Finance Stochast. 8, 45–71 (2004)
DOI: 10.1007/s00780-003-0109-0

Finance and Stochastics

On the Malliavin approach to Monte Carlo approximation of conditional expectations

Bruno Bouchard[1], **Ivar Ekeland**[2], **Nizar Touzi**[3]

[1] Université Paris VI, PMA, and CREST, Boite Courrier 188, 75252 Paris, Cedex 05, France (e-mail: bouchard@ccr.jussieu.fr)
[2] University of British Columbia, Department of Mathematical Economics, Vancouver V6T, 1Z2, Canada (e-mail: ekeland@math.ubc.ca)
[3] CREST and CEREMADE, 15 bd Gabriel Péri, 92245 Malakoff, Cedex, France (e-mail: touzi@ensae.fr)

Abstract. Given a multi-dimensional Markov diffusion X, the Malliavin integration by parts formula provides a family of representations of the conditional expectation $E[g(X_2)|X_1]$. The different representations are determined by some *localizing functions*. We discuss the problem of variance reduction within this family. We characterize an exponential function as the unique integrated mean-square-error minimizer among the class of separable localizing functions. For general localizing functions, we prove existence and uniqueness of the optimal localizing function in a suitable Sobolev space. We also provide a PDE characterization of the optimal solution which allows to draw the following observation : the separable exponential function does not minimize the integrated mean square error, except for the trivial one-dimensional case. We provide an application to a portfolio allocation problem, by use of the dynamic programming principle.

Key words: Monte Carlo, Malliavin calculus, calculus of variations

JEL Classification: G10, C10

Mathematics Subject Classification (1991): 60H07, 65C05, 49-00

1 Introduction

Let X be a Markov process. Given n simulated paths of X, the purpose of this paper is to provide a Monte Carlo estimation of the conditional expectation

$$r(x) := E\left[g(X_2)|X_1 = x\right] , \tag{1.1}$$

The authors gratefully acknowledge for the comments raised by an anonymous referee, which helped understanding the existence result of Sect. 4.2 of this paper.

Manuscript received: August 2002; final version received: June 2003

i.e. the regression function of $g(X_2)$ on X_1. In order to handle the singularity due to the conditioning, one can use Kernel methods developed in the statistics literature, see e.g. [2]. However, the asymptotic properties of the Kernel estimators depend on the bandwidth of the Kernel function as well as the dimension of the state variable X. Therefore, using these methods in a Monte Carlo technique does not induce the $\sqrt{n}$ rate of convergence.

Malliavin integration by parts formula has been suggested recently in [7,6,10] and [9] in order to recover the $\sqrt{n}$ rate of convergence. Let us first discuss the case of the density estimator considered by [9]. The starting point is the expression of the density p of a smooth real-valued random variable G (see [12] Proposition 2.1.1) :

$$p(x) = E\left[1_{\{G>x\}}\delta\left(\frac{DG}{|DG|^2}\right)\right] , \tag{1.2}$$

where δ is the Skorohod integration operator, and D is the Malliavin derivative operator. Writing formally the density as $p(x) = E[\varepsilon_x(G)]$, where ε_x is the Dirac measure at point x, the above expression is easily understood as a consequence of an integration by parts formula (integrating up the Dirac). A remarkable feature of this expression is that it suggests a Monte Carlo estimation technique which does not require the use of Kernel methods in order to approximate the Dirac measure.

The same observation prevails for the case of the regression function $r(x)$ which can be written formally in :

$$r(x) = \frac{E\left[g(X_2)\varepsilon_x(X_1)\right]}{E\left[\varepsilon_x(X_1)\right]} .$$

Using the Malliavin integration by parts formula, [6] suggest an alternative representation of the regression function $r(x)$ in the spirit of (1.2). This idea is further developed in [10] when the process X is a multi-dimensional correlated Brownian motion.

An important observation is that, while (1.2) suggests a Monte Carlo estimator with $\sqrt{n}$ rate of convergence, it also provides the price to pay for this gain in efficiency : the right-hand side of (1.2) involves the Skorohod integral of the normalized Malliavin derivative $|DG|^{-2}DG$; in practice this requires an approximation of the continuous-time process DG and its Skorohod integral.

In this paper, we provide a family of such alternative representations in the vector-valued case. As in [6], we introduce *localizing functions* $\varphi(x)$ in order to catch the idea that the relevant information, for the computation of $r(x)$, is located in the neighborhood of x. The practical relevance of such localizing functions is highlighted in [10].

