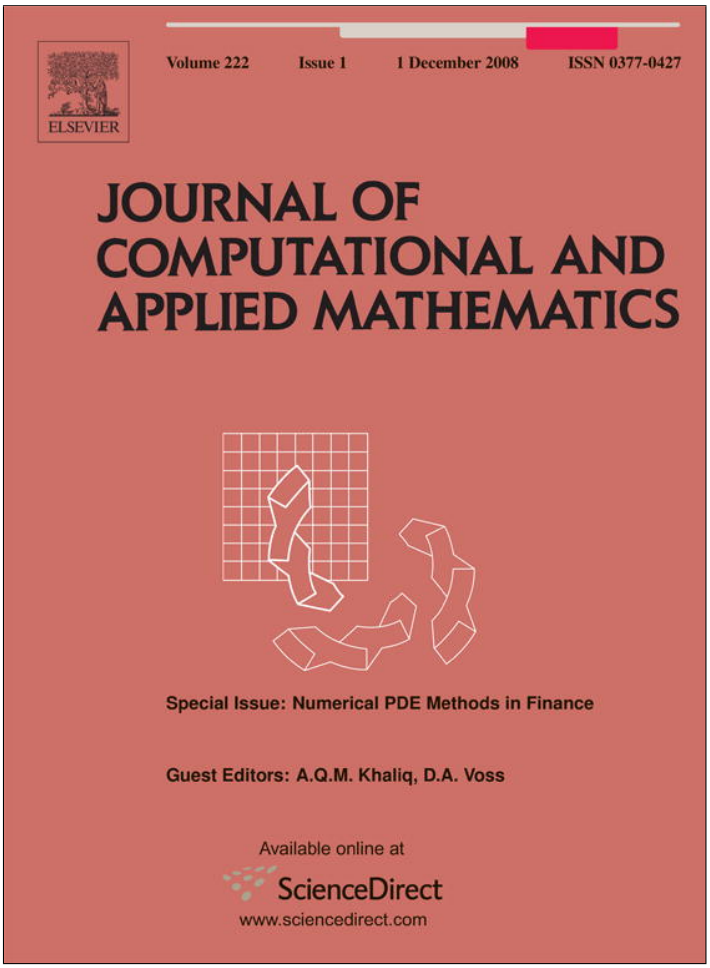

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Journal of Computational and Applied Mathematics 222 (2008) 42–53
JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS
www.elsevier.com/locate/cam

First exit time probability for multidimensional diffusions: A PDE-based approach

P. Patie[a,b,1], C. Winter[c,*]

[a] *RiskLab, Department of Mathematics, ETH Zurich, CH-8092 Zurich, Switzerland*
[b] *Department of Mathematical Statistics and Actuarial Science, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland*
[c] *SAM, ETH Zurich, CH-8092 Zurich, Switzerland*

Received 24 April 2006; received in revised form 5 April 2007

Abstract

First exit time distributions for multidimensional processes are key quantities in many areas of risk management and option pricing. The aim of this paper is to provide a flexible, fast and accurate algorithm for computing the probability of the first exit time from a bounded domain for multidimensional diffusions. First, we show that the probability distribution of this stopping time is the unique (weak) solution of a parabolic initial and boundary value problem. Then, we describe the algorithm which is based on a combination of the sparse tensor product finite element spaces and an *hp*-discontinuous Galerkin method. We illustrate our approach with several examples. We also compare the numerical results to classical Monte Carlo methods.

MSC: 60G40; 60J60; 65M60

Keywords: First exit time; Multidimensional diffusion; Finite element method

1. Introduction

First exit time distributions for multidimensional random processes are key quantities in many fields of sciences, such as mathematical physics, neurology, and also in mathematical finance. For instance, in the latter case, they are required for the pricing of some path-dependent options and for estimating the risk of default in the structural approach. Unfortunately, closed form solutions to this problem are not attainable except in a few specific cases. In this paper, we suggest a flexible and efficient numerical scheme based on PDE techniques, for computing the distribution of the first exit time from a bounded domain for multidimensional diffusions. We start by showing that the probability distribution of this stopping time can be represented as the unique weak solution of the backward Kolmogorov equation associated with the process subject to appropriate initial and boundaries conditions. It is well known that the strong solution of an initial-boundary value problem admits a probabilistic representation due to the Feynman-Kac formula; see [2, Theorem 2.3]. However, because of the incompatibility of the initial and boundary conditions

* Corresponding author.
E-mail addresses: patie@stat.unibe.ch (P. Patie), chwinter@math.ethz.ch (C. Winter).
[1] Supported by Credit Suisse Group, Swiss Re and UBS AG through RiskLab, Switzerland.

0377-0427/$ - see front matter
doi:10.1016/j.cam.2007.10.043

in our problem, we cannot get such a solution. To overcome this difficulty, we derive a variational formulation of the problem which will be also used for the numerical approach. Under some smoothness conditions on the coefficients of the diffusion and by using well known results on variational problems, we derive the existence and uniqueness of a weak solution. At a second stage due to smoothness of the free terms, we deduce that the solution is locally strong, and thus admits a Feynman-Kac type representation. We conclude by a uniqueness and a limiting argument that the first exit time probability is in fact the weak solution of the variational form of the Kolmogorov equation. For solving this problem numerically, straightforward application of standard schemes fails due to the so-called 'curse of dimension': The number of degrees of freedom on a tensor product finite element mesh of width h in dimension d grows like $O(h^{-d})$ as $h \to 0$. This can be avoided by using sparse tensor product spaces and an *hp*-discontinuous Galerkin (DG) time stepping procedure as in [6]. The resulting algorithm requires $O(h^{-1}|\log h|^{2d+6})$ operations. Compared to classical simulation techniques where Monte Carlo simulations are combined with discretization schemes for SDE's, our approach turns out to be more accurate and faster even for high-dimensional problems. It is also more flexible in the sense that we do not need the knowledge of the transition densities of the process unlike for simulation algorithms developed recently; see e.g. [5]. Moreover, the PDE approach provides the density at no additional cost using the ansatz functions for the time discretization and computes the distribution of the stopping time for any starting point of the diffusion in only one run.

The outline of the paper is as follows. We start by describing the set-up of the problem and fix the notation. In Section 2, we show that, under some regularity conditions on the coefficients of the diffusions and on the domain, the first exit time probability is the unique weak solution of the backward Kolmogorov equations subject to specific initial and boundary conditions. Section 3 is devoted to the description of the numerical algorithm which includes sparse tensor product finite element spaces and an *hp*-discontinuous Galerkin time integration. Finally, we give numerical results and compare the PDE approach with classical Monte Carlo methods.

