.8}$$

and its derivatives $c'(x), ..., c^{(p)}(x)$ at a point $x = x_0$ are to be estimated from a sample $({}_mX, {}_mV)$, $m = 1, ..., M$, from (X, V). For notational simplicity we here suppose that X and V are one-dimensional. Assume that the $(p+1)$-th derivative of $c(x)$ at x_0 exists. Then the unknown regression function $c(x)$ can be locally approximated by a polynomial of order p due to the weighted least squares regression problem (see [18])

$$\sum_{m=1}^{M} K_h({}_mX - x_0) \cdot \left[{}_mV - \sum_{j=0}^{p} \beta_j ({}_mX - x_0)^j \right]^2 \to \min_{\beta_0,...,\beta_p}, \tag{4.9}$$

where h is a bandwidth, and K_h is a kernel function. If $\hat{\beta}_j$, $j = 0, ..., p$, is the solution to problem (4.9) then,

$$\hat{c}(x_0) = \hat{\beta}_0, ..., \hat{c}^{(j)}(x_0) = j!\hat{\beta}_j, \; j = 0, ..., p. \tag{4.10}$$

The local polynomial regression approach allows for different alternatives. For example, to get an estimate for $\partial_s^k(x)$ one can use a local regression method directly to the sample (4.4) or one can use a polynomial regression method of order $p \geq 1$ to the sample (4.3) and obtain both estimates $\hat{u}_s(x)$ and $\hat{\partial}_s^k(x)$. Apparently the second alternative for estimating $\partial_s^k(x)$ is less accurate but not so expensive since sample (4.3) is simpler. If we apply a polynomial regression method of order $p \geq 1$ to sample (4.4), we can obtain estimates both for deltas and gammas. The latter is very noteworthy, since known methods for estimating gammas are very expensive. Moreover we note that local regression methods do not involve the rather difficult problem of choice of basis functions. Admittedly, it involves the problem of bandwidth selection however. But this difficulty can in principle be solved during numerical calculations. Over all, the local regression method in the situation where we need $\hat{u}_s(x)$ and $\hat{\partial}_s^k(x)$ only at one position (s, x) looks promising.

5. Direct least squares methods for deltas

Due to (2.5), $u_t(x)$ is continuous in t for $T_i < t < T_{i+1}$, $i = 0, ..., I-1$, $u_{T_{i+1}-0}(x) = u_{T_{i+1}}(x)$, and if

$$f_i(x) > u_{T_i+0}(x) = E\left(\frac{u_{i+1}(X_{i+1}^{i,x})}{B_{i+1}^{i,x,1}} \right),$$

then

$$u_{T_i+0}(x) < u_{T_i}(x) = f_i(x).$$

Together with (2.1) we have on $(T_i, T_{i+1}]$

$$du = rudt + \sum_{j=1}^{d} \frac{\partial u}{\partial x^j} \sum_{k=1}^{d} c^{jk} dW^k(t) \tag{5.1}$$

or

$$u_t(X_t) = u_{T_i+0}(x) + \int_{T_i}^{t} r(s,X_s)u_s(X_s)ds + \int_{T_i}^{t} \sum_{j=1}^{d} \frac{\partial u_s(X_s)}{\partial x^j} \sum_{k=1}^{d} c^{jk}(s,X_s)dW^k(s), \tag{5.2}$$

where $X_s = X_s^{T_i,x}$.

Let h be a small positive number and $T_i \le t < t+h \le T_{i+1}$. We assume that X_t is supposed to be known exactly (see Section 2). In the context below, this means that the time step Δt used in a numerical integration scheme for computing X_t (e.g. the Euler scheme) is much smaller than h: $\Delta t \ll h$. From (5.2) we get the following approximate equality

$$u_{t+h}(X_{t+h}) \simeq u_t(X_t) + r(t,X_t)u_t(X_t)h + [c^{\intercal}(t,X_t)\partial_t(X_t)]^{\intercal}\Delta_t W(h), \tag{5.3}$$

where $\Delta_t W(h) = W(t+h) - W(t)$. Of course, if $t = T_i$, one should set u_{T_i+0} instead of u_{T_i} in (5.3).