The main contribution of this paper is the discussion of the variance reduction issue related to the family of localizing functions. We first restrict the family to the class of separable functions $\varphi(x) = \prod_i \varphi_i(x^i)$. We prove existence and uniqueness of a solution to the problem of minimization of the integrated mean square error in this class. The solution is of the exponential form $\varphi_i(x^i) = e^{-\eta^i x^i}$ where the η^i's are positive parameters characterized as the unique solution of a system of non-linear equations. In the one-dimensional case, this result has been obtained heuristically by [9].

We also study the problem of minimizing the integrated mean square error within a larger class of all localizing functions. We first prove existence and uniqueness in a suitable Sobolev space. We then provide a PDE characterization of the solution with appropriate boundary conditions. An interesting observation is that separable localizing functions do not solve this equation, except for the one dimensional case.

The estimation method devised in this paper is further explored in [4] in the context of the simulation of backward stochastic differential equations.

The paper is organized as follows. Section 2 introduces the main notations together with some preliminary results. Section 3 contains the proof of the family of alternative representations of the conditional expectation. The variance reduction issues are discussed in Sect. 4. Numerical experiments are provided in Sect. 5. Finally, Sect. 6 provides an application of this technique to a popular stochastic control problem in finance, namely find the optimal portfolio allocation in order to maximize expected utility from terminal wealth.

2 Preliminaries

We start by introducing some notations. Throughout this paper we shall denote by $\mathcal{J}_k$ the subset of $\mathbb{N}^k$ whose elements $I=(i_1,\dots,i_k)$ satisfy $1\le i_1<\ldots<i_k\le d$. We extend this definition to $k=0$ by setting $\mathcal{J}_0=\emptyset$.

Let $I=(i_1,\dots,i_m)$ and $J=(j_1,\dots,i_n)$ be two arbitrary elements in $\mathcal{J}_m$ and $\mathcal{J}_n$. Then $\{i_1,\dots,i_m\}\cup\{j_1,\dots,j_n\}=\{k_1,\dots,k_p\}$ for some $\max\{n,m\}\le p\le\min\{d,m+n\}$, and $1\le k_1<\ldots<k_p\le d$. We then denote $I\vee J:=(k_1,\dots,k_p)\in\mathcal{J}_p$.

2.1 Malliavin derivatives and Skorohod integrals

Let $(\Omega,\mathcal{F},P)$ be a complete probability space equipped with a d-dimensional standard Brownian motion $W=(W^1,\dots,W^d)$. Since we are interested in the computation of the regression function (1.1), we shall restrict the time interval to $\mathbf{T}:=[0,2]$. We denote by $\mathbb{F}:=\{\mathcal{F}_t,\,t\in\mathbf{T}\}$ the $P-$completion of the filtration generated by W. Throughout this paper, we consider a Markov process X such that X_1 and X_2 belong to the Sobolev spaces $\mathbb{D}^{k,p}$ $(p,\ k\ge 1)$ of $k-$times Malliavin differentiable random variables satisfying :

$$||X||_{\mathbb{D}^{k,p}}:=\left[E(|X|^p)+\sum_{j=1}^{k}E\left(||D^jX||^p_{L^p(\mathbf{T}^j)}\right)\right]^{1/p}<\infty$$

where

$$||D^jX||_{L^p(\mathbf{T}^j)}=\left(\int_{\mathbf{T}^j}|D_{t_1}\cdots D_{t_j}X|^p dt_j\ldots dt_1\right)^{1/p}.$$

Given a matrix-valued process h, with columns denoted by h^i, and a random variable F, we denote

$$S_i^h(F) := \int_{\mathbf{T}} F(h_t^i)^* dW_t \text{ for } i = 1, \ldots, d, \text{ and } S_I^h(F) := S_{i_1}^h \circ \ldots \circ S_{i_k}^h(F)$$

for $I = (i_1, \ldots, i_k) \in \mathcal{J}_k$, whenever these stochastic integrals exist in the Skorohod sense. Here * denotes transposition. We extend this definition to $k = 0$ by setting $S_\emptyset^h(F) := F$. Similarly, for $I \in \mathcal{J}_k$, we set :

$$S_{-I}^h(F) := S_{\bar{I}}^h(F) \text{ where } \bar{I} \in \mathcal{J}_{d-k} \text{ and } I \vee \bar{I} \text{ is the unique element of } \mathcal{J}_d .$$

2.2 Localizing functions

Let φ be a C_b^0, i.e. continuous and bounded, mapping from $\mathbb{R}^d$ into $\mathbb{R}$. We say that φ is a *smooth localizing function* if

$$\varphi(0) = 1 \text{ and } \partial_I \varphi \in C_b^0 \text{ for all } k = 0, \ldots, d \text{ and } I \in \mathcal{J}_k .$$

Here, $\partial_I \varphi = \partial^k \varphi / \partial x_{i_1}, \ldots \partial x_{i_k}$. For $k = 0$, $\mathcal{J}_k = \emptyset$, and we set $\partial_\emptyset \varphi := \varphi$. We denote by $\mathcal{L}$ the collection of all such localization functions.

