Monte Carlo Greeks for financial products via approximative transition densities

Jörg Kampen[1], Anastasia Kolodko[1], and John Schoenmakers[1]

December 13, 2007

Abstract

In this paper we introduce efficient Monte Carlo estimators for the valuation of high-dimensional derivatives and their sensitivities ("Greeks"). These estimators are based on an analytical, usually approximative representation of the underlying density. We study approximative densities obtained by the WKB method. The results are applied in the context of a Libor market model.

Keywords: Financial derivatives, sensitivities, Monte-Carlo methods, WKB expansions.

2000 AMS subject classification: 60H10, 62G07, 65C05

1 Introduction

Valuation methods for high-dimensional derivative products are typically based on Monte Carlo simulation of the underlying process. The dynamics of the underlyings are usually given via a (jump-)diffusion SDE. In case of a diffusion SDE, the underlying process may be simulated using an Euler scheme or a (weak) second order scheme e.g. see Kloeden & Platen [24] or Milstein & Tretyakov [30]. For simulation of jump-diffusions see e.g. Cont & Tankov [7], and Glasserman & Merener [16] for simulation of (Libor) interest rate models with jumps.

The evaluation of option sensitivities, 'Greeks' in financial terms, comes down to the the computation of expressions of the form $\frac{\partial}{\partial\lambda}E(f(X^\lambda))$ (and possibly higher order derivatives), where f is a pay-off function, X is the state of an underlying process depending on some parameter λ. For example, the first and second order derivatives with respect the initial state are called Deltas and Gammas, respectively. In the literature the evaluation of Greeks has been treated by several methods (a nice overview about classical and recent literature is provided in Elie, Fermanian, Touzi [11]). Classical finite difference approaches have been studied by L' Ecuyer & Perron [10], Broadie & Glasserman [4], Milstein & Schoenmakers [29], Milstein & Tretyakov [31], Detemple, Garcia, Rindisbacher [8], and Giles & Glasserman [15]. These approaches are quite

[1]Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstr. 39, D-10117 Berlin, Germany. `kampen[kolodko][schoenma]@wias-berlin.de`.
Supported by the DFG Research Center MATHEON 'Mathematics for Key Technologies' in Berlin.

general and easy to implement as they do not require particular knowledge of the distribution of the underlying. However, they require full blown simulation of the corresponding system of stochastic differential equations and, in order to be efficient, some degree of regularity with respect to the pay-off function. In case the transition kernel of X is known or known in a good approximation, the latter drawback can be avoided by differentiating this kernel with respect to the sensitivity parameter λ, see Fries & Kampen [13] and Fries & Joshi [14]. The typical difficulty that the distribution of the underlying is only known for very special cases was overcome by Fournié, Lasry, Lebuchoux, Lions, Touzi [12], who used the Malliavin integration-by-parts formula in order to express Greeks in the form $E(f(X^{\lambda})\pi)$, where the random variable π is called a *Greek weight.* In a more recent alternative approach Elie, Fermanian, Touzi [11] construct Greek estimators which are based on variance minimizing choices of Greek weights. As a matter of fact, the Malliavin method does not lead to this optimal weight in general. In order to avoid straightforward SDE simulation in the context of the Libor market interest rate model, and so reducing simulation costs, Kurbanmuradov, Sabelfeld & Schoenmakers [27] considered lognormal approximations for the transition density, whereas Hunter, Jäckel, & Joshi [19], and Pelsser, Pietersz & van Regenmortel [32] propose specific drift approximations.

In an ideal situation, the density of the underlying process X^{λ} at a fixed point in time is known explictly and an efficient method to sample from it is available. Usually, however, neither of this is true. Even if the transition density is known, we will show that calculation of sensitivities, based on kernel differentiation for instance (as in [13]), may cause problems (high variance) in case the kernel under consideration is 'highly peaked ', for example due to small maturities, low volatilities, or high dimensionality of the underlying system. In this paper we therefore choose for a rather general approach with the following objectives.

- Developing efficient variance bounded probabilistic representations for price sensitivities, based on an analytical approximation of the underlying density and a possibly rougher approximative standard density (e.g. a lognormal density) which is basically used as an importance sampler.

- Construction of a "good" analytical approximation for the density of the underlying process by using (convergent) WKB[1] methods;

We underline that, in principle, the way of constructing an analytical approximation of the transition density is not essential for the developed Greek estimators. In this article we exploit the use of WKB approximations as a generic convergent method. In special cases, however, construction of high accuracy transition kernels may be possible by other means (see [27] for example).

The structure of the paper is as follows. In Section 2 we set up the model class for which we exemplify our methods and specify the financial products (including Bermudan callables) for which prices and sensitivities are to be determined. In Section 3 we introduce probabilistic representations for integral functionals of kernel type and their derivatives. As a particular result we prove that the corresponding estimator for the derivatives has non-exploding variance

[1] The historical origin of the name is the work of **W**entzel, **K**ramers, **B**rioullin in the context of semiclassical solutions of the Schrödinger equation. The meaning of WKB has broadened since; nowadays, it refers to analytic expansions of exponential form.

for sharply peaked kernels in contrast to some existing weighted Monte Carlo schemes. This estimator thus allows for efficient Monte Carlo estimation of option sensitivities, in particular with respect to underlyings (Deltas), even in situations where the densities are sharply peaked (for instance when volatilities are small). The general probabilistic representations introduced in Section 3 are applied to the computation of Deltas for Bermudan callable products in Section 4. Section 5 deals with the WKB-theory of densities of diffusion equations (densities of processes which have continuous paths). A remark about extensions to analytic expansions of densities of Feller processes is included. These extensions will be considered in detail in Kampen [22]. In Section 5.1 we summarize some results concerning pointwise valid WKB-representations of densities obtained in Kampen [21]. Since in practice only finitely many terms of a WKB expansion can be computed, it will be necessary to use a truncated form of the WKB-representation for actual computations. In Section 5.2. we analyze the effect of this truncation error on approximations of solutions of Cauchy problems and their derivatives.The case of non-autonomous diffusion models is discussed in Section 5.3. The results of Sections 2-5 are applied in Section 6 to the Libor market model. In Section 6.1. we compute explicitly the first three coefficients of the WKB representation of the Libor model density. In Section 6.2 we compute prices and Deltas in a case study of European swaptions.

2 Basic setup

Let $X = (X^1, ..., X^n)$ be a Markovian process of financial quantities in $\mathbb{R}^n_+$ ($\mathbb{R}_+ := \{x : x > 0\}$) under a given pricing measure P, connected with a given discounting numeraire B, $B > 0$, on some filtered probability space. For example, X may represent a system of asset prices or (Libor) interest rates. A popular framework for the system (X, B) is, for instance, the class of jump-diffusions (e.g. Cont & Tankov [7]). For simplicity however, we mainly consider in the present article ordinary diffusions, but, note that the main results generally extend to jump processes as well (see Remark 12 for example).

With respect to an n-dimensional standard Wiener process $W = (W^1, ..., W^n)^\top$ on the probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[t_0,T]}, P)$, where as usual $(\mathcal{F}_t)$ is the P-augmentation of the filtration generated by W, we assume that X is governed by the stochastic differential equation (SDE),

$$\frac{dX^i}{X^i} = \mu(t, X)dt + \sum_{j=1}^{n} \sigma^{ij}(t, X)dW^j, \quad 1 \leq i, j \leq n. \tag{1}$$

It is assumed that $\mu(t, x)$ and the matrix $\sigma(t, x) = \left(\sigma^{ij}(t, x)\right)$, $t \in [t_0, T]$, $x \in \mathbb{R}^n_+$ are such that for all $x_0 \in \mathbb{R}^n_+$, there exists a unique solution $t \to X_t \in \mathbb{R}^n_+$ of (1) for $t_0 \leq t \leq T$ satisfying $X_{t_0} = x_0 =: X^{t_0,x_0}_{t_0}$. It is further assumed that the Markov process X has a transition density

$$p(t, x, s, y), \quad t_0 \leq s \leq t \leq T, \quad x, y \in \mathbb{R}^n_+, \tag{2}$$

which is differentiable with respect to x, y, s, and t, up to any order. In order to guarantee the existence and uniqueness of (1), and the existence of the transition density (2) as stated, it is sufficient to require that the functions $\mu(\cdot, \cdot)$ and

$\sigma(\cdot,\cdot)$ are bounded and have bounded derivatives up to any order, and that the volatility matrix $\sigma(t,x)$ is regular with

$$0 < \lambda_1 \leq \left| \left(\sigma\sigma^\top\right)(t,x) \right| \leq \lambda_2 \tag{3}$$

for all (t,x), $t \in [t_0,T]$, $x \in \mathbb{R}^n_+$, and some $0 < \lambda_1 < \lambda_2$ (see for example Bally & Talay [3]).

Let us take (w.l.o.g.) $B_0 = 1$ and consider a contingent claim with pay-off function of the form $f\ (X_\tau)B_\tau$ at some $(\mathcal{F}_\cdot)$-stopping time τ. By general arguments (e.g. Duffie [9]), the price of this claim at time t_0 is given by

$$u(t_0,x_0) = E\ f(X_\tau^{t_0,x_0}).$$

For deterministic τ, say $\tau \equiv T$, we have a European claim, and for $t_0 \leq t \leq T$ its discounted value process can be represented by

$$u_t := u(t,X_t) := E^{\mathcal{F}_t} f(X_T) = \int p(t,X_t,T,y) f(y) dy,$$

where

$$u(t,x) = \int p(t,x,T,y) f(y) dy \tag{4}$$

is the unique solution of the Cauchy problem

$$\frac{\partial u}{\partial t} + \frac{1}{2}\sum_{i,j=1}^{n} x^i x^j \left(\sigma\sigma^\top\right)^{ij}(t,x)\frac{\partial^2 u}{\partial x^i \partial x^j} + \sum_{i=1}^{n} x^i \mu(t,x)\frac{\partial u}{\partial x^i} = 0, \tag{5}$$

$$u(T,x) = f(x).$$

The density kernel $p(\cdot,\cdot,T,y)$ is the unique (weak) solution of (5) with $p(T,x,T,y)$ $= \delta(x-y)$, where δ is the Dirac-delta function in Schwarz distribution sense.

Of particular importance are Bermudan callable contracts. A Bermudan contract starting at t_0, is specified by a set of exercise dates $\{t_1, t_2, ..., t_\mathcal{I}\}$, where $t_0 < t_1 < ... < t_\mathcal{I} < T$, and corresponding (discounted) pay-off functions $f_i(x)$, $1 \leq i \leq \mathcal{I}$. According to the contract, the holder has the right to call (once) a cash-flow $f_i(X_{t_i}^{t_0,x_0})B_{t_i}^{t_0,x_0,1}$ (with $B_0^{t_0,x_0,1} = 1$) at an exercise date t_i of his choice. It is well known (e.g. Duffie [9]) that the discounted price of this contract at time t, $t_0 \leq t \leq T$, assuming that no exercise took place before t, is given by

$$u(t,x) := \sup_{\tau \in \mathcal{T}_{i,\mathcal{I}}} Ef_\tau(X_\tau^{t,x}) = Ef_{\tau_*^{t,x}}(X_{\tau_*^{t,x}}^{t,x}), \quad t_{i-1} < t \leq t_i, \tag{6}$$

where $x = X_t^{t_0,x_0}$, $\mathcal{T}_{i,\mathcal{I}}$ the set of stopping times τ taking values in $\{t_i, t_{i+1}, ..., t_\mathcal{I}\}$, and $\tau_*^{t,x}$ is an optimal stopping time. In particular the process $u(t,X_t)$ is a supermartingale and is called the Snell envelope of the (discounted) cash-flow process $f_i(X_{t_i})$.

3 Probabilistic representations and their estimators

In this section we consider for a given smooth function $u : \mathbb{R}^n_+ \to \mathbb{R}_+$ and a smooth kernel function $p : \mathbb{R}^n_+ \times \mathbb{R}^n_+ \to \mathbb{R}_+$, probabilistic representations for the integral

$$
\begin{aligned}
I(x) &:= \int p(x,y)u(y)dy, \quad \text{and its gradient} \\
\frac{\partial I}{\partial x}(x) &= \int \frac{\partial}{\partial x}p(x,y)u(y)dy, \quad \text{with} \quad \frac{\partial}{\partial x} := \left(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}\right).
\end{aligned} \tag{7}
$$

Here and in the following sufficient (uniform) integrability conditions are assumed to be fulfilled, for instance, in order to guarantee that (7) is valid.