Suppose $\hat{u}_t(x)$ and $\hat{u}_{t+h}(x)$ are known (e.g., they are evaluated by one of the procedures in [5]), ${}_mX_s$, $m = 1,\dots,M$, are trajectories starting from (t_0,x_0), and ${}_mW(s)$ are the corresponding realizations of the Wiener process. We then look for $\partial_t(x) = (\partial_t^1(x),\dots,\partial_t^d(x))^{\intercal}$ in the form

$$\partial_t^k(x) \sim \sum_{l=1}^{L_k} \alpha_l^k \psi_l^k(x),\ k = 1,\dots,d, \tag{5.4}$$

where L_k and ψ_l^k may depend on t. Substituting in (5.3) ${}_mX_t$, ${}_mX_{t+h}$, $\hat{u}_t$, $\hat{u}_{t+h}$ instead of X_t, X_{t+h}, u_t, u_{t+h}, then ${}_mW(t+h) -_m W(t)$ instead of $\Delta_t W(h)$, and $\partial_t({}_mX_t)$ in the form (5.4), we obtain M relations, whence $\hat{\alpha}_l^k$ can be found by the method of least squares, i.e.

$$\hat{\alpha}_l^k = \arg\min_{\alpha_l^k} \frac{1}{M} \sum_{m=1}^{M} \left[{}_m\Delta\hat{u} - ({}_mr)({}_m\hat{u})h - \sum_{i=1}^{d} ({}_mZ^i)({}_m\Delta W^i) \right]^2, \tag{5.5}$$

where

$$ {}_m\Delta\hat{u} = \hat{u}_{t+h}({}_mX_{t+h}) - \hat{u}_t({}_mX_t),\ {}_mr = r(t,{}_mX_t),\ {}_m\hat{u} = \hat{u}_t({}_mX_t),$$
$$ {}_m\Delta W^i =_m W^i(t+h) -_m W^i(t),$$
$$ {}_mZ^i = \sum_{k=1}^{d} ({}_mc^{ki}) \sum_{l=1}^{L_k} \alpha_l^k ({}_m\psi_l^k),\ \text{with } {}_mc^{ki} = c^{ki}(t,{}_mX_t),\ {}_m\psi_l^k = \psi_l^k({}_mX_t).$$

Using the technique from, e.g. [18] (see also [17]), it is not difficult to prove that $\hat{\partial}_t^k(x) = \sum_{l=1}^{L_k} \hat{\alpha}_l^k \psi_l^k(x)$ converge to the true deltas when the estimation error of $\hat{u}$ goes to zero, $h \to 0$, $M \to \infty$, and $L_k \to \infty$. In our case the error analysis is much simpler than in [17]. In contrast to [17], we consider only a one-step error, we consider both $\hat{u}_t(x)$ and $\hat{u}_{t+h}(x)$ to be accurately known, and finally the expression inside the square brackets in (5.5) is linear with respect to α_l^k.

As an alternative least squares problem we may look for

$$Z_t(x) := c^{\mathsf{T}}(t, X_t)\partial_t(X_t)$$

in the form

$$Z_t^k(x) \sim \sum_{l=1}^{L} \beta_l^k \zeta_l(x),\ k = 1, ..., d, \tag{5.6}$$

where L and ζ_l do not depend on k. The solution $\hat{\beta}_l^i$ satisfies the minimization problem

$$\hat{\beta}_l^i = \arg\min_{\beta_l^i} \frac{1}{M} \sum_{m=1}^{M} \left[{}_m\Delta\hat{u} - ({}_mr)({}_m\hat{u})h - \sum_{i=1}^{d}\sum_{l=1}^{L} \beta_l^i({}_m\zeta_l)({}_m\Delta W^i) \right]^2 . \tag{5.7}$$

Both minimization problem (5.5) and (5.7) lead to a linear algebraic system of order $d \times L_k$. We now show that the linear system corresponding to problem (5.7) can, in a good approximation, be split into d linear systems each of order L. The error due to this splitting tends to zero if $M \to \infty$.