With these notations, we introduce the set $\mathbf{H}(X)$ as the collection of all matrix-valued $L^2(\mathcal{F}_2)$ processes h satisfying

$$\int_{\mathbf{T}} D_t X_1 h_t dt = I_d \text{ and } \int_{\mathbf{T}} D_t X_2 h_t dt = 0 \tag{2.1}$$

(here I_d denotes the identity matrix) and such that :

$$S_I^h(\varphi(X_1)) \text{ is well-defined in } \mathbb{D}^{1,2} \text{ for all } I \in \mathcal{J}_k , \ k \leq d \text{ and } \varphi \in \mathcal{L} . \tag{2.2}$$

We shall assume all over this paper that
Standing Assumption : $\mathbf{H}(X) \neq \emptyset$.
We next report useful properties for the rest of the paper.

Lemma 2.1 *Consider an arbitrary process $h \in \mathbf{H}(X)$. Then, for all bounded $f \in C_b^1$ and for all real valued r.v. $F \in \mathbb{D}^{1,2}$ with $E[F^2 \int_{\mathbf{T}} |h_t|^2 dt] < \infty$:*
(i) $\int_{\mathbf{T}} D_t (f(X_2)) h_t dt = 0$ *and therefore* $E\left[f(X_2) \int_{\mathbf{T}} F h_t^* dW_t \right] = 0.$
(ii) $\int_{\mathbf{T}} f(X_1) F h_t^* dW_t = f(X_1) \int_{\mathbf{T}} F h_t^* dW_t - \nabla f(X_1) F,$

Proof The first identity in (i) is a direct consequence of the chain rule formula together with (2.1). The second identity follows from the Malliavin integration by parts formula. To see that (ii) holds, we apply a standard result (see e.g. Nualart (1995) p. 40)

$$\int_{\mathbf{T}} f(X_1) F h_t^* dW_t = f(X_1) \int_{\mathbf{T}} F h_t^* dW_t - \nabla f(X_1) F \int_{\mathbf{T}} D_t X_1 h_t dt .$$

The required result follows from (2.1). □

2.3 Examples

Example 2.1 (Markov diffusion) Let X be defined by the stochastic differential equation :

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t \ , \tag{2.3}$$

together with an initial condition X_0. Here, b, σ and σ^{-1} are C_b^∞ vector and matrix-valued functions. Under the above condition X belongs to the set $\mathbb{L}_{\mathbf{T}}^{k,p}$ $(p,\ k \geq 1)$ of processes X such that $X_t \in \mathbb{D}^{k,p}$ for all $t \in \mathbf{T}$ and satisfying :

$$||X||_{\mathbb{L}_{\mathbf{T}}^{k,p}} := \left[E(\int_{\mathbf{T}} |X_t|^p dt) + \sum_{j=1}^{k} E\left(\int_{\mathbf{T}} ||D^j X_t||^p_{L^p(T^j)} dt \right) \right]^{\frac{1}{p}} < \infty \ .$$

We denote by $\mathbb{L}_{\mathbf{T}}^\infty := \cap_{p\geq 1} \cap_{k \geq 1} \mathbb{L}_{\mathbf{T}}^{k,p}$. We similarly define $\mathbb{D}^\infty$. Notice that $f(X) \in \mathbb{L}_{\mathbf{T}}^\infty$ whenever $f \in C_b^\infty$. In particular, $\sigma^{-1}(X) \in \mathbb{L}_{\mathbf{T}}^\infty$ (see [12] Proposition 1.5.1).

The first variation process of X is the matrix-valued process defined by :

$$Y_0 = I_d \text{ and } dY_t \ = \ \nabla b(X_t) Y_t dt + \sum_{i=1}^{d} \nabla \sigma^i(X_t) Y_t dW_t^i \ , \tag{2.4}$$

where ∇ is the gradient operator, and σ^i is the $i-$th column vector of σ. By [12] Lemma 2.2.2, the processes Y and Y^{-1} also belong to $\mathbb{L}_{\mathbf{T}}^\infty$.