Remark 1. In (7), kernel p (which may or may not be a density in the second argument) and function u have to be distinguished from the respective definitions in Section 2, although they may be related. For fixed t, T, $0 \le t \le T$, one could take (see (4-5)), $p(x,y) := p(t,x,T,y)$ and $u(x) := u(t,x)$ for example.

Let ζ be an $\mathbb{R}_+^n$-valued random variable on some probability space with density ϕ, $\phi > 0$. Then, obviously,

$$
I(x) = E\ p(x,\zeta)\frac{u(\zeta)}{\phi(\zeta)} \tag{8}
$$

is a probabilistic representation for (7) which may be estimated by the unbiased Monte Carlo estimator

$$
\widehat{I}(x) := \frac{1}{M}\sum_{m=1}^{M} p(x,{}_m\zeta)\frac{u({}_m\zeta)}{\phi({}_m\zeta)}, \tag{9}
$$

where for $m = 1, ..., M$, ${}_m\zeta$ are i.i.d. samples from a distribution with density ϕ. By taking gradients in (8) we readily obtain the probabilistic representation

$$
\frac{\partial I}{\partial x}(x) = E\ \frac{\partial}{\partial x}p(x,\zeta)\frac{u(\zeta)}{\phi(\zeta)}, \tag{10}
$$

with corresponding estimator,

$$
\widehat{\frac{\partial I}{\partial x}}(x) := \frac{1}{M}\sum_{m=1}^{M} \frac{\partial}{\partial x}p(x,{}_m\zeta)\frac{u({}_m\zeta)}{\phi({}_m\zeta)}. \tag{11}
$$

While as a rule (9) is an effective estimator for $I(x)$ for a proper choice of ϕ, unfortunately the gradient estimator (11) has a serious drawback: If the kernel $p(x,\cdot)$ is sharply peaked (nearly proportional to a 'delta-function'), its variance may be extremely high. This fact is demonstrated by the following stylistic example of a multi-asset model, which is nevertheless realistic in orders of magnitude.

Example 2. Consider for fixed $x_0 \in \mathbb{R}_+^n$, parameters $s > 0$, and $\sigma > 0$, the n-dimensional lognormal density

$$
p(s,\sigma;x_0,y) := \frac{1}{(2\pi\sigma^2 s)^{n/2}}\prod_{i=1}^{n}\frac{\exp\left[-\frac{1}{2\sigma^2 s}\ln^2\frac{y^i}{x_0^i}\right]}{y^i}. \tag{12}
$$

In (12) $p(s,\sigma;x_0,\cdot)$ is the density of the random variable

$$
(x_0^1 e^{\sigma\sqrt{s}\xi^1}, ..., x_0^n e^{\sigma\sqrt{s}\xi^n}),
$$

where ξ^i , $i = 1, ..., d$, are i.i.d. standard normal random variables. Thus, for small s and σ, $p(s, \sigma; x_0, \cdot)$ is peaked ('delta-shaped') around x_0. Let us now take $\phi(\cdot) := p(s, \sigma; x_0, \cdot)$ in (8) and (10), respectively, and $u \equiv ||x_0||$ (a constant of order x_0 in magnitude). Clearly, estimator (9) equals $||x_0||$ almost surely and so has zero variance. However, estimator (11) is not deterministic and we have

$$\begin{aligned}\widehat{\frac{\partial I}{\partial x^j}}(x_0) &:= \frac{1}{M} \sum_{m=1}^{M} \frac{||x_0||}{p(s, \sigma; x_0, {}_m\zeta)} \frac{\partial}{\partial x^j} p(s, \sigma; x_0, {}_m\zeta) \\ &= \frac{||x_0||}{M} \sum_{m=1}^{M} \frac{\partial}{\partial x^j} \ln p(s, \sigma; x_0, {}_m\zeta) \\ &= \frac{||x_0||}{M} \sum_{m=1}^{M} \frac{\ln \frac{{}_m\zeta^j}{x_0^j}}{\sigma^2 s x_0^j} = \frac{||x_0||}{M} \sum_{m=1}^{M} \frac{{}_m\xi^1}{\sigma \sqrt{s} x_0^j}.\end{aligned}$$

Hence, $E\left[\widehat{\frac{\partial I}{\partial x^j}}(x_0)\right] = 0$ as should be, but,

$$\operatorname{Var}\left[\widehat{\frac{\partial I}{\partial x^j}}(x_0)\right] = \frac{||x_0/x_0^j||^2}{M} \frac{1}{\sigma^2 s} \tag{13}$$

which explodes when $\sigma^2 s$ goes to zero!

Remark 3. In Fries & Kampen [13] estimators (9) and (11) are used for computing prices and sensitivities of European Libor options, respectively. In their numerical examples they used 50% (rather high) volatility in order to amplify Monte Carlo errors. While, indeed, a larger volatility generally gives rise to a large Monte Carlo error of (9), Example 2 shows that the opposite is true for estimator (11). For example, 50% volatility in combination with 0.5 yr. maturity corresponds to a (just moderate) variance factor $1/\left(\sigma^2 s\right) = 8.0$ in (13), while a more usual Libor volatility, e.g. 14%, and 0.5 y maturity would give a factor 102.0(!).

In the present paper we propose sensitivity estimators which are efficient on a broad time and volatility scale. As a result, the next theorem provides a tool for constructing sensitivity (gradient) estimators with non-exploding variance.

Theorem 4. *Let λ be a reference density on $\mathbb{R}^n$ with $\lambda(z) \neq 0$ for all z (for example, the standard normal density). Let ξ be an $\mathbb{R}^n$-valued random variable on some probability space, with density λ and $g : \mathbb{R}^n_+ \times \mathbb{R}^n \to \mathbb{R}^n_+$ be a smooth enough map which has at least continuous derivatives with $|\partial g(x, z)/\partial z| \neq 0$, such that for each $x \in \mathbb{R}^n_+$ the random variable $\zeta^x := g(x, \xi)$ has a density $\phi(x, \cdot)$ on $\mathbb{R}^n_+$. Then, for (7) we have the probabilistic representation*

$$\frac{\partial I}{\partial x}(x) = E \frac{\partial}{\partial x} \frac{p(x, \zeta^x) u(\zeta^x)}{\phi(x, \zeta^x)} = E \frac{\partial}{\partial x} \frac{p(x, g(x, \xi)) u(g(x, \xi))}{\phi(x, g(x, \xi))}, \tag{14}$$

with corresponding Monte Carlo estimator

$$\widehat{\frac{\partial I}{\partial x}}(x) = \frac{1}{M} \sum_{m=1}^{M} \frac{\partial}{\partial x} \frac{p(x, g(x, {}_m\xi)) u(g(x, {}_m\xi))}{\phi(x, g(x, {}_m\xi))}. \tag{15}$$

Let $\|\cdot\|_\alpha := \sqrt[\alpha]{E\,|\cdot|^\alpha}$ *where* $|\cdot|$ *denotes either a vector norm or a compatible matrix norm. Then it holds*

$$E\left|\frac{\partial}{\partial x}\frac{p(x,g(x,\xi))u(g(x,\xi))}{\phi(x,g(x,\xi))}\right|^2 \le 2M_2^2M_3^2M_4^2+4M_1^2M_4^2M_5^2+4M_1^2M_3^2M_4^2M_6^2, \tag{16}$$

hence the second moments of the Monte Carlo samplers for the components of $\partial I/\partial x$ *are bounded by the right-hand-side of (16), if for fixed* $x\in\mathbb{R}_+^n$*, there are constants* $\alpha_1,...,\alpha_6>1$ *and* $M_1,...,M_6>0$ *with*

$$\frac{1}{\alpha_4}+\frac{1}{\alpha_1}+\frac{1}{\alpha_5}=1,\quad \frac{1}{\alpha_4}+\frac{1}{\alpha_2}+\frac{1}{\alpha_3}=1,\quad \frac{1}{\alpha_4}+\frac{1}{\alpha_1}+\frac{1}{\alpha_6}+\frac{1}{\alpha_3}=1,$$

such that,

$$\begin{gathered} \|u(g(x,\xi))\|_{2\alpha_1}\le M_1,\quad \left\|\frac{\partial u}{\partial y}(g(x,\xi))\right\|_{2\alpha_2}\le M_2,\\ \left\|\frac{\partial g}{\partial x}(x,\xi)\right\|_{2\alpha_3}\le M_3,\quad \left\|\frac{p(x,g(x,\xi))}{\phi(x,g(x,\xi))}\right\|_{2\alpha_4}\le M_4, \\ \left\|\left(\frac{1}{p}\frac{\partial p}{\partial x}-\frac{1}{\phi}\frac{\partial \phi}{\partial x}\right)(x,g(x,\xi))\right\|_{2\alpha_5}\le M_5,\ \left\|\left(\frac{1}{p}\frac{\partial p}{\partial y}-\frac{1}{\phi}\frac{\partial \phi}{\partial y}\right)(x,g(x,\xi))\right\|_{2\alpha_6}\le M_6. \end{gathered} \tag{17}$$

Proof. For any bounded measurable $\psi:\mathbb{R}_+^n\to\mathbb{R}$, we have

$$\int \psi(g(x,z))\lambda(z)dz = \int \psi(g(x,z))\phi(x,g(x,z))\left|\frac{\partial g(x,z)}{\partial z}\right|dz.$$

Therefore, the densities ϕ and g are connected via the relationship

$$\phi(x,g(x,z))\left|\frac{\partial g(x,z)}{\partial z}\right| = \lambda(z). \tag{18}$$

By (18), the right-hand-side of (14) equals

$$\begin{aligned} \frac{\partial}{\partial x}E\,\frac{p(x,g(x,\xi))u(g(x,\xi))}{\phi(x,g(x,\xi))} &= \frac{\partial}{\partial x}\int \frac{p(x,g(x,z))u(g(x,z))}{\phi(x,g(x,z))}\lambda(z)dz \\ &= \frac{\partial}{\partial x}\int p(x,g(x,z))u(g(x,z))\left|\frac{\partial g(x,z)}{\partial z}\right|dz \\ &= \frac{\partial}{\partial x}\int p(x,y)u(y)dy = \frac{\partial I}{\partial x}(x). \end{aligned}$$

To prove the moment estimation (16), we observe that

$$\begin{aligned} &E\left|\frac{\partial}{\partial x}\frac{p(x,\zeta^x)u(\zeta^x)}{\phi(x,\zeta^x)}\right|^2 = E\left|\frac{\partial}{\partial x}\frac{p(x,g(x,\xi))u(g(x,\xi))}{\phi(x,g(x,\xi))}\right|^2 \\ &= E\left|u(g(x,\xi))\frac{\partial}{\partial x}\frac{p(x,g(x,\xi))}{\phi(x,g(x,\xi))}+\frac{p(x,g(x,\xi))}{\phi(x,g(x,\xi))}\frac{\partial u}{\partial y}(g(x,\xi))\frac{\partial g}{\partial x}(x,\xi)\right|^2 \\ &\le 2E\,\frac{p^2(x,g(x,\xi))}{\phi^2(x,g(x,\xi))}\left|\frac{\partial u}{\partial y}(g(x,\xi))\right|^2\left|\frac{\partial g}{\partial x}(x,\xi)\right|^2 \\ &\quad + 2E\,u^2(g(x,\xi))\left|\frac{\partial}{\partial x}\frac{p(x,g(x,\xi))}{\phi(x,g(x,\xi))}\right|^2 \\ &=: 2(I)+2(II). \end{aligned}$$

Then by Hölders inequality, $(I) \leq$

$$\sqrt[\alpha_2]{E\left|\frac{\partial u}{\partial y}(g(x,\xi))\right|^{2\alpha_2}}\sqrt[\alpha_3]{E\left|\frac{\partial g}{\partial x}(x,\xi)\right|^{2\alpha_3}}\sqrt[\alpha_4]{E\frac{p^{2\alpha_4}(x,g(x,\xi))}{\phi^{2\alpha_4}(x,g(x,\xi))}} \leq M_2^2 M_3^2 M_4^2.$$