We have

$$\begin{aligned} \frac{1}{M} \sum_{m=1}^{M} \left[{}_m\Delta\hat{u} - ({}_mr)({}_m\hat{u})h - \sum_{i=1}^{d} ({}_mZ^i)({}_m\Delta W^i) \right]^2 &= \frac{1}{M} \sum_{m=1}^{M} ({}_m\Delta\hat{u} - h({}_mr)({}_m\hat{u}))^2 \\ -\frac{2}{M} \sum_{m=1}^{M}\sum_{i=1}^{d} ({}_m\Delta\hat{u} - h({}_mr)({}_m\hat{u}))({}_mZ^i)({}_m\Delta W^i) &+ \frac{1}{M} \sum_{m=1}^{M}\sum_{i,j=1}^{d} ({}_mZ^i)({}_mZ^j)({}_m\Delta W^i)({}_m\Delta W^j), \end{aligned} \tag{5.8}$$

where

$${}_mZ^i = \sum_{l=1}^{L} \beta_l^i({}_m\zeta_l) = \sum_{l=1}^{L} \beta_l^i \zeta_l({}_mX_t).$$

Due to the central limit theorem, we have approximately

$$\frac{1}{M} \sum_{m=1}^{M}\sum_{i,j=1}^{d} ({}_mZ^i)({}_mZ^j)({}_m\Delta W^i)({}_m\Delta W^j) \simeq h \sum_{i=1}^{d} E(Z^i)^2,$$

where the error tends to zero if $M \to \infty$. By the same reasoning, we have

$$\frac{1}{M} \sum_{m=1}^{M}\sum_{i=1}^{d} ({}_mZ^i)^2 h \simeq h \sum_{i=1}^{d} E(Z^i)^2,$$

hence the last summand in (5.8) can be replaced by $\frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{d}({}_mZ^i)^2 h$. Now it is not difficult to see that problem (5.7) is close to the following minimization problem

$$\hat{\beta}_l^i = \arg\min_{\beta_l^i} \frac{1}{M} \sum_{m=1}^{M}\sum_{i=1}^{d} \left[\frac{{}_m\Delta\hat{u} - ({}_mr)({}_m\hat{u})h}{h}({}_m\Delta W^i) - \sum_{l=1}^{L} \beta_l^i({}_m\zeta_l) \right]^2 . \tag{5.9}$$

In turn, (5.9) is equivalent to d separate problems

$$\hat{\beta}_l^i = \arg\min_{\beta_l^i} \frac{1}{M} \sum_{m=1}^{M} \left[\frac{{}_m\Delta\hat{u} - ({}_m r)({}_m\hat{u})h}{h} ({}_m\Delta W^i) - \sum_{l=1}^{L} \beta_l^i ({}_m\zeta_l) \right]^2, \; i = 1, \dots, d. \tag{5.10}$$

Each of the i-th problem (5.10), $i = 1, \dots, d$, gives a linear algebraic system of order L with respect to β_l^i, $l = 1, \dots, L$.

In the approach based on (5.4) we get all the deltas at once, but the approach based on (5.6) allows for more efficient computation.

Concerning the selection of a suitable h, it is difficult (or even impossible) to give some recommendations on a theoretical basis. In practice, however, this issue can be settled during computations. In any case it is clear that h cannot be chosen too small because of the terms $({}_m\Delta\hat{u})({}_m\Delta W^i)/h$ in (5.10). Heuristically, these terms have the error $\rho \sim \varepsilon/\sqrt{h}$ if the error of $\hat{u}$ is of $\sim \varepsilon$. Hence, if e.g., $h \sim \varepsilon^\kappa$, $\kappa > 0$, then $\rho \sim \varepsilon^{1-\kappa/2}$, i.e., κ should be less than 2.