1.1. The first exit time problem

Let $\left(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\right)$ be a filtered probability space and $\mathbf{X} := (X^1, \ldots, X^d)$, with $\mathbf{x} := (x_1, \ldots, x_d)$, be the solution to the following system of stochastic differential equations:

$$\mathrm{d}X_t^i = b_i(\mathbf{X}_t)\mathrm{d}t + \sum_{j=1}^{d} \sigma_{ij}(\mathbf{X}_t)\mathrm{d}W_t^j, \quad X_0^i = x_i \in \mathbb{R}, \tag{1}$$

where $W := (W^1, \ldots, W^d)^T$ is a d-dimensional standard Brownian motion with $d \geq 1$ and the coefficients b_i and σ_{ij} are smooth enough; see e.g. [7]. We denote the infinitesimal generator of $\mathbf{X}$ by $\mathcal{G}$, which has the form

$$\mathcal{G}f(\mathbf{x}) = \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} a_{ij}(\mathbf{x})\frac{\partial^2 f(\mathbf{x})}{\partial x_i x_j} + \sum_{i=1}^{d} b_i(\mathbf{x})\frac{\partial f(\mathbf{x})}{\partial x_i} \tag{2}$$

where $a = \sigma\sigma^\top$ and $f \in C_K^2(\mathbb{R}^d)$, the space of functions on $\mathbb{R}^d$ with compact support which are twice continuously differentiable.

Let A be an open bounded Borel subset of $\mathbb{R}^d$ with boundary denoted by ∂A and define the stopping time

$$\tau_A = \inf\{u \geq 0; \mathbf{X}_u \notin A\}. \tag{3}$$

This is the time of first exit of $\mathbf{X}$ from the domain A where we assume that ∂A is smooth enough. Let us denote by $\mathbb{Q}(t, \mathbf{x})$ the probability that $\mathbf{X}$ starting from $\mathbf{x}$ did not exit the domain A before t, i.e.

$$\mathbb{Q}(t, \mathbf{x}) = 1 - \mathbb{P}_{\mathbf{x}}(\tau_A < t). \tag{4}$$

Formally, $\mathbb{Q}$ coincides with the solution of the following backward Kolmogorov equation associated with the process $\mathbf{X}$:

$$\frac{\partial u}{\partial t}(t, \mathbf{x}) = \mathcal{G}u(t, \mathbf{x}) \quad \text{on } (t, \mathbf{x}) \in \mathbb{R}^+ \times A \tag{5a}$$

$$u(0, \mathbf{x}) = 1, \quad \mathbf{x} \in A \tag{5b}$$

$$u(t, \mathbf{x}) = 0, \quad \mathbf{x} \in \partial A, \ t > 0. \tag{5c}$$

1.2. Notation

Throughout the paper, we use the following notation. Let V and H separable Hilbert spaces with dense injection $V \overset{d}{\hookrightarrow} H$ and norms $\|.\|_V$ and $\|.\|_H$, respectively. Denoting by $(.,.)_H$ the scalar product on H and identifying H with H^*, the antidual of H, we get the Gelfand triple

$$V \overset{d}{\hookrightarrow} H \cong H^* \overset{d}{\hookrightarrow} V^*.$$

We denote by $(.,.)_{V^*,V}$ the usual duality pairing in $V^* \times V$ and by $\|.\|_{V^*}$ the norm in V^*. Next, for any $T > 0$, we consider the spaces

$$\mathcal{V} = L^2(0, T; V) \quad \text{and} \quad \mathcal{H} = L^2(0, T; H)$$

where if K is a Hilbert space, $L^2(0, T; K)$ denotes the class of measurable functions $f : (0, T) \to K$ for the Lebesgue measure such that $\int_0^T \|f(t)\|_K^2 \mathrm{d}t < \infty$. Then, we have

$$\mathcal{V}^* = L^2(0, T; V^*).$$

Furthermore, $H_0^1(A)$ is defined by $H_0^1(A) = \{v \in H^1(A);\ v_{|\partial A} = 0\}$ with the norm

$$\|v\|^2_{H_0^1(A)} = \sum_{i=1}^{d} \left\| \frac{\partial v}{\partial x_i} \right\|^2_{L^2(A)} + \|v\|^2_{L^2(A)}.$$

For any $T > 0$, we denote by Q_T the cylinder $A \times (0, T)$ and introduce the Hölder spaces $\mathcal{H}^{l,l/2}(Q_T)$, for a nonintegral positive number l. $\mathcal{H}^{l,l/2}(Q_T)$ is the Banach space of functions $u(x, t)$ that are continuous in Q_T, together with all derivatives of the form $\frac{\partial^r}{\partial t^r}\frac{\partial^s}{\partial x^s}$ for $2r + s < l$, and have a finite norm

$$\|u\|^{(l)}_{Q_T} = \langle u \rangle^{(l)}_{Q_T} + \sum_{j=0}^{[l]} \langle u \rangle^{(j)}_{Q_T}$$

where

$$\langle u \rangle^{(0)}_{Q_T} = \max_{Q_T} |u|, \qquad \|u\|^{(l)}_{Q_T} = \sum_{2r+s=l} \left| \frac{\partial u^r}{\partial t^r} \frac{\partial^s u}{\partial x^s} \right|^{(0)}_{Q_T}.$$

2. First exit time probability as the weak solution of the Kolmogorov equation

In what follows, we provide sufficient conditions on the coefficients of the diffusion and on the boundary of the domain A which ensure that the first exit time probability $\mathbb{Q}$ is the (weak) solution of the parabolic problem (5). Although it seems to be a classical result, to the best of our knowledge, we did not find any rigorous proof of such a representation.