For the second term we have $(II) =$

$$\begin{aligned}
&E\,u^2(g(x,\xi))\,\frac{p^2(x,g(x,\xi))}{\phi^2(x,g(x,\xi))}\left|\frac{\frac{\partial p}{\partial x}(x,g(x,\xi))}{p(x,g(x,\xi))}-\frac{\frac{\partial \phi}{\partial x}(x,g(x,\xi))}{\phi(x,g(x,\xi))}\right.\\
&\left.+\left(\frac{\frac{\partial p}{\partial y}(x,g(x,\xi))}{p(x,g(x,\xi))}-\frac{\frac{\partial \phi}{\partial y}(x,g(x,\xi))}{\phi(x,g(x,\xi))}\right)\frac{\partial g}{\partial x}(x,\xi)\right|^2\\
&\leq 2E\,u^2(g(x,\xi))\,\frac{p^2(x,g(x,\xi))}{\phi^2(x,g(x,\xi))}\left|\frac{\frac{\partial p}{\partial x}(x,g(x,\xi))}{p(x,g(x,\xi))}-\frac{\frac{\partial \phi}{\partial x}(x,g(x,\xi))}{\phi(x,g(x,\xi))}\right|^2\\
&+2E\,u^2(g(x,\xi))\,\frac{p^2(x,g(x,\xi))}{\phi^2(x,g(x,\xi))}\left|\frac{\frac{\partial p}{\partial y}(x,g(x,\xi))}{p(x,g(x,\xi))}-\frac{\frac{\partial \phi}{\partial y}(x,g(x,\xi))}{\phi(x,g(x,\xi))}\right|^2\left|\frac{\partial g}{\partial x}(x,\xi)\right|^2\\
&\leq 2M_1^2M_4^2M_5^2+2M_1^2M_3^2M_4^2M_6^2,
\end{aligned}$$

again by Hölders inequality. ■

Remark 5. If in (17) the random variables $u(g(x,\xi))$, $\frac{\partial u}{\partial y}(g(x,\xi))$, and so on, have moments of high enough order, Theorem 4 guarantees that the variance of estimator (15) is controlled via the moment estimates (17). The most delicate bound in (17) is M_5 in fact. Indeed, if one takes $g(x,\xi) \equiv g(x_0,\xi)$ estimator (15) collapses to (11), and in Example 2 where $\phi(x,y) \equiv p(x_0,y)$ in fact, we see that M_5 cannot taken to be small when $\sigma^2 s$ is small, i.e. when p is highly peaked around x. In contrast, if for fixed x, $\phi(x,\cdot)$ is approximately proportional to $p(x,\cdot)$ and $\partial \ln\phi(x,\cdot)/\partial x \approx \partial \ln p(x,\cdot)/\partial x$ (both with respect to the weight function $\phi(x,\cdot)$), a small M_5 may exist. Note that for $\phi(\cdot,\cdot)$ exactly proportional to $p(\cdot,\cdot)$, we may take $M_5 = 0$.

Remark 6. It can be shown that Theorem 4 can be extended to probabilistic representations and corresponding estimators for higher order derivatives,

$$\frac{\partial I}{\partial x^\beta}(x) = E\,\frac{\partial}{\partial x^\beta}\frac{p(x,\zeta^x)u(\zeta^x)}{\phi(x,\zeta^x)} = E\,\frac{\partial}{\partial x^\beta}\frac{p(x,g(x,\xi))u(g(x,\xi))}{\phi(x,g(x,\xi))},$$

with corresponding Monte Carlo estimator

$$\widehat{\frac{\partial I}{\partial x^\beta}}(x) = \frac{1}{M}\sum_{m=1}^{M}\frac{\partial}{\partial x^\beta}\frac{p(x,g(x,_m\xi))u(g(x,_m\xi))}{\phi(x,g(x,_m\xi))}, \tag{19}$$

where $\beta := (\beta_1,\ldots,\beta_n)$, $\beta_i \in \{0,1,2,\ldots\}$ is a multi-index with (formally) $\partial x^\beta = \partial x_1^{\beta_1}\partial x_2^{\beta_2}\cdots\partial x_n^{\beta_n}$. Loosely speaking, the variance of the higher order derivative estimator (19) can be bounded from above by an expression like (16) involving (i) sufficiently high moments of the derivatives,

$$y \to \frac{\partial u}{\partial y^\gamma}, \quad \text{and} \quad z \to \frac{\partial g(x,z)}{\partial z^\gamma}$$

for fixed x, $\gamma \leq \beta$ (component wise), with respect to weight functions $y \to \phi(x, y)$ and $z \to \lambda(z)$, respectively, and, (ii) for fixed x, $L^q(\mathbb{R}^n_+, \phi(x, y)dy)$-norms of

$$y \to \frac{\partial}{\partial x^\gamma}\left(\frac{\phi(x,y)}{p(x,y)}\right), \quad \text{and} \quad y \to \frac{\partial}{\partial y^\gamma}\left(\frac{\phi(x,y)}{p(x,y)}\right), \quad \gamma \leq \beta,$$

for q large enough.

Remark 7. In the next section we consider financial applications where $I(x)$ is the price of a derivative contract considered in dependence of the argument x which may stand for the underlying process or some parameter (vector) which affects the dynamics of the underlying process (e.g. volatilities). Moreover, we there give a recipe how to construct a lognormal density approximation ϕ, corresponding to a particular normal reference density and an exponential type transformation (in Theorem 4, λ and g respectively).

Remark 8. Theorem 4 can be can be generalized to the case where both p and ϕ live on a common submanifold rather than on the whole state space. Via such a generalization it would be possible to extend our results to factor reduced situations in the spirit of Fries & Kampen [13] and Fries & Joshi [14]. However, this is considered beyond the scope of the present article and therefore we restrict ourselves to the full-factor case.

4 Sensitivities for Bermudan options

Theorem 4 may be applied in general for computing sensitivities ("Greeks") of derivative products. For estimator (11) the danger of exploding variance is typically the largest when derivatives of prices with respect to underlyings (Deltas, Gammas) are considered. We therefore consider in this section only (first order) derivatives with respect to the underlying process, hence Deltas.

Let $\tau : \Omega \to \mathbb{R}_+$ be a given stopping time with respect to the filtration $(\mathcal{F}_\cdot)$. As usual we may think of Ω as being the space of functions $\omega : [0, \infty) \to \mathbb{R}^n$, which are continuous from the right and have limits from the left, and define $\tau^{s,x}(\omega) := s + \tau(X^{s,x}_{s+(\cdot)}(\omega))$. We now consider the Bermudan contract introduced in Section 2. For fixed t, t^+, $t_0 \leq t \leq t^+ \leq t_1$, $x \in \mathbb{R}^n_+$, we have $\tau^{t,x}_* = \tau^{t^+, X^{t,x}_{t^+}}_*$ since $\tau^{t,x}_* \geq t_1$, and we thus may write

$$\begin{aligned} u(t,x) &:= Ef(X^{t,x}_{\tau^{t,x}_*}) = EE^{\mathcal{F}_{t^+}} f(X^{t^+, X^{t,x}_{t^+}}_{\tau^{t^+, X^{t,x}_{t^+}}_*}) \\ &= \int p(t, x, t^+, y) dy Ef(X^{t^+, y}_{\tau^{t^+, z}_*}) \\ &= \int p(t, x, t^+, y) u(t^+, y) dy, \end{aligned}$$

by the Chapman-Kolmogorov equation.

For each t, t^+ as above, let $\phi(t, x, t^+, y)$, $g(t, x, t^+, y)$, and reference density $\lambda(t, t^+, z)$ be as in Theorem 4. We then have the probabilistic representation

$$u(t,x) = E\, \frac{p(t,x,t^+, g(t,x,t^+,\xi))}{\phi(t,x,t^+, g(t,x,t^+,\xi))} f(X^{t^+, g(t,x,t^+,\xi)}_{\tau^{t^+, g(t,x,t^+,\xi)}_*}), \tag{20}$$

with Monte Carlo estimator

$$\widehat{u}(t,x) := \frac{1}{M}\sum_{m=1}^{M}\frac{p(t,x,t^+,g(t,x,t^+,{}_m\xi))}{\phi(t,x,t^+,g(t,x,t^+,{}_m\xi))}f(X^{t^+,g(t,x,t^+,\ {}_m\xi)}_{\tau_*^{t^+,g(t,x,t^+,\ {}_m\xi)}}), \tag{21}$$

and for the gradients (Deltas) we have the probabilistic representation

$$\Delta_i := \frac{\partial u}{\partial x^i}(t,x) = E\ \frac{\partial}{\partial x^i}\left(\frac{p(t,x,t^+,g(t,x,t^+,\xi))}{\phi(t,x,t^+,g(t,x,t^+,\xi))}f(X^{t^+,g(t,x,t^+,\xi)}_{\tau_*^{t^+,g(t,x,t^+,\xi)}})\right) \tag{22}$$

with Monte Carlo estimator

$$\widehat{\Delta}_i := \frac{1}{M}\sum_{m=1}^{M}\frac{\partial}{\partial x^i}\left(\frac{p(t,x,t^+,g(t,x,t^+,{}_m\xi))}{\phi(t,x,t^+,g(t,x,t^+,{}_m\xi))}f(X^{t^+,g(t,x,t^+,\ {}_m\xi)}_{\tau_*^{t^+,g(t,x,t^+,\ {}_m\xi)}})\right), \tag{23}$$

where ${}_m\xi$, $m = 1, ..., M$, are i.i.d. samples from the reference density λ. Indeed, by pre-conditioning on $\mathcal{F}_{t^+}$ and then taking expectations we see that (21) and (23) are unbiased Monte Carlo estimators for the price (20) and 'deltas' (22), respectively. Moreover, if ϕ is close to p in the sense of Theorem 4, it is not difficult to see that also gradient estimator (23) has non-exploding variance when $t^+ \downarrow t$.

Estimators (21) and (23) are useful if one has an analytic approximation $\widehat{p}(t,x,t^+,y)$ of the density $p(t,x,t^+,y)$ and known densities $\phi(x,\cdot)$ for $x \in \mathbb{R}^n_+$. The approximation $\widehat{p}$ may be obtained by some specific method, for example by a WKB expansion as presented in Section 5, or some lognormal approximation as proposed in Kurbanmuradov, Sabelfeld & Schoenmakers [27] for the Libor market model. Of course the density ϕ has to be chosen with some care. If it is possible to sample directly from $\widehat{p}$ (e.g. in case of a log-normal approximation) we may take $\phi = \widehat{p}$. If not, (e.g. in the case of a WKB expansion) one may take for ϕ a (not necessarily very accurate) lognormal approximation of the density p.

A canonical lognormal approximation for $p(t,x,t^+,z)$ is obtained by freezing X in the coefficients of (1) at the initial time. We thus obtain

$$\begin{aligned} \overset{\text{lgn}}{X}{}^{t,x;i}_{t^+} &:= x^i \exp\left(-\frac{1}{2}\sum_{j=1}^{n}\int_t^{t^+}(\sigma^{ij})^2(s,x)ds + \int_t^{t^+} r(s,x)ds \right.\\ &\qquad \left. + \sum_{j=1}^{n}\int_t^{t^+}\sigma^{ij}(s,x)dW^j_s\right) =: x_i\exp(\xi_i). \end{aligned} \tag{24}$$

Here, $(\xi_i)_{i=1}^n$ is a Gaussian random vector with

$$E\xi_i = -\frac{1}{2}\sum_{j=1}^{n}\int_t^{t^+}(\sigma^{ij})^2(s,x)ds + \int_t^{t^+} r(s,x)ds =: \mu^{i;t,t^+,x}, \quad 1 \le i \le n,$$

and

$$Cov(\xi_i,\xi_j) = \sum_{l=1}^{n}\int_t^{t^+}\sigma^{il}(s,x)\sigma^{jl}(s,x)\,ds =: \sigma^{ij;t,t^+,x}, \quad 1 \le i,j \le n.$$

Clearly, the density ϕ is then given by

$$\phi(t,x,t^+,y) := \frac{\psi_{\mu^{t,t^+,x},\sigma^{t,t^+,x}}(\ln\frac{y^1}{x^1},\ln\frac{y^2}{x^2},...,\ln\frac{y^n}{x^n})}{y^1y^2\cdots y^n}, \tag{25}$$

$y^i>0,\ 1\le i\le n$, with $\psi_{\mu^{t,t^+,x},\sigma^{t,t^+,x}}$ being the density of the n-dimensional normal distribution $\mathcal{N}_n(\mu^{t,t^+,x},\sigma^{t,t^+,x})$ with $\mu^{t,t^+,x} := (\mu^{i;t,t^+,x})_{1\le i\le n}$ and $\sigma^{t,t^+,x} := (\sigma^{ij;t,t^+,x})_{1\le i,j\le n}$.