6. Testing

All procedures of evaluating prices and in particular sensitivities of Bermudan options contain a lot of errors of different nature and it is almost impossible to take them satisfactorily into account. We therefore propose a test procedure which allows for estimating the quality of either procedure. Equalities (5.1)-(5.2) or the equivalent equality

$$\begin{aligned} du &= \frac{1}{B_t}(u(X_t) - \sum_{j=1}^{d} \frac{\partial u}{\partial x^j}(X_t) X_t^j) dB + \sum_{j=1}^{d} \frac{\partial u}{\partial x^j}(X_t) dX^j \\ &= \frac{1}{B_t}(u(X_t) - [\partial_t(X_t)]^\intercal X_t) dB + [\partial_t(X_t)]^\intercal dX \end{aligned} \tag{6.1}$$

give good opportunities for such testing.

Let X be the solution of (2.1) starting from x_0 at t_0 and $H_t(x)$ be a d-dimensional vector-function. Consider the hedging strategy $\varphi := (H_t^0, H_t(X_t))$ with

$$H_t^0 = \frac{1}{B_t}(V_t - H_t^\intercal(X_t) X_t),$$

where V_t is determined by the equation

$$dV = \frac{1}{B_t}(V - H_t^\intercal(X_t) X_t) dB + H_t^\intercal(X_t) dX, \; V_{t_0} = u_{t_0}(x_0).$$

Clearly, the corresponding portfolio has the form

$$V_t(\varphi) = H_t^0 B_t + H_t^\intercal(X_t) X_t = \frac{1}{B}(V_t - H_t^\intercal(X_t) X_t) B_t + H_t^\intercal(X_t) X_t$$

and the strategy φ is self-financing. It is not difficult to prove that $V_t = u_t$ if and only if $H_t(x) = \partial_t(x)$.

Let $\hat{u}_t(X_t)$ and $\hat{\partial}_t(X_t)$ be constructed due to some procedure. Consider the equation

$$dV = \frac{1}{B_t}(V - [\hat{\partial}_t(X_t)]^\top X_t)dB + [\hat{\partial}_t(X_t)]^\top dX,\ V_{t_0} = \hat{u}_{t_0}(x_0), \tag{6.2}$$

and the portfolio

$$V_t = \frac{1}{B_t}(V_t - [\hat{\partial}_t(X_t)]^\top X_t)B_t + [\hat{\partial}_t(X_t)]^\top X_t. \tag{6.3}$$

Let T_ν be the first time from $\{T_0,...,T_I\}$ for which $\hat{u}$ coincides with f, i.e., $\hat{u}_{T_i}(X_{T_i}) > f_i(X_{T_i})$ for $i = 0,...,\nu-1$ and $\hat{u}_{T_\nu}(X_{T_\nu}) = f_\nu(X_{T_\nu})$. Now let us compare V_t and $\hat{u}_t(X_t)$ on the interval $[T_0,T_\nu]$. Tightness of these quantities is a necessary condition for the considered procedure to be 'good'. As a criterion of tightness one may take, for instance, the difference at $T_\nu:\ V_{T_\nu} - \hat{u}_{T_\nu}(X_{T_\nu}) = V_{T_\nu} - f_\nu(X_{T_\nu})$, or some integral criterion on the time interval $[T_0,T_\nu]$. Of course, we assume that $\hat{u}$ is a satisfactorily accurate estimate of u. The above mentioned testing is rather efficient because we need $\hat{u}_t(x)$ and $\hat{\partial}_t(x)$ only on an individual trajectory.