2.1. The variational formulation of the backward Kolmogorov equation

In this part, we recall briefly the variational formulation of abstract parabolic problems and provide the assumptions under which existence and uniqueness of the weak solution are ensured. Consider for $F \in \mathcal{V}^*$ and $u_0 \in H$ the abstract parabolic problem

$$\frac{\mathrm{d}}{\mathrm{d}t}(u(t,.), v)_V + g(u(t,.), v) = (F, v)_{V^*,V} \quad \forall v \in V,\ 0 < t \le T, \tag{6a}$$

$$(u(0,.), v)_H = (u_0, v)_H \quad \forall v \in V. \tag{6b}$$

The derivative $\frac{\mathrm{d}}{\mathrm{d}t}$ is understood in the weak sense. We are now ready to state the following general result for the existence and uniqueness of a weak solution of the parabolic problem; see [4, Theorem 4.4.1 and Remark 4.4.3]

Proposition 1. *Assume that the bilinear form $g(.,.)$ in (6a) is*

(1) *continuous, i.e. there exits a constant $c > 0$ such that*

$$\|g(u,v)\|_H \leq c\|u\|_V\|v\|_V \quad \forall u, v \in V, \tag{7}$$

(2) *coercive on V, i.e. there exist positive constants λ and α such that*

$$g(v,v) + \lambda\|v\|_H^2 \geq \alpha\|v\|_V^2 \quad \forall v \in V. \tag{8}$$

Then for any $F \in \mathcal{V}^*$, there exists a unique $u \in \mathcal{V}$, a solution of (6a) and (6b).

Remark 2. Note that (6a) implies that $\frac{\mathrm{d}}{\mathrm{d}t}u \in \mathcal{V}^*$, so that

$$u \in C^0([0,T], H) \cap H^1\left(0, T; V^*\right). \tag{9}$$

We make the following assumptions.

Assumption 3. (1) The coefficients $\sigma_{i,j}$, $1 \leq i, j \leq d$, in (1) are $(k-1)$ times continuously differentiable on A.
(2) The coefficients b_i, $1 \leq i \leq d$, in (2) are $(k-1)$ times continuously differentiable on A.
(3) There exists a constant $\alpha > 0$ such that

$$\sum_{i=1}^{d}\sum_{j=1}^{d} a_{i,j}(\mathbf{x})\xi_i\xi_j \geq \alpha\sum_{i=1}^{d}\xi_i^2, \quad \forall \xi_i \in \mathbb{R},\ \mathbf{x} \in A. \tag{10}$$

(4) ∂A is piecewise smooth (as defined in [3]).

We take $H = L^2(A)$ and $V = H_0^1(A)$ so that $V^* = H^{-1}(A)$ and specify the bilinear form $g(.,.)$ defined on $H_0^1(A) \times H_0^1(A)$ by

$$g(u,v) = \frac{1}{2}(a\nabla u, \nabla v)_{L^2(A)} - ((b - \operatorname{div} a)\nabla u, v)_{L^2(A)}.$$

Now the Kolmogorov variational problem reads

$$\frac{\mathrm{d}}{\mathrm{d}t}(u(t,.), v)_{L^2(A)} + g(u(t,.), v) = 0 \quad \forall v \in H_0^1(A), \tag{11a}$$

$$u(0,.) = 1. \tag{11b}$$

Theorem 4. *Under Assumption 3 the bilinear form $g(.,.)$ satisfies the continuity and coerciveness conditions. As a consequence, the problem (11a) and (11b) admits a unique variational solution*

$$u \in L^2(0,T; H_0^1(A)) \cap C^0([0,T]; L^2(A)) \cap H^1(0,T; H^{-1}(A)). \tag{12}$$

Proof. Let $u \in H_0^1(A)$; then we have

$$g(u,u) = \frac{1}{2}(a\nabla u, \nabla u)_{L^2(A)} - ((b - \nabla a)\nabla u, u)_{L^2(A)}.$$

By using the Cauchy inequality, we obtain, for any $c_2 > 0$,

$$(b_j\frac{\partial u}{\partial x_j}, u)_{L^2(A)} \leq \left\|b_j\right\|_{L^\infty(A)}\left(c_2\left\|u\right\|_{L^2(A)}^2 + \frac{1}{4c_2}\left\|\frac{\partial u}{\partial x_j}\right\|_{L^2(A)}^2\right).$$

Combining this with (10), we get

$$g(u,u) \geq \frac{\alpha}{2}\|\nabla u\|_{L^2(A)} - \|b - \nabla a\|_{L^\infty(A)}\left(c_2\left\|u\right\|_{L^2(A)}^2 + \frac{1}{4c_2}\left\|\nabla u\right\|_{L^2(A)}^2\right).$$

We conclude by choosing $c_2 \geq \frac{\|b-\nabla a\|_{L^\infty(A)}}{2\alpha}$ to get the coerciveness conditions (8) and obtain the Theorem from Proposition 1.

Remark 5. By integration by parts, one observes that u satisfies (5) in the sense of distribution in Q_T, (5b) holds in the topology of $L^2(A)$ and the boundary condition (5c) is contained in the fact that $u \in L^2\left(0, T; H_0^1(A)\right)$ and A is regular.

2.2. Regularity of the weak solution

Thanks to the smoothness of the free terms of the problems and the regularity of A, the weak solution has some regularity. We have the following classical (local) regularity result; see e.g. [3, Theorem III.12.1].

Proposition 6. *Let Assumption 3 holds and fix α such that $[\alpha] = 1$. Then, the unique solution of the variational problem (11a) and (11b) belongs to*

$$u \in \mathcal{H}^{2+\alpha,1+\alpha/2}(Q_T) \cap C^0([0, T] \times A). \tag{13}$$

Moreover, for any $t_0 > 0$,

$$u \in C^0([t_0, T]; C^2(A) \cap C^0(\bar{A})). \tag{14}$$

2.3. The probabilistic representation

We end this section with the following result.

Theorem 7. *Let Assumption 3 holds. Then, the unique weak solution $u(t, \mathbf{x})$ of the variational Kolmogorov equation (11a) and (11b), belongs to*

$$\mathcal{H}^{2+\alpha,1+\alpha/2}(Q_T) \cap C^0([0, T] \times A) \tag{15}$$

and coincides with $\mathbb{Q}(t, \mathbf{x}) = 1 - \mathbb{P}_{\mathbf{x}}(\tau_A < t)$ for all $x \in \bar{A}$, $0 < t \leq T$.

Proof. Let u be the weak solution of (11a) and (11b). From Proposition 6, we have, for any $t_0 > 0$, $u \in C^0([t_0, T]; C^2(A) \cap C^0(\bar{A}))$. We deduce, from [2, Theorem II.2.3] and Remark 5, that u has the following probabilistic representation, setting $\tau_A^t = \tau_A \wedge t$, for any $t > 0$ and any $\mathbf{x} \in A$,

$$u(t, \mathbf{x}) = \mathbb{E}_{\mathbf{x}}[\mathbb{I}_{\{\tau_A^t = t\}}] = \mathbb{Q}(t, \mathbf{x}). \tag{16}$$

That is, we have, for any $t > 0$ and $\mathbf{x} \in \bar{A}$, $u(t, \mathbf{x}) = \mathbb{Q}(t, \mathbf{x})$. From (13), we know that $u \in C^0([0, T] \times A)$ which implies that $\lim_{t \searrow 0} u(t, \mathbf{x}) = 1$ for any $\mathbf{x} \in A$. Moreover, by the uniqueness of both the probabilistic representation and the solution of (5a) and (5b), we get, for any $\mathbf{x} \in A$, that

$$\lim_{t \searrow 0} \mathbb{Q}(t, \mathbf{x}) = \lim_{t \searrow 0} u(t, \mathbf{x}) = 1.$$

The proof of the Theorem is then completed by observing that u and $\mathbb{Q}$ coincide on $\{t = 0\} \times A \cup \{t > 0\} \times \bar{A}$.