For practical applications it is most useful to discretize estimator (23) to

$$\begin{aligned}\widehat{\Delta}_i^h := \frac{1}{M}\sum_{m=1}^{M}\frac{1}{2h}\Bigg(&\frac{p(t,x+\mathfrak{h}_i,t^+,g(t,x+\mathfrak{h}_i,t^+,{}_m\xi))}{\phi(t,x+\mathfrak{h}_i,t^+,g(t,x+\mathfrak{h}_i,t^+,{}_m\xi))}f(X^{t^+,g(t,x+\mathfrak{h}_i,t^+,\ {}_m\xi)}_{\tau_*^{t^+,g(t,x+\mathfrak{h}_i,t^+,\ {}_m\xi)}})\\ &-\frac{p(t,x-\mathfrak{h}_i,t^+,g(t,x-\mathfrak{h}_i,t^+,{}_m\xi))}{\phi(t,x-\mathfrak{h}_i,t^+,g(t,x-\mathfrak{h}_i,t^+,{}_m\xi))}f(X^{t^+,g(t,x-\mathfrak{h}_i,t^+,\ {}_m\xi)}_{\tau_*^{t^+,g(t,x-\mathfrak{h}_i,t^+,\ {}_m\xi)}})\Bigg),\end{aligned} \tag{26}$$

where $\mathfrak{h}_i := h(\delta_{i1},\ldots,\delta_{in})$ (δ_{ij} being the Kronecker symbol), for small enough $h>0$. Without further details we note that according to Milstein and Tretyakov (2004) in a related context, it is efficient to take $h\approx x/\sqrt{M}$.

As an alternative, it is also possible to expand the derivatives in (23) path-wise. Since the set of exercise dates is discrete, for almost sure all ω the function $y\to\tau_*^{t^+,y}(\omega)$ is constant in the region $\{y :\mid y-y_0\mid<\epsilon_0\}$ for fixed y_0 and ϵ_0 small enough (y_0 and ϵ_0 may depend on ω). This can be seen from the fact that the random variable $X^{t^+,y}_{\tau_*^{t^+,y}}(\omega)$, which has a density in $\mathbb{R}^n$, lays almost sure in the interior of the exercise region, and, the fact that, almost surely, for a fixed exercise date t_i, the map $y\to X^{t^+,y}_{t_i}$ is continuous (e.g. see Protter (1990)). Thus, by differentiating (23) path-wise we obtain

$$\begin{aligned}\widehat{\Delta}_i := &\frac{1}{M}\sum_{m=1}^{M} f(X^{t^+,g(t,x,t^+,\ {}_m\xi)}_{\tau_*^{t^+,g(t,x,t^+,\ {}_m\xi)}})\frac{\partial}{\partial x^i}\left(\frac{p(t,x,t^+,g(t,x,t^+,{}_m\xi))}{\phi(t,x,t^+,g(t,x,t^+,{}_m\xi))}\right)\\ &+\frac{1}{M}\sum_{m=1}^{M}\frac{p(t,x,t^+,g(t,x,t^+,{}_m\xi))}{\phi(t,x,t^+,g(t,x,t^+,{}_m\xi))}\frac{\partial f}{\partial z}(X^{t^+,g(t,x,t^+,\ {}_m\xi)}_{\tau_*^{t^+,g(t,x,t^+,\ {}_m\xi)}})\cdot\\ &\cdot\partial_y X^{t^+,y}_{\tau_*^{t^+,g(t,x,t^+,\ {}_m\xi)}}(g(t,x,t^+,\ {}_m\xi))\frac{\partial g(t,x,t^+,\ {}_m\xi)}{\partial x^i},\end{aligned} \tag{27}$$

where

$$\frac{\partial}{\partial x^i}\frac{p(t,x,t^+,y)}{\phi(t,x,t^+,y)},\qquad \frac{\partial g(t,x,t^+,\ {}_m\xi)}{\partial x^i},\qquad \frac{\partial f}{\partial z}$$

can in principle be expressed analytically, and the vector process

$$\partial_y X_s^{t^+,y}(\cdot) := \frac{\partial X_s^{t^+,y}}{\partial y}(\cdot),\qquad s\ge t^+,$$

can in principle be simulated via a variational system of SDEs (e.g. see Protter [34], Milstein & Schoenmakers [29], Giles & Glasserman [15]).

In this paper we will prefer the discretized version (26) of (23) for our applications. The algorithm is as follows. We first choose an $h>0$, and sample ${}_m\xi$ for $m=1,\ldots,M$ from the reference (usually normal) density. Next we simulate for each m a pair of trajectories ${}_mX^{\pm}$, which start in ${}_mg^{\pm} := g(t, x \pm h, t^+, {}_m\xi)$ at t^+, and end at the optimal stopping times ${}_m\tau_*^{\pm} := \tau_*^{t^+, {}_mg^{\pm}}$. Of course the optimal exercise dates ${}_m\tau_*^{\pm}$ are generally unknown in practice, but we assume that we have good approximations ${}_m\tau^{\pm}$ at hand, which are constructed via some well known procedure. For example, in a pre-computation we may construct an exercise boundary via a regression method (e.g. Longstaff & Schwartz [28]), or as an alternative, we may use the policy iteration method of Kolodko & Schoenmakers [25], see also Bender & Schoenmakers [5]. As discussed above, for a particular ω we have ${}_m\tau_*^{+} = {}_m\tau_*^{-}$ provided that h is small enough. For this reason we take in our simulations simply ${}_m\tau^- ={}_m \tau^+$, where τ is some approximation of the optimal exercise policy. This pragmatic assumption is justified if the probability of the event ${}_m\tau^- \neq_m \tau^+$ is small enough, i.e. h is small enough. For each m we compute also the values ${}_mp^{\pm} := p(t, x \pm \mathfrak{h}, t^+, {}_mg^{\pm})$ and ${}_m\phi^{\pm} := \phi(t, x \pm \mathfrak{h}, t^+, {}_mg^{\pm})$, and finally compute the estimate

$$\widehat{\Delta}_i^h := \frac{1}{M} \sum_{m=1}^{M} \frac{1}{2h} \left(\frac{{}_mp^+}{{}_m\phi^+} f({}_mX^+) - \frac{{}_mp^-}{{}_m\phi^-} f({}_mX^-) \right). \tag{28}$$

Remark 9. In the previous sections vector and matrix components are denoted by superscripts, so that time parameters of processes can be denoted by subscripts. In the next sections we depart from this convention and use subscripts for vector and matrix components.

5 WKB approximations for transition densities

5.1 Recap of WKB theory

We summarize some results concerning WKB-expansions of parabolic equations (cf. Kampen [21] for details). Let us consider the parabolic diffusion operator

$$\tfrac{\partial u}{\partial t} + Lu \equiv \tfrac{\partial u}{\partial t} + \tfrac{1}{2} \sum_{i,j} a_{ij} \tfrac{\partial^2 u}{\partial x_i \partial x_j} + \sum_i b_i \tfrac{\partial u}{\partial x_i}. \tag{29}$$

For simplicity of notation and without loss of generality it is assumed that the diffusion coefficients a_{ij} and the first order coefficients b_i in (29) depend on the spatial variable x only. In the following let $\delta t := T - t$, and let the functions

$$(x, y) \to d(x, y) \geq 0, \quad (x, y) \to c_k(x, y), \;\; k \geq 0,$$

be defined on $\mathbb{R}^n \times \mathbb{R}^n$, with d^2 and c_k, $k \geq 0$, being smooth. Then a set of (simplified) conditions sufficient for pointwise valid WKB-representations of the form

$$p(t, x, T, y) = \frac{1}{\sqrt{2\pi\delta t}^n} \exp\left(-\frac{d^2(x, y)}{2\delta t} + \sum_{k=0}^{\infty} c_k(x, y) \delta t^k \right), \tag{30}$$

for the solution $(t, x) \to p(t, x, T, y)$ of the final value problem

$$\begin{aligned} \frac{\partial p}{\partial t} + Lp &= 0, \quad \text{with final value} \\ p(T, x, T, y) &= \delta(x - y), \quad y \in \mathbb{R}^n \quad \text{fixed}, \end{aligned} \tag{31}$$

is given by

(A) The operator L is uniformly elliptic in $\mathbb{R}^n$, i.e. as in (3) the matrix norm of $(a_{ij}(x))$ is bounded below and above by $0 < \lambda < \Lambda < \infty$ uniformly in x,

(B) the smooth functions $x \rightarrow a_{ij}(x)$ and $x \rightarrow b_i(x)$ and all their derivatives are bounded.

For more subtle (and partially weaker conditions) we refer to Kampen [21]. If we add the uniform boundedness condition

(C) there exists a constant c such that for each multiindex α and for all $1 \leq i, j, k \leq n$,
$$\left|\frac{\partial^\alpha a_{jk}}{\partial x^\alpha}(x)\right|, \ \left|\frac{\partial^\alpha b_i}{\partial x^\alpha}(x)\right| \leq c \exp\left(c|x|^2\right), \tag{32}$$

then the Taylor expansions of the functions d and c_k around $y \in \mathbb{R}^n$ are equal to d and $c_k, k \geq 0$ globally. I.e. we have the power series representations

$$d^2(x,y) = \sum_\alpha d_\alpha(y)\delta x^\alpha \tag{33}$$

$$c_k(x,y) = \sum_\alpha c_{k,\alpha}(y)\delta x^\alpha, \qquad k \geq 0, \tag{34}$$

where $\delta x := x - y$. Note that (C) is implied by the stronger condition that all derivatives in (32) have a uniform bound. Summing up we have the following theorem:

Theorem 10. *If the hypotheses (A),(B) are satisfied, then the fundamental solution p has the representation*

$$p(\delta t, x, y) = \frac{1}{\sqrt{2\pi\delta t}^n} \exp\left(-\frac{d^2(x,y)}{2\delta t} + \sum_{k\geq 0} c_k(x,y)\delta t^k\right), \tag{35}$$

where d and c_k are smooth functions, which are unique global solutions of the first order differential equations (36),(37), *and* (39) *below. Especially,*

$$(\delta t, x, y) \rightarrow \delta t \ln p(\delta t, x, y) = -\frac{n}{2}\delta t \ln(2\pi\delta t) - \frac{d^2}{2} + \sum_{k\geq 0} c_k(x,y)\delta t^{k+1}$$

is a smooth function which converges to $-\frac{d^2}{2}$ as $\delta t \searrow 0$, where d is the Riemannian distance induced by the line element $ds^2 = \sum_{ij} a_{ij}^{-1} dx_i dx_j$, where with a slight abuse of notation (a_{ij}^{-1}) denotes the matrix inverse of (a_{ij}). If the hypotheses (A),(B) and (C) are satisfied, then in addition the functions d, $c_k, k \geq 0$ equal their Taylor expansion around y globally, i.e. we have (33)-(34).