If we are equipped at any position (t,x) with $\hat{u}_t(x)$ and $\hat{\partial}_t(x)$ constructed by one of the regression methods (Section 4) or by using the direct method of least squares (Section 5) basing on a sample set of trajectories, we may introduce some new criterions of tightness. In such cases we apply again equation (6.2) to any trajectory of the considered set or to a new independent set of trajectories. The latter is necessary if deltas are constructed via direct least squares. After computing ${}_lV_{T_{\nu(l)}}$ for every trajectory ${}_lX_t,\ l = 1,...,L$, it is natural to introduce, for example, the following criterions

$$Cr^2 := [E(V_{T_\nu} - f_\nu(X_{T_\nu}))^2]^{1/2} \simeq \left(\frac{1}{L}\sum_{l=1}^{L}({}_lV_{T_{\nu(l)}} - f_{\nu(l)}({}_lX_{T_{\nu(l)}}))^2\right)^{1/2},$$

$$Cr^+ := E(V_{T_\nu} - f_\nu(X_{T_\nu}))^+ \simeq \frac{1}{L}\sum_{l=1}^{L}({}_lV_{T_{\nu(l)}} - f_{\nu(l)}({}_lX_{T_{\nu(l)}}))^+ ,$$

$$Cr^- := E(f_\nu(X_{T_\nu}) - V_{T_\nu})^+ \simeq \frac{1}{L}\sum_{l=1}^{L}(f_{\nu(l)}({}_lX_{T_{\nu(l)}}) - {}_lV_{T_{\nu(l)}})^+ .$$

7. Calibration

Given market prices $\{C_1,\dots,C_n\}$ at $t = 0$ for a set of liquid benchmark options (typically call options with different strikes K_i and maturities T_i) one looks for a risk-neutral model $\mathbb{Q}$ which prices these options correctly:

$$C_i = E^{\mathbb{Q}}\left[B_i^{-1}(X_{T_i} - K_i)^+\right], \quad i = 1,\dots,n.$$

The idea is the following: one calibrates a risk-neutral model to a set of observed market prices of liquid options and then uses this model to price exotic, illiquid or 'Over The Counter' (OTC) options and to compute hedge ratios. If we parameterize our class of models by $\theta \in \Theta$ (e.g. (2.1) with $\sigma(t,X) = \sigma(t,X;\theta)$), then a practical solution to the calibration problem is to minimize the in-sample weighted quadratic pricing error:

$$O(\theta) = \sum_{i=1}^{n} w_i |C^{\theta}(T_i, K_i) - C_i|^2, \tag{7.1}$$

where $C^{\theta}(T_i, K_i)$ are the model prices and w_i are nonnegative weights. Given that the number of calibration constraints (option prices) is finite, there may be many models reproducing the option prices with equal precision, which means that the error object function (7.1) may have many local minima. Typically, the error landscape may have flat regions in which the error has a low sensitivity to variations in model parameters. Since a good hedging performance is a desired feature of the model to be calibrated (which is not taken into account in (7.1)), it is natural to consider instead of O another objective function

$$O^H(\theta) = \sum_{i=1}^{n} w_i |C^{\theta}(T_i, K_i) - C_i|^2 + \alpha Cr^2(\theta), \quad \alpha > 0$$

and define

$$\hat{\theta} = \arg \inf_{\theta \in \Theta} O^H(\theta),$$

where the “lack of hedge” $Cr^2(\theta)$ can be cheaply computed using regression as described in the previous section.

REFERENCES

[1] L. Andersen, M. Broadie (2001). A primal-dual simulation algorithm for pricing multidimensional American options. Working paper, Columbia Business School, New York.

[2] D. Belomestny, C. Bender, J. Schoenmakers (2006). True upper bounds via non-nested Monte Carlo, to appear in Mathematical Finance.

[3] D. Belomestny, G.N. Milstein (2006). Monte Carlo evaluation of American options using consumption processes. International Journal of Theoretical and Applied Finance, **9**, No. 4, 1-27.

[4] D. Belomestny, G.N. Milstein (2005). Adaptive simulation algorithms for pricing American and Bermudan options by local analysis of financial market. WIAS-Preprint No. 1022, Berlin.

[5] D. Belomestny, G.N. Milstein, V. Spokoiny (2006). Regression methods in pricing American and Bermudan options using consumption processes. WIAS-Preprint No. 1145, Berlin.

[6] C. Bender, J. Schoenmakers (2006). An iterative method for multiple stopping: convergence and stability. Advances in Appl. Prob., Volume 38, Number 3, 729-749.