Remark 8. We point out that we could actually consider more general diffusions and/or weaken the assumption on the boundary of the domain A. We could for instance consider time-dependent coefficients for the diffusion (1). In this case, the Assumption 3 should hold uniformly in time.

3. Numerical algorithm

As mentioned in the Introduction, a straightforward application of standard numerical schemes for solving (5) fails due to the so-called 'curse of dimension': The number of degrees of freedom on a tensor product finite element mesh of width h in dimension d grows like $O(h^{-d})$ as $h \to 0$. Therefore, Monte Carlo methods are mostly used where the error decreases like $O(M^{-1/2})$ if one uses M simulations. This holds for any $d \geq 1$, but only in a probabilistic sense. We use the algorithm described in [6] which provides an error of $O(N^{-p})$ where p is the degree of the finite elements which can be any integer ≥ 1. The method is based on the following two observations.

(i) To reduce the number of degrees of freedom in high dimensions, so-called sparse tensor product finite element spaces are used. Their numbers of degrees of freedom grow like $O(h^{-1}\,|\log h|^{d-1})$ as $h \to 0$, instead of $O(h^{-d})$ for the full tensor product spaces. At the same time, the approximation rate in $H^1(A)$ for elements of degree $p \geq 1$ and smooth functions is $O(h^p)$, the same as for full tensor product spaces. This result requires more regularity than $H^{p+1}(A)$ for the approximated function, and the amount of extra regularity increases with d. In the problem mentioned above, the initial data u_0 are incompatible with the boundary conditions. However, the solution operator $E(t)$ of the parabolic problem is an analytic semigroup and increases the smoothness of the solution $u(\cdot, t)$ for $t > 0$. This parabolic smoothing effect suffices for optimal convergence of sparse space discretization at $T > 0$ for any d, even for initial data that are just in $L^2(A)$.

(ii) Even with a sparse space discretization the number $\hat{N}_L$ of spatial degrees of freedom is substantial if d is large. Reducing the number of time steps (and thus, the number of spatial problems to be solved) to pass from $t = 0$ to the final time T is therefore essential. Time analyticity of $E(t)$ implies analytic time regularity of the solution $u(t)$ for $t > 0$, but not uniformly in $(0, T)$. This allows us to construct hp-discontinuous Galerkin time stepping schemes with exponential convergence in the number of spatial problems.

3.1. Sparse tensor product spaces

In the interval $I = (0, 1)$, we define the mesh $\mathcal{T}^\ell$ given by the nodes $j2^{-\ell}$, $j = 0, \ldots, 2^\ell$, with the mesh width $h_\ell = 2^{-\ell}$. We define $\mathcal{V}^\ell$ as the space of piecewise polynomials of degree $p \geq 1$ on the mesh $\mathcal{T}^\ell$ which are in $C^{p'-1}([0, 1])$ with $1 \leq p' \leq p$ and vanish at the endpoints 0, 1. We write $N^\ell = \dim \mathcal{V}^\ell$, $M^\ell := N^\ell - N^{\ell-1}$, $N^{-1} := 0$; Then $N^\ell = O(2^\ell)$, $\ell = 0, 1, 2, \ldots$. We employ a wavelet basis ψ_j^ℓ, $j = 1, \ldots, M^\ell$, $\ell = 0, 1, 2, \ldots$ of $\mathcal{V}^\ell$ with the properties

$$\mathcal{V}^\ell = \operatorname{span}\{\psi_j^\ell \mid 0 \leq \ell \leq L;\, 1 \leq j \leq M^\ell\} \quad \text{and} \quad \operatorname{diam}(\operatorname{supp}\psi_j^\ell) \leq C2^{-\ell}. \tag{17}$$

Any function $v \in \mathcal{V}^L$ has the representation

$$v = \sum_{\ell=0}^{L} \sum_{j=1}^{M^\ell} v_j^\ell \psi_j^\ell \tag{18}$$

with $v_j^\ell = (v, \widetilde{\psi}_j^\ell)$ where $\widetilde{\psi}_j^\ell$ are the so-called dual wavelets. For $v \in H_0^1(I)$ one obtains the series

$$v = \sum_{\ell=0}^{\infty} \sum_{j=1}^{M^\ell} v_j^\ell \psi_j^\ell \tag{19}$$

which converges in $L^2(I)$ and in $H_0^1(I)$.

For $v \in L^2(I)$ we can define a projection $P_L : L^2(I) \to \mathcal{V}^L$ by truncating (19):

$$P_L v := \sum_{\ell=0}^{L} \sum_{j=1}^{M^\ell} v_j^\ell \psi_j^\ell, \quad P_{-1} := 0. \tag{20}$$

The increment or detail spaces $\mathcal{W}^\ell$ are defined by

$$\begin{cases} \mathcal{W}^\ell := \operatorname{span}\{\psi_j^\ell : 1 \leq j \leq M^\ell\}, & \ell = 1, 2, 3, \ldots \\ \mathcal{W}^0 := \mathcal{V}^0. \end{cases} \tag{21}$$

Then

$$\mathcal{V}^\ell = \mathcal{V}^{\ell-1} \oplus \mathcal{W}^\ell \quad \text{for } \ell \geq 1, \quad \text{and} \quad \mathcal{V}^\ell = \mathcal{W}^0 \oplus \cdots \oplus \mathcal{W}^\ell,\ \ell \geq 0, \tag{22}$$

and $Q_\ell := P_\ell - P_{\ell-1}$ is a projection from $L^2(I)$ to $\mathcal{W}^\ell$.