The recursion formulas for d and c_k, $k \geq 0$ are obtained by plugging the ansatz (30) into the parabolic equation (31), and ordering terms with respect to the monoms $\delta t^i = (T - t)^i$ for $i \geq -2$. By collecting terms of order δt^{-2} we obtain
$$d^2 = \frac{1}{4}\sum_{ij} d^2_{x_i} a_{ij} d^2_{x_j}, \tag{36}$$

where $d^2_{x_k}$ denotes the derivative of the function d^2 with respect to the variable x_k, with the boundary condition $d(x,y)=0$ for $x=y$. Collecting terms of order δt^{-1} yields

$$-\frac{n}{2}+\frac{1}{2}Ld^2+\frac{1}{2}\sum_i\left(\sum_j\left(a_{ij}(x)+a_{ji}(x)\right)\frac{d^2_{x_j}}{2}\right)\frac{\partial c_0}{\partial x_i}(x,y)=0, \tag{37}$$

where the boundary condition

$$c_0(y,y)=-\frac{1}{2}\ln\sqrt{\det\left(a_{ij}(y)\right)} \tag{38}$$

determines c_0 uniquely for each $y\in\mathbb{R}^n$. Finally, for $k+1\geq 1$ we obtain

$$\begin{aligned}&(k+1)c_{k+1}(x,y)+\tfrac{1}{2}\textstyle\sum_{ij}a_{ij}(x)\Big(\frac{d^2_{x_i}}{2}\frac{\partial c_{k+1}}{\partial x_j}+\frac{d^2_{x_j}}{2}\frac{\partial c_{k+1}}{\partial x_i}\Big)\\ &=\tfrac{1}{2}\textstyle\sum_{ij}a_{ij}(x)\sum_{l=0}^{k}\frac{\partial c_l}{\partial x_i}\frac{\partial c_{k-l}}{\partial x_j}+\frac{1}{2}\sum_{ij}a_{ij}(x)\frac{\partial^2 c_k}{\partial x_i\partial x_j}+\sum_i b_i(x)\frac{\partial c_k}{\partial x_i},\end{aligned} \tag{39}$$

with boundary conditions

$$c_{k+1}(x,y)=R_k(y,y)\ \text{if}\quad x=y, \tag{40}$$

R_k being the right side of (39). For some classical models in finance a global transformation of the diffusion operator to the Laplace operator is possible (at the price of more complicated first order terms however). We observe this in the case of a Libor market model (Section 6). The requirement that a transformation $y(x)$ of an operator with second order coefficients $a_{ij}(x)=(\sigma\sigma^\top)_{ij}(x)$ leads to a Laplacian with respect to second order terms in y is equivalent to

$$\sum_{ml}(\sigma\sigma^\top)_{ml}(x)\frac{\partial y_k}{\partial x_l}\frac{\partial y_j}{\partial x_m}=\delta_{jk} \tag{41}$$

where δ_{jk} denotes the Kronecker delta. If σ is invertible it follows directly that the transformation $y(x)$ satisfies the first order matrix equation

$$\left(\frac{\partial y_k}{\partial x_l}\right)=\left(\sigma^{-1}_{kl}(x)\right). \tag{42}$$

The latter equation determines the transformation (up to constants, of course) but cannot be integrated in general, and if it can not explicitly in general. However, a necessary and sufficient condition for integrability of (42) in terms of σ can be given, where we restrict ourselves to the case of invertible σ.

Proposition 11. *There is a global coordinate transformation for the operator (29) such that the second order part of the transformed operator equals the Laplacian, iff $a_{ij}=(\sigma\sigma^\top)_{ij}$ for a (square) matrix function σ which satisfies*

$$\sum_{l=1}^{n}\frac{\partial\sigma_{ik}(x)}{\partial x_l}\sigma_{lj}(x)=\sum_{l=1}^{n}\frac{\partial\sigma_{ij}(x)}{\partial x_l}\sigma_{lk}(x),\quad x\in\mathbb{R}^n. \tag{43}$$

The latter fact is also observed and proved in Ait-Sahalia [1]. If the condition of Proposition 11 is satisfied, then coordinate transformation leads to second order coefficients of the form $a_{ij} \equiv \delta_{ij}$, so that the solution of (36) becomes

$$d^2(x,y) = \sum_i (x_i - y_i)^2.$$

If conditions (A), (B), (C), and (43) hold, then in the transformed coordinates, explicit formulas for the coefficient functions $c_k, k \geq 0$ can be computed via the formulas

$$c_0(x,y) = \sum_i (y_i - x_i) \int_0^1 b_i(y + s(x-y))ds, \tag{44}$$

and

$$c_{k+1}(x,y) = \int_0^1 R_k(y + s(x-y), y)s^k ds, \tag{45}$$

with R_k being the right-hand-side of (39) where $a_{ij} = \delta_{ij}$. Similar formulas are obtained in Ait-Sahalia [1]. In Kampen [21] it is shown in addition how the coefficients c_k can be computed explicitly in terms of power series approximations of the diffusion coefficients a_{ij} and b_i. However, in high dimensional models such as the Libor market model direct computation of the coefficients c_k seems more feasible as it turns out that the computation up to the coefficient c_1 is sufficient for our purposes. We conclude this Section with a final remark concerning possible generalizations to Feller processes.

Remark 12. Pointwise converging analytic expansion can be obtained for a large class of densities p_F of Feller processes. We refer to remark 11 in [23].

5.2 Error estimates

We now study the approximation error of a truncated WKB expansion (and its derivatives), which is essential for the convergence of the Monte Carlo schemes. In this respect we will show how the derivatives (up to second order) of the product value function with respect to the underlyings computed by means of a truncated WKB-expansion converge in supremum norm and Hölder norms. Let us consider a WKB-approximation of the fundamental solution p of the form

$$p_l(t,x,T,y) = \frac{1}{\sqrt{2\pi\delta t}^n} \exp\left(-\frac{d^2(x,y)}{2\delta t} + \sum_{k=0}^{l} c_k(x,y)\delta t^k\right), \tag{46}$$

i.e. we assume that the coefficients d^2 and $c_k,\ 0 \leq k \leq l$ have been computed up to order l (recall that $\delta t = T - t$ for the sake of brevity). Let us denote the domain of the Cauchy problem by $D = (0,T) \times \mathbb{R}^n$. For integers $n \geq 0$ and real numbers $\delta \in (0,1)$ let $C^{m+\delta/2,n+\delta}(D)$ be the space of m (n) times differentiable functions such that the mth (nth) derivative with respect to time (space) is Hölder continuous with exponent $\frac{\delta}{2}$ (δ). Furthermore, $|.|_{m+\delta/2,n+\delta}$ denote the natural norms associated with these function spaces. It is well-known that in case of our assumptions (A) and (B) the fundamental solution p satisfies the a priori estimate

$$|p(t,x,T,y)| \leq C(T-t)^{-n/2} \exp\left(-\frac{\lambda_0|x-y|^2}{2(T-t)}\right), \tag{47}$$

for some generic constant C and some λ_0 which is less or equal than the lower bound λ in assumption (A) above. We call a WKB-approximation p_l of the fundamental solution p admissible, if it satisfies an a priori estimate of the form (47). The WKB-approximation p_0 is always admissible while the proof of theorem 9 (cf. [21]) that p_l is admissible if $l \geq l_0$ where l_0 is some natural number depending on the coefficient functions and can be computed by comparison of the WKB-expansion and the Levy-expansion. For lower l admissibility has to be ensured for each model. In the Libor market model admissibility for $l = 1$ is ensured. The following consequence of Safanov's theorem (cf. Krylov [26]).

Theorem 13. *Assume that* $(A), (B),$ *and* (C) *are satisfied and let* $h \in C^{2+\delta}(\mathbb{R}^n)$ *and* $f \in C^{\delta/2,\delta}(D)$. *If*

$$c \leq -\lambda \quad \textit{for some} \quad \lambda > 0, \tag{48}$$

then the Cauchy problem

$$\left\{ \begin{array}{l} \frac{\partial w}{\partial t} + \frac{1}{2}\sum_{ij} a_{ij}(x)\frac{\partial^2 w}{\partial x_i \partial x_j} + \sum_i b_i(x)\frac{\partial w}{\partial x_i} + c(x)w = f(t,x) \ \textit{in } D \\ \\ w(T,x) = h(x) \ \textit{for } x \in \mathbb{R}^n \end{array} \right. \tag{49}$$

has a unique solution w, *and there exists a constant* c *depending only on* δ, n λ, Λ *and* $K = \max\{|a|_\delta, |b|_\delta, |c|_\delta\}$ *such that*

$$|w|_{1+\delta/2,2+\delta} \leq c\left[|f|_{\delta/2,\delta} + |h|_{2+\delta}\right]. \tag{50}$$

In order to analyze the truncation error of the Cauchy problem with data h we consider the function

$$u^{\Delta}(t,x) = u(t,x) - u_l(t,x),$$

where

$$u(t,x) = \int_{\mathbb{R}^n} h(y)p(t,x,T,y)dy,$$

and

$$u_l(t,x) = \int_{\mathbb{R}^n} h(y)p_l(t,x,T,y)dy.$$

In order u_l in (5.2) to be a good approximation (and well-defined) we say that u_l is admissible if it satisfies the a priori estimate (47) above, and we say that u_l is admissible if p_l is admissible. It is then possible to derive different error estimates in strong norms depending on which Greeks we want to control on which level of regularity.

Theorem 14. *Assume that conditions (A), (B), and (C) hold and that* $h \in C^{2+\delta}(\mathbb{R}^n)$ *and assume that* u_l *is strongly admissible. Then*

$$|u(t,x,y) - u_l(t,x,y)|_{1+\delta/2,2+\delta} \in O(t^{l-\frac{\delta}{2}}).$$

Proof. Let $w(t,x) = e^{-rt}u^{\Delta}(t,x)$ with r constant and $w_l(t,x) = e^{-rt}u_l(t,x)$. Since

$$\frac{\partial u^{\Delta}}{\partial t} + Lu^{\Delta} = -\frac{\partial u_l}{\partial t} - Lu_l =: f_{u_l}(t,x), \text{ we have}$$

$$\frac{\partial w}{\partial t} + Lw + rw = \frac{\partial w_l}{\partial t} + Lw_l + rw_l =$$

$$e^{rt}\left(-\frac{\partial u_l}{\partial t} - Lu_l\right) =: e^{rt}f_{u_l}(t,x).$$

Admissibility of u_l and an argument similar to that of Krylov ensures that the right side of the latter equation can be measured in the norm $|.|_{\delta/2,\delta}$. Hence we can apply the estimate (50) to the function

$$w(t,x) = e^{-rt}u^{\Delta}(t,x)$$

for a constant $r > 0$ and we get (after dividing by e^{rt}

$$|u^{\Delta}|_{1+\delta/2,2+\delta} \leq c|f_{u_l}|_{\delta/2,\delta}. \tag{51}$$

In order to compute the term on the right side of (51) we can plug (46) into the left-hand side of (31) the parabolic equation satisfied by the exact fundamental solution p. However in order to see how the higher order terms behave exactly we plug in

$$p(t,x,y) = \frac{1}{\sqrt{2\pi t}^n}\exp\left(-\frac{d^2(x,y)}{2t} + \sum_{k=0}^{l} c_k(x,y)t^k + R_{l+1}(t,x,y)\right), \tag{52}$$

where

$$R_{l+1}(t,x,y) = \sum_{k=l+1}^{\infty} c_k(x,y)t^k \in O(t^{l+1}). \tag{53}$$

We get

$$\tfrac{\partial p}{\partial t} + \tfrac{1}{2}\textstyle\sum_{ij} a_{ij}\tfrac{\partial^2 p_l}{\partial x_i \partial x_j} + \textstyle\sum_i b_i \tfrac{\partial p}{\partial x_i} =$$

$$= t^l\Bigg((l+1)c_{l+1} + \tfrac{\partial}{\partial t}R_{l+1} + \tfrac{1}{2}\textstyle\sum_{ij} a_{ij}\Big(- \tfrac{d^2_{x_i}}{2}\left(\tfrac{\partial}{\partial x_j}\tfrac{R_{l+1}}{t}\right) - \tfrac{d^2_{x_j}}{2}\left(\tfrac{\partial}{\partial x_i}\tfrac{R_{l+1}}{t}\right)$$

$$+ (c_{l,x_i} + R_{l+1,x_i})(c_{l,x_j} + R_{l+1,x_j})\Big) + L(c_l + R_l)\Bigg)p \in O(t^l)p$$

Applying a priori estimates for p we get the result. ■

Remark 15. A more intricate analysis shows that for practical purposes it is possible to remove the admissibility condition above, if we approximate the Cauchy problem by a Dirichlet problem with a large but spatially bounded domain (a natural step from the numerical point of view). However, since this involves an additional analysis of an integral equation corresponding to the boundary condition we go not into the details here. Generalizations to estimates which include Taylor-expansions of the WKB-coefficients will also be considered elsewhere. The assumption $h \in C^{2+\delta}(\mathbb{R}^n)$ can be weakened to Hölder continuous pay-offs if we abstain from contolling the Θ Greek (sensitivity with respect to time) up to maturity. The case where c_k are computed up to $k = 1$ is the first case where the truncation error for first and second derivatives converges to zero (in supremum norm with order $O(\delta t)$ and in Hölder- extension of supremum norm with order $O(\delta t)^{1-\frac{\delta}{2}}$). This implies that our Monte Carlo computation scheme for the Greeks converges.