[7] P. Boyle, M. Broadie, P. Glasserman (1997). Monte Carlo methods for security pricing. Journal of Economic Dynamics and Control, **21**, 1267-1321.

[8] M. Broadie, P. Glasserman (1997). Pricing American-style securities using simulation. J. of Economic Dynamics and Control, **21**, 1323-1352.

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

[10] 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.

[11] D. Duffie, Dynamic Asset Pricing Theory. Princeton University Press 2001

[12] D. Egloff (2005). Monte Carlo algorithms for optimal stopping and statistical learning. Ann. Appl. Probab. **15**, no. 2, 1396-1432.

[13] R. Elie, J.-D. Fermanian, N. Touzi (2007). Kernel estimation of Greek weights by parameter randomization. To appear in Ann. Appl. Probab.

[14] J. Fan, I. Gijbels (1996). Local Polynomial Modelling and Its Applications. Chapman & Hall.

[15] P. Glasserman (2004). Monte Carlo Methods in Financial Engineering. Springer.

[16] M. Giles, P. Glasserman (2006). Smoking adjoints: fast Monte Carlo Greeks. Risk, January, 88-92.
[17] E. Gobet, J.-P. Lemor, X. Warin (2005). A regression-based Monte Carlo method to solve backward stochastic differential equations. Ann. Appl. Probab. **15**, no. 3, 2172-2202.
[18] L. Györfi, M. Kohler, A. Krzyżak, H. Walk (2002). A Distribution-Free Theory of Nonparametric Regression. Springer.
[19] J. Kampen, A. Kolodko, J. Schoenmakers (2007). Monte Carlo Greeks for financial products via approximative Greenian kernels. WIAS-Preprint No. 1208, Berlin.
[20] M. Haugh, L. Kogan (2004). Pricing American options: a duality approach. Operations Research, **52**, No. 2, 258–270.
[21] Karatzas I., Shreve S.E. Methods of Mathematical Finance. Springer, 1998.
[22] A. Kolodko, J. Schoenmakers (2006). Iterative construction of the optimal Bermudan stopping time. Finance and Stochastics, **10**, No. 1, 27-49.
[23] N.V. Krylov (1977). Contollable processes of diffusion type. Nauka, Moscow.
[24] D. Lamberton, B. Lapeyre (1996). Introduction to Stochastic Calculus Applied to Finance. Chapman & Hall.
[25] F.A. Longstaff, E.S. Schwartz (2001). Valuing American options by simulation: a simple least-squares approach. Review of Financial Studies, **14**, 113-147.
[26] G.N. Milstein, J.G.M. Schoenmakers (2002). Monte Carlo construction of hedging strategies against multi-asset European claims. Stochastics and Stochastics Reports, **73**(1-2), 125-157.
[27] G.N. Milstein, M.V. Tretyakov (2005). Numerical analysis of Monte Carlo evaluation of Greeks by finite differences. J. of Computational Finance, **8**, No. 3, 1-33.
[28] G.N. Milstein, M.V. Tretyakov (2006). Practical variance reduction via regression for simulating diffusions. Rep. No. MA-06-019, Univ. of Leicester.
[29] V.V. Piterbarg (2004). Risk sensitivities of Bermuda swaptions. International Journal of Theoretical and Applied Finance, **7**, No. 4, 465-509.
[30] L.C.G. Rogers (2001). Monte Carlo valuation of American options. Mathematical Finance, **12**, 271-286.
[31] J. Schoenmakers (2005). Robust Libor Modelling and Pricing of Derivative Products. Chapman & Hall/CRC.
[32] A. N. Shiryayev (1978). Optimal Stopping Rules. Berlin Heidelberg New York: Springer-Verlag.
[33] A.N. Shiryayev (1999). Essentials of Stochastic Finance: Facts, Models, Theory. World Scientific.
[34] J. Tsitsiklis, B. Van Roy (1999). Regression methods for pricing complex American style options. IEEE Trans. Neural. Net., **12**, 694-703.