In $I^d = (0, 1)^d$, $d > 1$, we define the subspace V^L as the tensor product of the one-dimensional spaces

$$V^L := \mathcal{V}^L \otimes \cdots \otimes \mathcal{V}^L \tag{23}$$

which can be written using (22) as

$$V^L = \sum_{0\le \ell_i \le L} \mathcal{W}^{\ell_1} \otimes \cdots \otimes \mathcal{W}^{\ell_d}. \tag{24}$$

The space V^L has $O(2^{Ld})$ degrees of freedom and is too costly if d is large. We shall use the sparse tensor product space

$$\begin{aligned}\widehat{V}^L &:= \operatorname{span}\{\psi_{j_1}^{\ell_1}(x_1)\ldots\psi_{j_d}^{\ell_d}(x_d) | 1 \le j_i \le M^{\ell_i}, \ell_1 + \cdots + \ell_d \le L\} \\ &= \sum_{0\le \ell_1+\cdots+\ell_d \le L} \mathcal{W}^{\ell_1} \otimes \cdots \otimes \mathcal{W}^{\ell_d}.\end{aligned} \tag{25}$$

As $L \to \infty$, we have $N_L := \dim(V^L) = O(2^{dL})$, and $\widehat{N}_L := \dim(\widehat{V}^L) = O(L^{d-1}\,2^L)$, i.e. the spaces $\widehat{V}^L$ have considerably smaller dimension than V^L. On the other hand, they do have similar approximation properties to V^L, provided the function to be approximated is sufficiently smooth.

3.2. Discontinuous Galerkin time stepping

We first analyze the time discretization of the parabolic problem (5)

$$\begin{aligned}&u'(t) - \mathcal{G}u(t) = 0, \quad t \in (0, T], \\ &u(0) = u_0,\end{aligned}$$

with $u_0 = 1$.

Let $\mathcal{M}$ be a partition of $(0, T)$ into $M(\mathcal{M})$ time steps $\{I_m\}_{m=1}^M$, $I_m = (t_{m-1}, t_m)$, $1 \le m \le M$, of size $k_m = t_m - t_{m-1}$. Define the one-sided limits of $u \in \mathcal{H}$ (or $\mathcal{V}$) as

$$\begin{aligned}u_m^+ &:= \lim_{s\to 0^+} u(t_m + s), \quad 0 \le m \le M-1, \\ u_m^- &:= \lim_{s\to 0^+} u(t_m - s), \quad 1 \le m \le M\end{aligned} \tag{26}$$

and $[u]_m := u_m^+ - u_m^-$, $1 \le m \le M-1$.

The weak solution $u \in L^2(0, T; \mathcal{V}) \cap H^1(0, T; \mathcal{V}^*)$ of (5) satisfies

$$B_{\mathrm{DG}}(u, v) = (u_0, v_0^+)$$

for all

$$v \in C_b(\mathcal{M}; \mathcal{V}) := \{u : (0, T) \to \mathcal{V} : u|_{I_m} \in C^0(\bar{I}_m; \mathcal{V}), I_m \in \mathcal{M}\}$$

where

$$B_{\mathrm{DG}}(u, v) := \sum_{m=1}^M \int_{I_m} \{(u', v)_{\mathcal{V}^*\times\mathcal{V}} + g(u, v)\}\mathrm{d}t + (u_0^+, v_0^+)_{\mathcal{H}} + \sum_{m=2}^M ([u]_{m-1}, v_{m-1}^+)_{\mathcal{H}}.$$

With time step k_m we associate an order $r_m \ge 0$, and define the semidiscrete space

$$\mathcal{S}^{\vec{r}}(\mathcal{M}; \mathcal{V}) = \{u : (0, T) \to \mathcal{V} : u|_{I_m} \in \mathcal{P}^{r_m}(I_m; \mathcal{V}), 1 \le m \le M\}$$

with the order vector $\vec{r} := (r_1, \ldots, r_M)$. The number of unknown coefficient functions in $\mathcal{V}$ of $u \in \mathcal{S}^{\vec{r}}(\mathcal{M}; \mathcal{V}) := \mathcal{S}^{\vec{r}}(\mathcal{M}) \otimes \mathcal{V}$ is

$$\dim(\mathcal{S}^{\vec{r}}(\mathcal{M})) = \sum_{m=1}^M (r_m + 1).$$

For simplicity we will consider uniform orders, i.e. $r_m = r$ for all m, and write $\mathcal{S}^{\vec{r}} = \mathcal{S}^r(\mathcal{M}; \mathcal{V})$. It is shown in [6] that for geometric meshes $\mathcal{M}$ the approximation error decreases exponentially in $N = \dim(\mathcal{S}^r(\mathcal{M}))$. A mesh $\{I_m\}_{m=1}^M$ in

$(0, T)$ is geometric with M time steps $I_m = (t_{m-1}, t_m)$, $m = 1, \ldots, M$, and grading factor $\eta \in (0, 1)$, if

$$t_0 = 0, \quad t_m = T\eta^{M-m}, \ 1 \le m \le M.$$

3.2.1. Fully discrete problem

We now also discretize the space $A = (0, 1)^d$ with the sparse grid subspace $\widehat{V}^L$ of V of mesh width $h = 2^{-L}$, $L > 0$. The discontinuous Galerkin time stepping scheme is given by:

Find $\widehat{U}^L \in \mathcal{S}^r(\mathcal{M}; \widehat{V}^L)$ such that

$$B_{\mathrm{DG}}(\widehat{U}^L, \widehat{W}) = (u_0, \widehat{W}_0^+) \quad \forall \widehat{W} \in \mathcal{S}^r(\mathcal{M}; \widehat{V}^L).$$

In each of the M time steps amounts to the solution of a linear system of size

$$(r + 1)\widehat{N}_L = (r + 1)O(h^{-1} |\log h|^{d-1}) \tag{27}$$

which depends on the time steps k and h and which we now derive.

Let $\mathcal{Q} := \mathcal{P}_r(I_m, \hat{V}^L)$, equipped with the norm of $L^2([t_{m-1}, t_m]) \otimes V$. In time step m of the algorithm we have to determine $\hat{U}_m^L := \hat{U}^L|_{I_m} \in \mathcal{Q}$ which satisfies, for all $W \in \mathcal{Q}$,

$$\int_{t_{m-1}}^{t_m} [((\hat{U}_m^L)', W) + g(\hat{U}_m^L, W)]\mathrm{d}t + (\hat{U}_m^L(t_{m-1}), W(t_{m-1})) = (\hat{U}_{m-1}^L(t_{m-1}), W(t_{m-1})) \tag{28}$$

where the expression $\hat{U}_0^L(t_0)$ is defined to mean the initial value $\hat{u}_0^L$.