5.3 Extension to diffusions with time-dependent coefficients

The results of Section 5.1 and 5.2. extend to time-dependent diffusions without great difficulties. We write down the formulas for c_0 and c_{k+1}. The justification can be found in [23]. We have

$$c_0(t,x,y) = \sum_i (y_i - x_i) \int_0^1 b_i(t, y + s(x-y))ds, \text{ and} \tag{54}$$

$$c_{k+1}(t,x,y) = \int_0^1 R_k(t, y + s(x-y), y) s^k ds, \tag{55}$$

where

$$R_k(t,x,y) \equiv \frac{\partial c_k}{\partial t} + \frac{1}{2}\sum_i \sum_{l=0}^{k} \frac{\partial c_l}{\partial x_i}\frac{\partial c_{k-l}}{\partial x_i} + \frac{1}{2}\sum_i \frac{\partial^2 c_k}{\partial x_i^2} + \sum_i \mu_i(t,x)\frac{\partial c_k}{\partial x_i}. \tag{56}$$

6 Applications to the Libor market model

We consider a Libor market model with respect to a tenor structure $0 < T_1 \ldots < T_{n+1}$ in the terminal measure P_{n+1} (induced by the terminal zero coupon bond $B_{n+1}(t)$). The dynamics of the forward Libors $L_i(t)$, defined in the interval $[0, T_i]$ for $1 \le i \le n$, are governed by the following system of SDE's (e.g., see Jamshidian [20]),

$$dL_i = -\sum_{j=i+1}^{n} \frac{\delta_j L_i L_j \, \gamma_i^\top \gamma_j}{1 + \delta_j L_j} dt + L_i \, \gamma_i^\top dW^{(n+1)} =: \mu_i(t, L) + L_i \, \gamma_i^\top dW^{(n+1)}, \tag{57}$$

where $\delta_i = T_{i+1} - T_i$ are day count fractions and $t \to \gamma_i(t) = (\gamma_{i,1}(t), \ldots, \gamma_{i,d}(t))$, $0 \le t \le T_i$, are bounded and smooth enough deterministic volatility vector functions. We denote the matrix with rows $\gamma_i^\top$ by Γ and assume that Γ is invertible. In (57), $(W^{(n+1)}(t) \mid 0 \le t \le T_n)$ is a standard d-dimensional Wiener process under the measure P_{n+1} with d, $1 \le d \le n$, being the number of driving factors. In what follows we consider the full-factor Libor model with $d = n$ in the time interval $[0,\, T_1)$.

6.1 WKB approximations for the Libor kernel

Let us transform the dynamics of (57) to $K_i := \ln L_i$, $1 \le i \le n$,

$$dK_i = \frac{1}{L_i} dL_i - \frac{1}{2L_i^2} d\langle L_i \rangle = \left(-\frac{\gamma_i^\top \gamma_i}{2} + \mu_i(t, e^{L_1}, \ldots, e^{L_n}) \right) dt + \gamma_i^\top dW^{(n+1)}. \tag{58}$$

Note that the coefficients of the generator corresponding to process K are bounded and satisfy the WKB assumptions (A) (B) in Section 5.1. Hence, we may apply a WKB approximation to the transition density of the process (58).

By the transformation $Y := \Gamma^{-1}K$ we obtain the process

$$dY_i = \mu_i^Y(t,Y)dt + dW_i^{(n+1)}, \quad 1 \le i \le n, \quad \text{where}$$

$$\mu_i^Y(t,Y) = (MY)_i + V_i + \sum_{j=1}^n \Gamma_{ij}^{-1}\mu_j(t, e^{(\Gamma Y)_1}, \ldots, e^{(\Gamma Y)_n}), \quad \text{and} \tag{59}$$

$$M_{ij} = \sum_{l=1}^n \frac{d\Gamma_{il}^{-1}}{dt}\Gamma_{lj}, \quad V_i = -\sum_{j=1}^n \Gamma_{ij}^{-1}\frac{|\gamma_j|^2}{2},$$

for which the generator has a Laplacian diffusion term, which leads to technically more convenient expressions in the respective WKB expansion. However, in the time dependent case we recommend using WKB density approximation of K rather then Y, as due to the (unbounded) linear term MY in (59) we might need higher order expansions for getting satisfactory results.

The situation of time independent γ (hence bounded μ^Y) is exemplified in case study Section 6.2, where the transition density p^Y is approximated and subsequently transformed to an approximated transition density p^L of the Libor process. Below we spell out the ingredients for computing the corresponding WKB coefficients c_0 and c_1 according to (44) and (45) to be exploited in Section 6.2. Using the notations

$$F_l(s,x,y) := \frac{1}{(\Gamma(x-y))_l}\ln\frac{1+\delta_l e^{(\Gamma x)_l}}{1+\delta_l e^{(\Gamma y)_l}}, \quad 1 \le l \le n,$$

and $a := (\gamma_i^\top \gamma_j)_{i,j=1}^n$, we may write,

$$c_0(s,x,y) = \sum_{i=1}^n V_i(y_i - x_i) + \sum_{i=1}^n\sum_{j=1}^n \Gamma_{ij}^{-1}(y_i - x_i)\sum_{l=j+1}^n a_{jl}F_l(s,x,y), \tag{60}$$

$$\begin{aligned}\frac{\partial c_0}{\partial x_p}(s,x,y) &= -V_p + \sum_{j=1}^n \Gamma_{pj}^{-1}\sum_{l=j+1}^n a_{jl}F_l(s,x,y) - \\ &\quad \sum_{i=1}^n\sum_{j=1}^n \Gamma_{ij}^{-1}(y_i - x_i)\sum_{l=j+1}^n a_{jl}\frac{\partial F_l(s,x,y)}{\partial x_p},\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 c_0}{\partial x_p^2}(s,x,y) &= +2\sum_{j=1}^n \Gamma_{pj}^{-1}\sum_{l=j+1}^n a_{jl}\frac{\partial F_l(s,x,y)}{\partial x_p} - \\ &\quad \sum_{i=1}^n\sum_{j=1}^n \Gamma_{ij}^{-1}(y_i - x_i)\sum_{l=j+1}^n a_{jl}\frac{\partial^2 F_l(s,x,y)}{\partial x_p^2},\end{aligned}$$

where

$$\frac{\partial F_l(s,x,y)}{\partial x_p} = \frac{\Gamma_{lp}(s)}{(\Gamma(x-y))_l}\left(\frac{\delta_l e^{(\Gamma x)_l}}{1+\delta_l e^{(\Gamma x)_l}} - F_l(s,x,y),\right)$$

and

$$\begin{aligned}\frac{\partial^2 F_l(s,x,y)}{\partial x_p^2} &= \frac{2\Gamma_{lp}^2}{(\Gamma(x-y))_l^2}\left(F_l(s,x,y) - \frac{\delta_l e^{(\Gamma x)_l}}{1+\delta_l e^{(\Gamma x)_l}}\right) \\ &\quad + \frac{\Gamma_{lp}^2}{(\Gamma(x-y))_l}\frac{\delta_l e^{(\Gamma x)_l}}{(1+\delta_l e^{(\Gamma x)_l})^2}.\end{aligned}$$

By L'Hospital's Rule we get,

$$\begin{aligned}\lim_{x\to y} F_l(s,x,y) &= \frac{\delta_l e^{(\Gamma y)_l}}{1+\delta_l e^{(\Gamma y)_l}},\\ \lim_{x\to y} \frac{\partial F_l(s,x,y)}{\partial x_p} &= \frac{\Gamma_{lp}\delta_l e^{(\Gamma y)_l}}{2(1+\delta_l e^{(\Gamma y)_l})^2},\\ \lim_{x\to y} \frac{\partial^2 F_l(s,x,y)}{\partial x_p^2} &= \frac{\Gamma_{lp}^2\delta_l e^{(\Gamma y)_l}(1-\delta_l e^{(\Gamma y)_l})}{3(1+\delta_l e^{(\Gamma y)_l})^3}.\end{aligned}$$

We finally obtain $p^L(s,u,t,v)$ by density transformation formula,

$$p^L(s,u,t,v) = p^Y(s, S_s^{-1}(u), t, S_t^{-1}(v)) \left| \frac{\partial S_t^{-1}(v)}{\partial v} \right|$$

with

$$S_t^{-1}(v) := \Gamma^{-1}(t)(\ln v_1, \ldots, \ln v_n)^\top.$$

For simplicity in the case study below we assume that the matrix Γ is upper triangular and does not depend on t. We then have,

$$\begin{aligned} p^L(s,u,t,v) &= \frac{1}{\sqrt{2\pi(t-s)}^n} \prod_{i=1}^n \frac{\Gamma_{ii}^{-1}}{v_i} \exp\left(-\frac{\left(\Gamma^{-1}(\ln \frac{v_1}{u_1}, \ldots, \ln \frac{v_n}{u_n})^\top\right)^2}{2(t-s)} \right.\\ &\qquad \left. + \sum_{k=0}^{\infty} c_k(s, S^{-1}(u), S^{-1}(v))(t-s)^k \right). \end{aligned}$$

6.2 Case study

We now illustrate the estimators in Section 3 and Section 4 by computing European and Bermudan swaptions and Deltas in a Libor market model. A (payer) swaption contract with maturity T_i and strike θ with principal \$1 gives the right to contract at T_i for paying a fixed coupon θ and receiving floating Libor at the settlement dates $T_{i+1}, \ldots, T_n$. The discounted payoff of the contract is thus

$$S_i(L_{T_i}^{0,L(0)}) = \frac{1}{B_{n+1}(T_i)} \sum_{j=i}^n B_{j+1}(T_i) \left(\delta_j L_{T_i}^{0,L(0);j} - \theta\right)^+. \tag{61}$$

Here we use the notation

$$L_{T_i}^{0,L(0);j} := L_j(T_i), \quad 1 \le j \le n.$$

in line with Section 2 and Section 4.

For our experiments we take in Libor model (57), $\delta_i \equiv 0.5$, flat 3.5% initial Libor curve, constant volatility loadings, $\gamma_i(t) \equiv 0.2 e_i$, where e_i are n-dimensional unit vectors decomposing an input correlation matrix ρ,

$$\rho_{ij} = \exp\left[\frac{|j-i|}{n-1} \ln \rho_\infty\right], \quad 1 \le i,j \le n, \tag{62}$$

with $\rho_\infty = 0.3$ (for more general correlation structures we refer to Schoenmakers [36]).

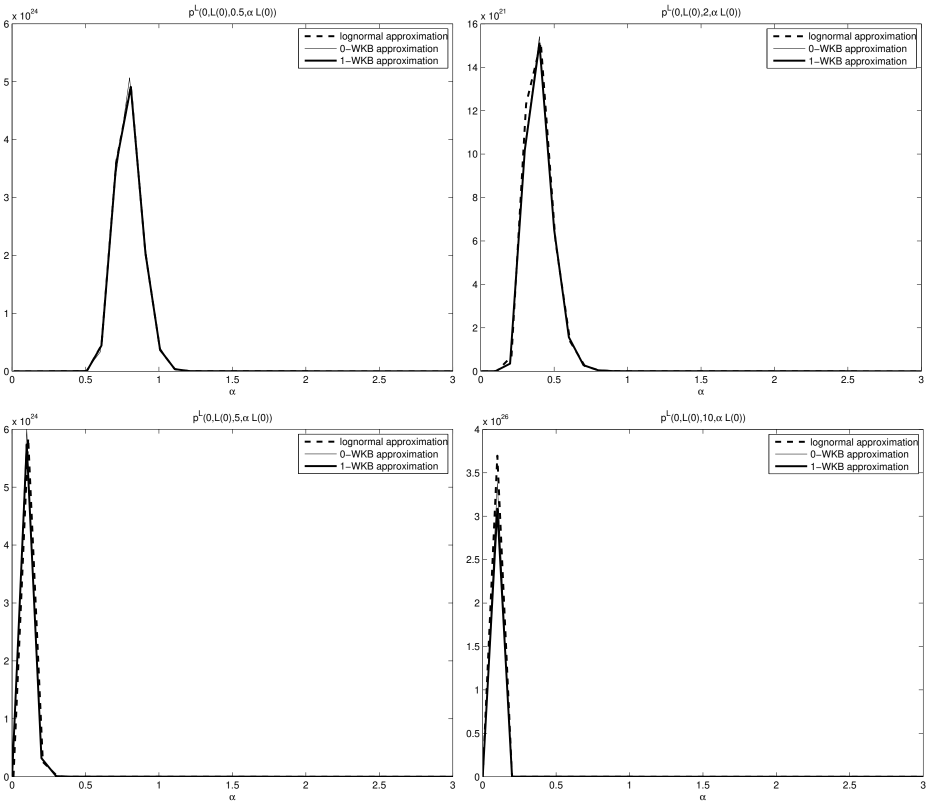

Figure 1: WKB approximations of the Libor transition density (cross-section) $p^L(0, L(0), t, \alpha L(0))$ for different t (thin line for p_0^L, bold line for p_1^L) and lognormal approximation p_{lgn}^L (dashed line).