Let $\{\phi\}_{j=0}^{r_m}$ be a basis of the polynomial space $\mathcal{P}_{r_m}(-1, 1)$. Then, the time shape functions on time interval I_m are given by $\phi_j \circ F_m^{-1}$ where the mapping $F_m : (-1, 1) \to I_m$ is given by

$$t = F_m(\tau) = \frac{1}{2}(t_{m-1} + t_m) + \frac{1}{2}k_m\tau, \qquad k_m = t_m - t_{m-1}, \quad \tau \in (-1, 1).$$

If we write $\hat{U}_m^L(x, t)$ and $\hat{W}$ in (28) as

$$\hat{U}_m^L(x, t) = \sum_{j=0}^{r_m} \hat{U}_{m,j}^L(x)(\phi_j \circ F_m^{-1})(t), \qquad \hat{W}(x, t) = \sum_{j=0}^{r_m} \hat{W}_{m,j}^L(x)(\phi_j \circ F_m^{-1})(t), \tag{29}$$

the variational problem (28) has the following form:

Find $(\hat{U}_{m,j}^L)_{j=0}^r \in (\hat{V}^L)^{r+1}$ such that for every $(\hat{W}_i)_{i=0}^r \in (\hat{V}^L)^{r+1}$

$$\sum_{i,j=0}^{r} \left(C_{ij} \cdot (\hat{U}_{m,j}^L, \hat{W}_i)_H + \frac{k_m}{2} G_{ij} \cdot g(\hat{U}_{m,j}^L, \hat{W}_i) \right) = \sum_{i=0}^{r} f_{m,i}(\hat{W}_i) \tag{30}$$

where

$$f_{m,i}(v) := \phi_i(-1)(\hat{U}_{m-1}^L(t_{m-1}), v)_H$$

$$C_{ij} := \int_{-1}^{1} \phi_j'\phi_i \mathrm{d}\tau + \phi_j(-1)\phi_i(-1), \qquad G_{ij} = \int_{-1}^{1} \phi_j\phi_i \mathrm{d}\tau. \tag{31}$$

Eq. (30) is a linear system of size $(r + 1)\hat{N}_L$ to be solved in each time step $m = 1, \ldots, M$. We will drop the subscript m for the sake of readability. Denoting by $\mathbf{M}$ and $\mathbf{S}$ the mass and stiffness matrices of $\hat{V}^L$ with respect to $(\cdot, \cdot)_H$ and $g(\cdot, \cdot)$, respectively, (30) takes the matrix form

$$\mathbf{R}\vec{u} = \vec{f}, \quad \mathbf{R} = \mathbf{C} \otimes \mathbf{M} + \frac{k}{2}\mathbf{G} \otimes \mathbf{S} \tag{32}$$

where $\vec{u}$ denotes the coefficient vector of $\hat{U}_m^L \in \mathcal{Q}$.

3.2.2. Derivation of a linear system

From now on we will use temporal shape functions $\phi_i(\tau) = (i+1/2)^{1/2} L_i(\tau)$. The system (32) of size $(r+1)\hat{N}_L$ can be reduced to solving $r+1$ linear systems of size $\hat{N}_L$. We use the Schur decomposition $\mathbf{C} = \mathbf{QTQ}^H$ with a unitary matrix $\mathbf{Q}$ and an upper triangular matrix $\mathbf{T}$ which has the eigenvalues $\lambda_1, \ldots, \lambda_{r+1}$ of $\mathbf{C}$ on the diagonal. Multiplying (32) by $\mathbf{Q}^H \otimes \mathbf{I}$ from the left gives

$$\left(\mathbf{T} \otimes \mathbf{M} + \frac{k}{2}\mathbf{I} \otimes \mathbf{S}\right) \vec{w} = \vec{g}$$

with

$$\vec{w} := (\mathbf{Q}^H \otimes \mathbf{I})\vec{x}, \qquad \vec{g} := (\mathbf{Q}^H \otimes \mathbf{I})\vec{f}.$$

This system is block upper triangular. With $\vec{w} = (\vec{w}_0, \ldots, \vec{w}_r)$ we obtain the solution by solving

$$\left(\lambda_{j+1}\mathbf{M} + \frac{k}{2}\mathbf{S}\right) \vec{w}_j = \vec{s}_j \quad \text{for } j = r, r-1, \ldots, 0 \tag{33}$$

where

$$\vec{s}_j := \vec{g}_j - \sum_{l=j+1}^{r} \mathbf{T}_{j+1,l+1}\mathbf{M}\vec{w}_l.$$

For each DG time step, we have to solve the $r+1$ linear systems in (33). Each of these linear systems is of the same type as in the backward Euler method where the matrix is $\mathbf{M} + k\mathbf{S}$. Therefore, an implementation of the DG method is very similar to an implementation of the backward Euler method.

Remark 9. We could also apply other time stepping schemes for the time discretization. But the *hp*-discontinuous Galerkin scheme uses the analytic time regularity of the solution $u(t)$ for $t > 0$. Especially for incompatible or non-smooth initial data, this is an advantage since the number of time points can be reduced using geometric meshes obtaining exponential convergence rates.

4. Numerical results

We first make a complexity comparison between the Monte Carlo method and the deterministic approach. Then, we show the outcomes of numerical experiments from both approaches in the one-dimensional case, by considering the first exit time of an interval for a Brownian motion with a linear drift, where analytic formulas are available. Results for the analogue multidimensional case are also presented. All computations are performed on a Dual-Core AMD Opteron(tm) Processor 2218 with 2.61 GHz using MATLAB 7.1.