The Libor transition kernel $p^L(s,x,t,y)$ shows to have a pronounced "delta-shaped" form. See Figure 1 for cross-sections $\alpha \to p^L(0, L(0), t, \alpha L(0))$ of its WKB approximations p_0^L and p_1^L, with c_0 only, and with c_0 and c_1, respectively. Because of this "delta-shape" it is very important for efficiency of the estimators in Section 4 to find a suitable proxy density ϕ. We take for ϕ the transition kernel of a lognormal approximation L^{lgn}, obtained from the Libor process (57) by freezing the coefficients at the initial time s,

$$L_t^{\text{lgn}\, s,x;\, i}(\xi) = x_i \exp(\xi_i), \tag{63}$$

where ξ is a n-dimensional Gaussian vector with

$$\begin{aligned} E\xi_i &= (t-s)\left(\frac{|\gamma_i|^2}{2} - \sum_{j=i+1}^{n} \frac{|\gamma_i||\gamma_j|\rho_{ij}\delta_j x_j}{1+\delta_j x_j}\right) =: \mu_i^{\text{lgn}}(s,t,x), \\ Cov(\xi_i,\xi_j) &= \Gamma_{ij}, \quad 1 \le i,j \le n. \end{aligned} \tag{64}$$

The transition density of L^{lgn} is then given by

$$\begin{aligned} p_{\text{lgn}}^L(s,u,t,v) \quad &= \quad \frac{1}{\sqrt{2\pi(t-s)}^n} \prod_{i=1}^{n} \frac{\Gamma_{ii}^{-1}}{v_i} \times \\ &\exp\left(-\frac{\left|\Gamma^{-1}((\ln\frac{v_1}{u_1}\ldots\ln\frac{v_n}{u_n}) - \mu^{\text{lgn}}(s,t,x))^T\right|^2}{2(t-s)}\right), \end{aligned}$$

with $|\cdot|$ denoting the Euclidean norm.

Example 1: European swaption

We test estimators (9) and (15) by pricing European swaptions and Deltas for different maturities T_1. We consider an at-the-money swaption over the period $[T_1, T_{19}]$. The payoff function is given by

$$u(\cdot) = B_{n+1}(0)S_1(\cdot),$$

where S_1 is defined in (61) with $n = 19$ and $\theta = 3.5\%$. We set $x = L(0)$ and

$$\phi(x,\cdot) = p_{\text{lgn}}^L(0, x, T_1, \cdot).$$

So, in order to sample in (9) from density ϕ, we simulate via (63)-(64) the lognormal samples

$${}_m\zeta = L_{T_1}^{\text{lgn}\, 0,L(0)}({}_m\xi), \quad m = 1,\ldots,M.$$

As a (more accurately approximated) Libor transition kernel, we use WKB approximation p_0^L and p_1^L,

$$p(x,\cdot) \approx p_0^L(0,x,T_1,\cdot) \quad \text{and} \quad p(x,\cdot) \approx p_1^L(0,x,T_1,\cdot),$$

and denote the corresponding estimators by $\widehat{I}_0$ and $\widehat{I}_1$, respectively. For computing Deltas, we take a finite difference analogue of (15),

$$\widehat{\frac{\partial I}{\partial x_i}}^{(h)}(x) := \frac{1}{M}\sum_{m=1}^{M}\frac{1}{2h}\left(\frac{p(x+\mathfrak{h}_i, g(x+\mathfrak{h}_i,\,{}_m\xi))u(g(x+\mathfrak{h}_i,\,{}_m\xi))}{\phi(x+\mathfrak{h}_i, g(x+\mathfrak{h}_i,\,{}_m\xi))}-\right.$$
$$\left.\frac{p(x-\mathfrak{h}_i, g(x-\mathfrak{h}_i,\,{}_m\xi))u(g(x-\mathfrak{h}_i,\,{}_m\xi))}{\phi(x-\mathfrak{h}_i, g(x-\mathfrak{h}_i,\,{}_m\xi))}\right),\ 1\le i\le n, \quad (65)$$

where

$$g(x,{}_m\xi) = \overset{\mathrm{lgn}}{L}{}_{T_1}^{0,x}({}_m\xi)$$

and $\mathfrak{h}_i = h(\delta_{i1},\ldots,\delta_{in})$.

The estimates due to (9) and (65) are compared with corresponding estimates due to "exact" Libor trajectories simulated by a log-Euler scheme with small time steps. For this we simulate M Libor trajectories using $\Delta t = \delta_i/5$, and take

$$\widehat{I}_{\mathrm{ex}} = \frac{1}{M}\sum_{m=1}^{M} u({}_mL_{T_1}^{0,x}), \quad (66)$$
$$\widehat{\frac{\partial I_{\mathrm{ex}}}{\partial x_i}}^{(h)} = \frac{1}{M}\sum_{m=1}^{M}\frac{u({}_mL_{T_1}^{0,x+\mathfrak{h}_i}) - u({}_mL_{T_1}^{0,x-\mathfrak{h}_i})}{2h}, \quad 1\le i\le n,$$

where the subscript $_{\mathrm{ex}}$ refers to "exact" Libor trajectories. Analogue to (66), we compute $\widehat{I}_{\mathrm{lgn}}$ and $\frac{\partial \widehat{I}_{\mathrm{lgn}}}{\partial x_i}$ due to the lognormal approximation L^{lgn}.

In Table 1 and Table 2, we give values of European swaptions and Deltas at time zero, computed via estimators $\widehat{I}_{\mathrm{ex}}$, $\widehat{I}_{\mathrm{lgn}}$, $\widehat{I}_0$, $\widehat{I}_1$ and $\widehat{\frac{\partial I_{\mathrm{ex}}}{\partial x_n}}^{(h)}$, $\widehat{\frac{\partial I_{\mathrm{lgn}}}{\partial x_n}}^{(h)}$, $\widehat{\frac{\partial I_0}{\partial x_n}}^{(h)}$, $\widehat{\frac{\partial I_1}{\partial x_n}}^{(h)}$, respectively, for different maturities T_1. For the computation we take $h = 3.5\times 10^{-5}$, $M = 5\times 10^5$ in order to keep standard deviations within 0.5% relative to the values. As we see, the WKB approximation with only two coefficients, c_0 and $\widetilde{c}_1$, provides a very close estimate of the European swaptions and Deltas, also for large maturities. The distance between $\widehat{I}_{\mathrm{ex}}$ and $\widehat{I}_1$ and $\widehat{\frac{\partial I_{\mathrm{ex}}}{\partial x_n}}$ and $\widehat{\frac{\partial I_1}{\partial x_n}}$ is smaller than 0.5% relative to the value.This is in contrast to canonical lognormal estimators $\widehat{I}_{\mathrm{lgn}}$ and $\frac{\partial \widehat{I}_{\mathrm{lgn}}}{\partial x_i}$, which gives an acceptable approximation only for $T_1 \le 2$.

Table 1. (values in basis points)

T_1	$\widehat{I}_{\mathrm{ex}}$ (SD)	$\widehat{I}_{\mathrm{lgn}}$ (SD)	$\widehat{I}_0$ (SD)	$\widehat{I}_1$ (SD)
1.0	178.9(0.4)	179.0(0.4)	181.6(0.4)	178.9(0.4)
2.0	245.3(0.6)	246.5(0.6)	251.4(0.6)	244.3(0.6)
5.0	351.3(1.0)	359.8(1.0)	376.4(1.1)	352.7(1.0)
10.0	429.6(1.5)	451.4(1.6)	495.6(1.7)	431.8(1.4)

Table 2. (values in basis points)

T_1	$\widehat{\frac{\partial I_{\text{ex}}}{\partial x_n}}^{(h)}$ (SD)	$\widehat{\frac{\partial I_{\text{lgn}}}{\partial x_n}}^{(h)}$ (SD)	$\widehat{\frac{\partial I_0}{\partial x_n}}^{(h)}$ (SD)	$\widehat{\frac{\partial I_1}{\partial x_n}}^{(h)}$ (SD)
1.0	1768.3(2.8)	1774.2(2.8)	1794.9(2.9)	1770.7(2.8)
2.0	1726.4(2.9)	1732.1(2.9)	1729.0(2.9)	1729.0(2.9)
5.0	1599.6(3.2)	1615.9(3.3)	1722.5(3.5)	1597.0(3.2)
10.0	1417.1(3.8)	1474.0(4.2)	1668.0(4.7)	1422.7(3.9)

Example 2: Bermudan swaption

For our next example we take a Bermudan at-the-money swaption with 10 annual exercise possibilities starting from T_1. The set of exercise dates of the contract is thus $\{T_1, T_3, T_5, \ldots, T_{19}\} =: \mathcal{T}$. We compute the values and Deltas for different T_1 using estimators (21) and (26), respectively, with

$$t = 0, \quad t^+ = T_1, \quad x = L(0).$$

As payoff function f in the estimators we here have

$$f_{T_\tau}(L_{T_\tau}^{t^+,x}) = B_{n+1}(0) S_\tau(L_{T_\tau}^{t^+,x}),$$

where S_τ is given by (61) with $n = 19$ and $\theta = 3.5\%$, and τ is a Bermudan stopping time. Similar to the case of European swaptions, we take in (21) and (26) the lognormally approximated Libor transition density,

$$\phi(t, x, t^+, \cdot) = p_{\text{lgn}}^L(t, x, t^+, \cdot).$$

Then,

$$g(t, x, t^+, {}_m\xi) = \overset{\text{lgn}}{L}{}_{t^+}^{t,x}({}_m\xi),$$

where $\overset{\text{lgn}}{L}{}_{t^+}^{t,x}({}_m\xi)$ is the m-th realization of the lognormal random variable (63)-(64).

For the Bermudan swaption, a very good approximation of the optimal stopping policy is constructed by Andersen's method (strategy II, see Andersen [2]),

$$\tau_A^{t^+,x} := \inf \Big\{ i : \; T_i \geq t^+, \, T_i \in \mathcal{T}, \, B_{n+1}(T_i) S_i(L_{T_i}^{t^+,x}) \geq H_i + \max_{j \geq i, T_j \in \mathcal{T}} E^i \left[B_{n+1}(T_j) S_j(L_{T_j}^{t^+,x}) \right] \Big\}. \tag{67}$$

The conditional expectations in (67), which are (discounted) swaption values in fact, can be computed accurately in closed-form (see, e.g., Schoenmakers [36]). Further in (67), H is a constant vector computed by backward optimization over a set of pre-simulated trajectories, as proposed by Andersen [2]. In Table 3, column 2, we display the Bermudan prices $\widehat{u}_{\text{ex}}^{\text{low}}$ due to stopping strategy τ_A. Upper estimations $\widehat{u}_{\text{ex}}^{\text{up}}$ are constructed from $\widehat{u}_{\text{ex}}^{\text{low}}$ by the dual approach, developed in Rogers [35] and Haugh and Kogan [18], see Table 3, column 3. As we see, the distance between lower and upper Bermudan estimates does not exceed 0.5% (relative to the values).