4.1. Complexity comparison

Let $\tau > 0$ be a given tolerance. It is known that for the Monte Carlo method the statistical error is given by $O\left(M^{-\frac{1}{2}}\right)$ where M is the number of simulations. Additionally, we have to discretize the SDE in time. For the Euler-Maruyama discretization the weak error is $O(k)$ where k is the number of time steps. To keep both errors at tolerance τ we have

$$M \geq \tau^{-2} \quad \text{and} \quad k \geq \tau^{-1}.$$

The corresponding work for one single point is

$$W_{\text{mc}} = O(M \cdot k) = O(\tau^{-3}).$$

For the finite element method with *hp*-discontinuous Galerkin time stepping the $\mathcal{H}_1$-error is $O\left(N^{-p-\delta}\right)$ where N is the number of points, p the degree of the finite elements and $\delta = \frac{p}{(p+1)d-1}$ (see [6]). Therefore, we have

$$N \geq \tau^{-\frac{1}{p+\delta}}$$

and the corresponding work for N points is

$$W_{\det} = O(N(\log N)^{2d+6}) = O\left(\tau^{-\frac{1}{p+\delta}} \left|\frac{1}{p+\delta}\log\tau\right|^{2d+6}\right).$$

4.2. The one-dimensional case

Let $-\infty < a_1 < a_2 < +\infty$ and set $A = (a_1, a_2)$, $b(x) = b \in \mathbb{R}$ and $\sigma(x) = \sigma > 0$. Then, the density of the first exit time of X, the (scaled) Brownian motion with a linear drift b, from A is well known to be, see e.g. [1],

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}\{\tau_A < t\} = \mathrm{e}^{-\sigma^2 t/2}(\mathrm{e}^{b(a_1-x)}\mathrm{ss}_t(a_2-x, a_2-a_1) + \mathrm{e}^{b(a_2-x)}\mathrm{ss}_t(x-a_1, a_2-a_1))$$

where the help function ss_t has the following series representation:

$$\mathrm{ss}_t(v, y) = \sum_{k=-\infty}^{\infty} \frac{y - v + 2ky}{\sqrt{2\pi}t^{\frac{3}{2}}}\mathrm{e}^{\frac{-(y-v+2ky)^2}{2t}}.$$

We compare a Monte Carlo approach with the deterministic algorithm described above. For the simulation we discretize the stochastic process:

$$\widetilde{X}(t + \mathrm{d}t) = \widetilde{X}(t) + b\mathrm{d}t + \sigma\sqrt{\mathrm{d}t}\widetilde{Z}$$

with $\widetilde{X}(0) = x_0$ and $\widetilde{Z} \sim N(0, 1)$. The random numbers are generated by the MATLAB function `randn`. We choose $\mathrm{d}t = 5 \times 10^{-4}$, a random sample of the size 10^5, and count the number of sample paths which exceed A. For x_0 we choose nine points in the interval A, equally spaced, and use the same random sequences for each point.

In the deterministic approach we discretize the space with $L = 6$ levels. For the hp-discontinuous Galerkin time stepping we use $\log(2^{L+1})$ time steps with increasing polynomial degree $r_m = m$. Additionally, we compute the backward Euler method with $\mathrm{d}t = 10^{-2}$.

The computed probabilities of all approaches and the corresponding absolute errors are shown in Fig. 1. We let $A = (-1, 1)$ and use the parameter $\sigma = 0.5$, $b = 0.1$ and $T = 1$. The elapsed time for the PDE approach with hp-discontinuous Galerkin scheme is 0.11 s, with the backward Euler method 0.30 s, and for the Monte Carlo approximation 162.07 s. The error of the Monte Carlo approximation is the highest. Additionally, the probability is only computed at 9 points whereas with the deterministic approach it is computed at 129. There is also a significant error reduction with the hp-discontinuous Galerkin method.

We can also calculate the first exit time density at $x = 0$, i.e. the time derivative of the computed probability solution. Recall that for the DG time stepping procedure we have

$$\hat{U}_m^L(x, t) = \sum_{j=0}^{r_m} \hat{U}_{m,j}^L(x)(\phi_j \circ F_m^{-1})(t)$$

on the time interval I_m. Therefore, the time derivative is obtained from

$$\frac{\partial \hat{U}_m^L(x, t)}{\partial t} = \sum_{j=0}^{r_m} \hat{U}_{m,j}^L(x)(\phi_j' \circ F_m^{-1})(t)\frac{2}{k_m}.$$

For the backward Euler and the Monte Carlo method we approximate the time derivative by the standard difference quotient. The first exit time density and the absolute error are plotted in Fig. 2. For the deterministic approach we get a smooth solution whereas the Monte Carlo approximation oscillates. Again, the hp-discontinuous Galerkin time stepping yields the best results.

4.3. The multidimensional case

In the n-dimensional case, with $n > 1$, we consider the solution to the following system of stochastic differential equations:

$$\mathrm{d}X_t^i = b\mathrm{d}t + \sigma\mathrm{d}W_t^i, \quad i = 1 \ldots n,$$

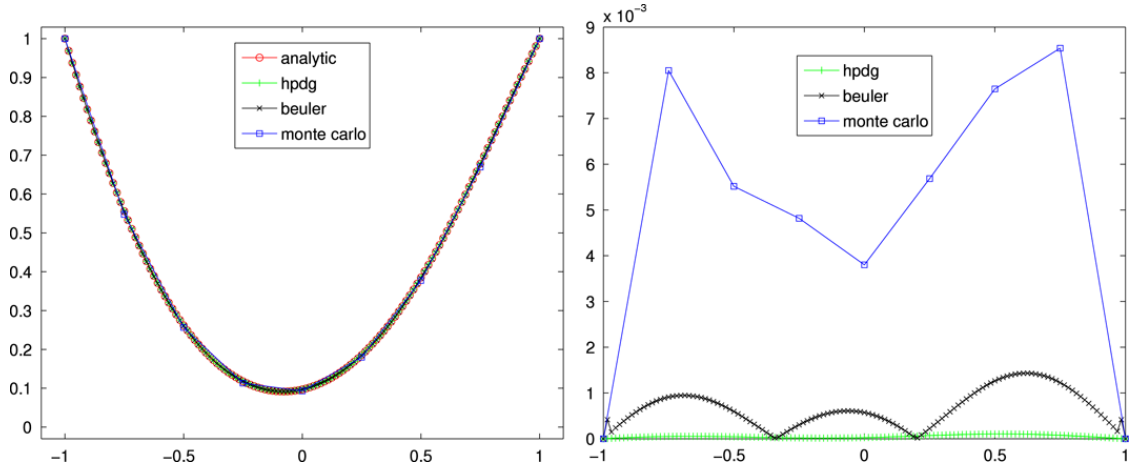

Fig. 1. Left: First exit time probability of a one-dimensional scaled Brownian motion with drift. Right: Absolute error with respect to the exact solution.

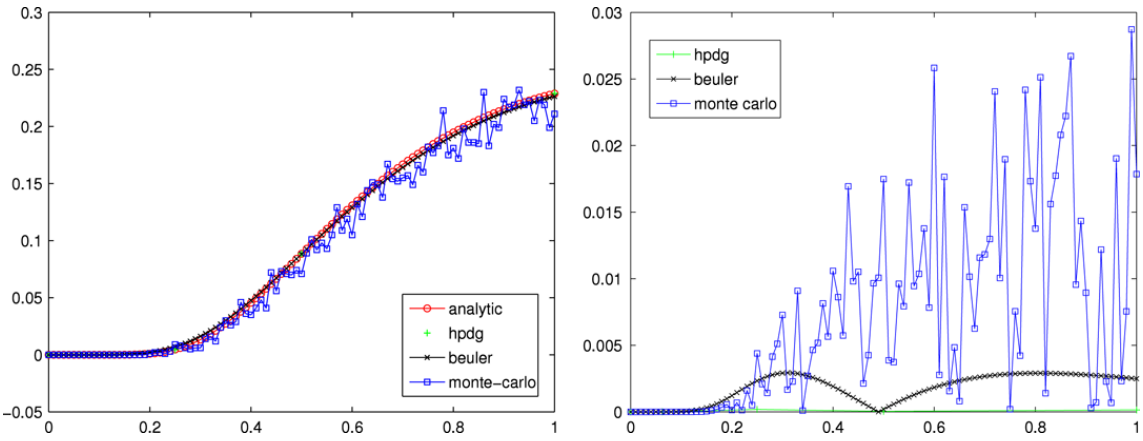

Fig. 2. Left: First exit time density of a one-dimensional scaled Brownian motion with drift. Right: Absolute error with respect to the exact solution.

where the W^i's are correlated Brownian motions with $\mathrm{d}\langle W^i, W^j\rangle_t = \rho \mathrm{d}t$, $i \neq j$, with parameters $\sigma = 0.5$, $\rho = 0.5$ and $b = 0.1$. We consider the domain $A = (-1, 1)^n$. The space is discretized with $L = 6$ levels. For the *hp*-discontinuous Galerkin time stepping we use again $\log(2^{L+1})$ time steps with increasing polynomial degree $r_m = m$. In the Monte Carlo method we again have $\mathrm{d}t = 5 \times 10^{-4}$, a random sample of the size 10^5 and count the number of sample paths which exceed A. For x_0 we choose nine points equally spaced on a curve where all dimensions but the first are fixed at $x_i = 0$. As a reference or "exact" solution we compute the probability with the deterministic approach and $L = 9$ levels.

In Fig. 3 we plot the difference between the computed probability and the reference solution in three dimensions. The PDE approach took 30 s to compute, the Monte Carlo method 360 s. Again the deterministic approach is more accurate and faster, although the probability is computed at about 2×10^6 points.

Similar results are obtained for five dimensions as shown in Fig. 4. Here, the elapsed time for the PDE approach is 674 s and the solution is computed at about 3×10^{10} points. The Monte Carlo method needs 871 s.

5. Conclusion

First, we showed that the first exit time probability for a multidimensional diffusion from a bounded domain is the unique weak solution of the associated backward Kolmogorov equation subject to specific initial and boundary conditions. Then, for solving this boundary value problem we used an algorithm based on a combination of the sparse tensor product finite element spaces and an *hp*-discontinuous Galerkin method. Numerical experiments confirmed the advantages of the presented algorithm over classical simulation techniques: It is faster and more accurate. Moreover,

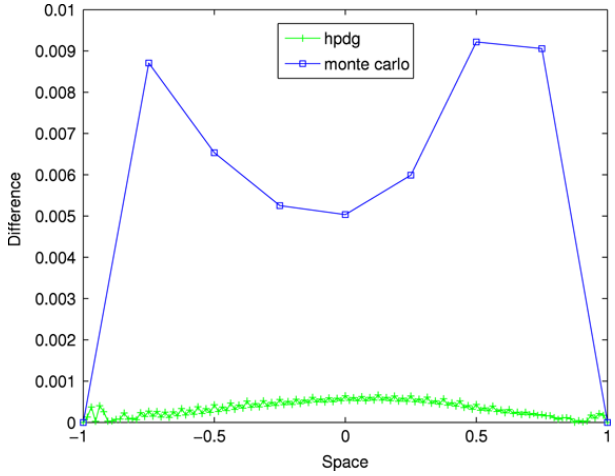

Fig. 3. Absolute error with respect to the reference solution for the 3D case.

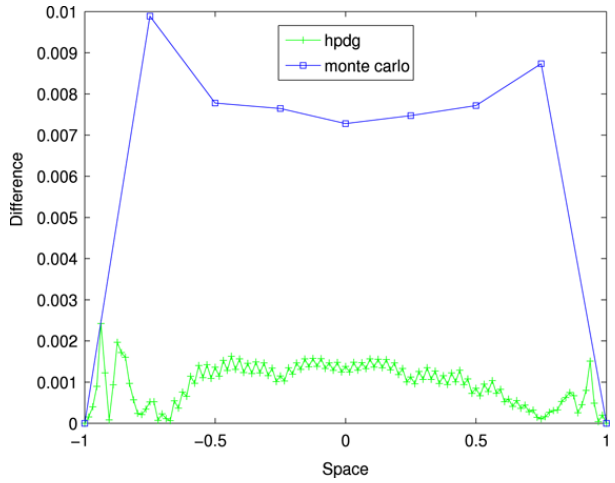

Fig. 4. Absolute error with respect to the reference solution for the 5D case.

the PDE-based approach has the additional advantage of providing the density and of computing the distribution of the stopping times for any starting point of the diffusion in only one run.

Acknowledgements

The authors are grateful to Prof. Schwab for many helpful discussions and also thank the referees for useful suggestions.

References

[1] A.N. Borodin, P. Salminen, Handbook of Brownian Motion — Facts and Formulae, Birkhaeuser Verlag, Basel, 2002.
[2] M. Freidlin, Functional Integration and Partial Differential Equations, in: Annals of Mathematics Studies, vol. 109, Princeton University Press, 1985.
[3] O.A. Ladyzenskaja, V.A. Solonnikov, N.N. Ural'ceva, Linear and Quasilinear Equations of Parabolic Type, in: Translations of Mathematical Monographs, vol. 23, American Mathematical Society, 1968.
[4] J.-L. Lions, E. Magenes, Non Homogeneous Boundary Value Problems and Applications, Springer-Verlag, Heidelberg, 1972.
[5] R. Mannella, Absorbing boundaries and optimal stopping in a stochastic differential equation, Phys. Lett. A 254 (1999) 257–262.
[6] T. von Petersdorff, C. Schwab, Numerical solutions of parabolic equations in high dimensions, M2AN Math. Model. Numer. Anal. 38 (2004) 93–127.
[7] D.W. Stroock, S.R.S. Varadhan, Multidimensional Diffusion Processes, Springer-Verlag, Berlin, 1979.