We next compute the estimates (21) and (26) via the stopping strategy τ_A. As in the previous example, we approximate the Libor transition density by WKB approximations p_0^L and p_1^L, respectively, and ϕ by the lognormal approximation p_{lgn}^L. We denote the estimates by $\widehat{u}_0$, $\widehat{u}_1$, $\widehat{u}_{\text{lgn}}$ and $\widehat{\frac{\partial u_0}{\partial x_n}}^{(h)}$, $\widehat{\frac{\partial u_1}{\partial x_n}}^{(h)}$, $\widehat{\frac{\partial u_{\text{lgn}}}{\partial x_n}}^{(h)}$, respectively. The results are compared with "exact" Bermudan values, given by the interval $[\widehat{u}_{\rm ex}^{\rm low}, \widehat{u}_{\rm ex}^{\rm up}]$, and with finite-difference approximations of Deltas,

$$\widehat{\frac{\partial u_{\rm ex}}{\partial x_i}}^{(h)} = \frac{1}{M}\sum_{m=1}^{M} \frac{f_{T_{\tau_A^{0,x}}}({}_mL^{0,x+\mathfrak{h}_i}_{T_{\tau_A^{0,x}}}) - f_{T_{\tau_A^{0,x}}}({}_mL^{0,x-\mathfrak{h}_i}_{T_{\tau_A^{0,x}}})}{2h}, \quad 1 \le i \le n,$$

with $\mathfrak{h}_i = h(\delta_{i1}, \ldots, \delta_{in})$, simulated with "exact" Libor trajectories. Here the "exact" trajectories are simulated with a log-Euler scheme using $\Delta t = \delta_i/10$.

As we see in Table 3 and Table 4, the WKB approximation with only two coefficients c_0 and c_1 involved, gives very close estimates of Bermudan swaption prices and Deltas, even for large values of T_1. We observe that the distance between the values simulated via "exact" Libor trajectories and the corresponding values due to the WKB approximation is smaller than 0.5% relative to the value. For these computations, $h = 3.5 \times 10^{-5}$, $M = 5 \times 10^5$ in order to keep standard deviations within 0.5% relative to the values.

Table 3. (values in basis points)

T_1	$\widehat{u}_{\rm ex}^{\rm low}$ (SD)	$\widehat{u}_{\rm ex}^{\rm up}$ (SD)	$\widehat{u}_{\rm lgn}$ (SD)	$\widehat{u}_0$ (SD)	$\widehat{u}_1$ (SD)
1.0	351.2(0.7)	352.5(1.0)	350.9(0.7)	354.7(0.7)	351.2(0.7)
2.0	388.4(0.8)	389.8(1.0)	388.2(0.8)	396.6(0.8)	387.3(0.8)
5.0	461.5(1.1)	463.4(1.3)	466.3(1.1)	492.9(1.1)	460.8(1.1)
10.0	523.7(1.6)	524.8(1.7)	543.6(1.7)	601.2(1.7)	523.6(1.5)

Table 4. (values in basis points)

T_1	$\widehat{\frac{\partial u_{\rm ex}}{\partial x_n}}^{(h)}$ (SD)	$\widehat{\frac{\partial u_{\rm lgn}}{\partial x_n}}^{(h)}$ (SD)	$\widehat{\frac{\partial u_0}{\partial x_n}}^{(h)}$ (SD)	$\widehat{\frac{\partial u_1}{\partial x_n}}^{(h)}$ (SD)
1.0	2709.2(3.5)	2720.9(3.5)	2747.2(3.5)	2709.2(3.5)
2.0	2631.1(3.5)	2630.5(3.5)	2700.7(3.6)	2628.6(3.5)
5.0	2392.9(3.7)	2407.7(3.8)	2561.9(4.0)	2398.0(3.8)
10.0	2101.5(4.4)	2152.5(4.7)	2443.4(5.3)	2111.5(4.4)

Remark 16. The values in Tables 1–4 are computed using a second order Taylor approximation of $c_1(x,y)$ around x, where $c_1(x,x)$, the derivatives $\frac{\partial c_1}{\partial y_i}(x,x)$ and $\frac{\partial^2 c_1}{\partial y_i \partial y_j}(x,x)$, are computed (using finite differences) prior to the Monte Carlo simulation.

Computational time

By using the new estimators we avoid step-by-step Euler simulation of the Libor process in the time interval $[0, T_1]$. Generally, the cost of Euler stepping up to T_1 is proportional to $T_1/\Delta t$, whereas the cost of the "direct estimators" (15) and (26) is independent of T_1. In particular, in the present Libor case, Euler stepping up to T_1 requires a cost proportional to $n^2 \frac{T_1}{\Delta t}$ times the cost of computing the (possibly virtual) pay-off at T_1. In comparison, the cost of simulating estimators (15) and (26) is proportional to n^2 times the cost of the pay-off at T_1.

In Figure 2 we compare for different T_1 the CPU time (per sample) needed for computing the values in Tables 2,4 using WKB based estimators (15) and (26) with the CPU time required for computing the estimates via straightforward Euler stepping of "exact" Libor trajectories up to T_1. We conclude that,

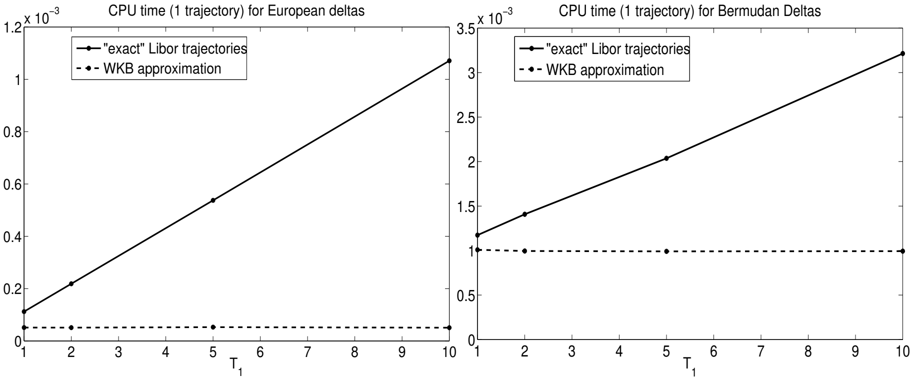

Figure 2: CPU time (seconds) for simulating European (left) and Bermudan (right) Deltas for different T_1 by log-Euler Libor simulation (solid line) and by WKB density approximation (with c_0 and c_1) (dash line).

particularly for larger T_1, the efficiency gain is quite high in the European case, and still considerable in the Bermudan case.

Remark 17. Fries & Kampen [13] and Fries & Joshi [14] propose simulation schemes which improve upon Euler SDE simulation and allow for taking larger time steps for obtaining the same accuracy. Assuming that such a scheme requires a time step of order, say $O(\sqrt{\Delta t})$, instead of $O(\Delta t)$ for the same accuracy, it is clear that, for example in the European case, the gain of our method with respect to this one is still order of $O(T/\sqrt{\Delta t})$.

References

[1] Ait-Sahalia, Y. Closed-form likelyhood expansions for multivariate diffusions. *Annals of Statistics*, forthcoming.

[2] Andersen, L. (1999) A simple approach to the pricing of Bermudan swaptions in a multifactor Libor market model. *Journal of Computational Finance*, **3**, 5-32

[3] Bally, V. and Talay, D. (1996) The law of the Euler scheme for stochastic differential equations I: convergence rate of the density. *Monte Carlo Methods Appl.*, **2**, 93–128.

[4] Broadie, M., Glasserman, P. Estimating Security Prices using Simulation, *Management Science,* vol. 42, 269-285, 1996.

[5] Bender, C., Schoenmakers, J. (2006) An iterative algorithm for multiple stopping: Convergence and stability. *Advances in Appl. Prob.*, Volume 38, Number 3 (2006), pp. 729-749

[6] Chen, N. and Glasserman, P.: Malliavin Greeks without Malliavin Calculus. *Stochastic Processes and their Applications*, Volume 117, Nr. 11 (2007), pp. 1689-1723.

[7] Cont, R. and Tankov, P. (2003). *Financial Modelling with Jump processes,* Chapman & Hall.

[8] Detemple, J., Garcia, R., Rindisbacher, M., Asymptotic Properties of Monte Carlo Estimators of Derivatives. *Management Science,* vol. 51, 11, 1657-1675, 2005.

[9] Duffie, D., *Dynamic Asset Pricing Theory.* Princeton: Princeton University Press 2001

[10] L' Ecuyer, P., Perron, G., On the Convergence Rates of IPA and FDC Derivative Estimators, *Operations Research,* vol. 42, 643-656, 1994

[11] Elie, R., Fermanian, J.-D., Touzi, N., Kernel estimation of Greek weights by parameter randomization, *Annals of Applied Probability,* to appear.

[12] Fourné, E., Lasry, J.M., Lebuchoux, J., Lions, P.L., Touzi, M., Applications of Malliavin Calculus to Monte Carlo Methods in Finance, *Finance and Stochastics,* vol. 3, 391-412, 1999.

[13] Fries, C., Kampen, J.: Proxy Simulation Schemes for generic robust Monte Carlo sensitivities, process oriented importance sampling and high accuracy drift approximation (with applications to the LIBOR market model), *Journal of Computational Finance,* Vol. 10, Nr. 2, 97-128, 2007.

[14] Partial Proxy Simulation Schemes for Generic and Robust Monte-Carlo Greeks. (2006) Fries, C., Joshi, M. (Working paper).

[15] Giles, M., and Glasserman, P. Smoking Adjoints: Fast Monte Carlo Greeks. *Risk*, vol. 19, 88-92, 2006.

[16] Glasserman, P. and Merener, N.: Numerical Solution of Jump-Diffusion LIBOR Market Models. *Finance and Stochastics*, 7, 1-27, 2003.

[17] Glasserman, P. and Zhao, X.: Fast Greeks by Simulation in Forward Libor Models. *Journal of Computational Finance,* Vol. 3, 1, 5-39, 1999.

[18] Haugh, M. B., Kogan, L. (2004) Pricing American options: A duality approach. *Operations Research*, **52**, 258-270

[19] Hunter, C., Jäckel, P. & Joshi, M.: Getting the Drift. *Risk* 14, 81-84 (July, 2001)

[20] Jamshidian, F.: LIBOR and swap market models and measures. Finance and Stochastics, 1, 293-330 (1997)

[21] Kampen, J.: *The WKB-Expansion of the fundamental solution of linear parabolic equations and its applications* . book manuscript, submitted to Memoirs of the American Mathematical Society, (downloadable at http://papers.ssrn.com/abstract=914122, Soc. Sc. Res. Net), 2006.

[22] Kampen, J.: Pointwise Convergent analytic expansions for densities of Feller processes (in preparation, Lecture SS 2006, Heidelberg).

[23] Kampen, J., Kolodko, A., Schoenmakers, J.: Monte Carlo Greeks for financial products via approximative Greenian Kernels WIAS-Preprint 1208.

[24] Kloeden, P.E., Platen, E., *Numerical solution of stochastic differential equations.* Springer Verlag Berlin, 1992.

[25] Kolodko, A., Schoenmakers, J. Iterative construction of the optimal Bermudan stopping time. *Finance and Stochastics*, **10**(1) (2006), 27-49.

[26] Krylov, N.V., *Lectures on Elliptic and Parabolic Equations in Hölder Spaces,* Graduate Studies in Mathematics, Vol. 12, American Mathematical Society, 1996.

[27] Kurbanmuradov, O., Sabelfeld, K., and Schoenmakers, J. Lognormal approximations to libor market models. *Journal of Computational Finance*, Vol. 6, 1, 69-100, 2002.

[28] Longstaff, F. A., Schwartz, R. S. Valuing American options by simulation: A simple least-square approach. *Review of Financial Studies*, **14** (2001), 113–147

[29] Milstein, G., and Schoenmakers, J. Monte Carlo construction of hedging strategies against multi-asset European claims. *Stochastics Stochastics Rep.*, 73(1-2):125–157, 2002.

[30] Milstein, G.N., Tretyakov, M.V., *Stochastic Numerics for Mathematical Physics*, Springer Verlag Berlin, 2004.

[31] Milstein, G.N., Tretyakov, M.V., Numerical Analysis of Monte Carlo Evaluation of Greeks by Finite Differences, *Journal of Computational Finance,* vol. 8, 3, 1-34, 2005.

[32] Pelsser, A., Pietersz, R., and van Regenmortel, M.: Fast Drift-Approximated Pricing in the BGM model. *Journal of Computational Finance,* Vol. 8(1) (2004), 93-124

[33] Piterbarg, V.V.: Risk sensitivities of Bermuda swaptions. *International Journal of Theoretical and Applied Finance,* **7**, No. 4, (2004) 465-509.

[34] Protter, P. *Stochastic Integration and Differential Equations.* Springer 1990.

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

[36] J. Schoenmakers. *Robust Libor Modelling and Pricing of Derivative Products.* Financial Mathematics. Chapman & Hall/CRC, 2005.