The Malliavin derivative is related to the first variation process by :

$$D_s X_t \ = \ Y_t Y_s^{-1} \sigma(X_s) 1_{\{s \leq t\}} \ ; \ \ s \geq 0 \ , \tag{2.5}$$

so that :

$$D_s X_1 \ = \ Y_1 Y_2^{-1} D_s X_2 1_{\{s \leq 1\}} \ ; \ \ s \geq 0 \ . \tag{2.6}$$

It follows that $\mathbf{H}(X)$ is not empty. Indeed, since X, Y, Y^{-1} and $\sigma^{-1}(X)$ are in $\mathbb{L}_{\mathbf{T}}^\infty$,

$$\hat{h}_t := (D_t X_2)^{-1} Y_2 Y_1^{-1} (1_{t \in [0,1)} - 1_{t \in [1,2]}) \tag{2.7}$$

defines a process in $\mathbb{L}_{\mathbf{T}}^\infty$ satisfying (2.1). Moreover, for each real-valued $F \in \mathbb{D}^\infty$, and $i = 1, \ldots, d$, $S_i^{\hat{h}}(F)$ is well defined and belongs to $\mathbb{D}^\infty$ (see [12] Property 2 p38 and Proposition 3.2.1 p. 158). By simple iteration of this argument, we also see that $\hat{h}$ satisfies (2.2).

Example 2.2 (Euler approximation of a Markov diffusion) Consider the Euler approximation $\bar{X}$ of (2.3) on the grid $0 = t_0 < t_1 < \ldots < t_N = 1 < \ldots < t_{2N} = 2$, $N \in \mathbb{N}$,

$$\begin{aligned} \bar{X}_{t_0} &= X_0 \\ \bar{X}_{t_{n+1}} &= \bar{X}_{t_n} + b\left(\bar{X}_{t_n}\right)(t_{n+1} - t_n) + \sigma\left(\bar{X}_{t_n}\right)\left(W_{t_{n+1}} - W_{t_n}\right) \ , n \leq 2N-1 \ . \end{aligned}$$

Recalling that b, σ are C_b^∞, we see that, for each $n \in \{0, \dots, t_{2N}\}$, $\bar{X}_{t_n} \in \mathbb{D}^\infty$, where the Malliavin derivatives can be computed recursively as follows :

$$\begin{aligned} D_t\bar{X}_{t_1} &= \sigma\left(\bar{X}_{t_0}\right) 1_{t\le t_1} \\ D_t\bar{X}_{t_{n+1}} &= D_t\bar{X}_{t_n} + \nabla b(X_{t_n}) D_t\bar{X}_{t_n}(t_{n+1}-t_n) \\ &\quad + \sum_{i=1}^d \nabla\sigma^i(\bar{X}_{t_n}) D_t\bar{X}_{t_n}(W^i_{t_{n+1}} - W^i_{t_n}) + \sigma(\bar{X}_{t_n}) 1_{t\in(t_n,t_{n+1}]} \,. \end{aligned}$$

Noticing that $D_t\bar{X}_{t_n} = 0$ for $t > t_n$ and recalling that $\sigma^{-1} \in C_b^\infty$, we see that :

$$\begin{aligned} \hat{h}_t &:= (1-t_{N-1})^{-1}\sigma^{-1}(\bar{X}_{t_{N-1}}) 1_{t\in(t_{N-1},1]} \\ &\quad - (2-t_{2N-1})^{-1}\sigma^{-1}(\bar{X}_{t_{2N-1}}) D_{\hat{t}}\bar{X}_2 \sigma^{-1}(\bar{X}_{t_{N-1}}) 1_{t\in(t_{2N-1},2]} \end{aligned} \tag{2.8}$$

where $\hat{t} \in (t_{N-1}, 1)$, satisfies (2.1) and (2.2).

Remark 2.1 Let $\hat{h}$ be the process defined in Example 2.1 or 2.2. Then, using [12] Proposition 3.2.1 p. 158, we see that, for any localizing function $\varphi \in \mathcal{L}$:

$$E\left[S_I^{\hat{h}}(\varphi(X_1))\right]^p \;<\; \infty \text{ for all } p \;\ge\; 1,\; I \in \mathcal{J}_k,\; k \le d\,.$$

3 Alternative representation of conditional expectations

The starting point of this paper is an alternative representation of the regression function $r(x)$, introduced in (1.1), which does not involve conditioning. This is a restatement of a result reported in [6] without proof, and further developed in [10] in the case where the process X is defined as a correlated Brownian motion, see also [9] for the one-dimensional case with $f \equiv 1$, and [12] Exercise 2.1.3 for $\varphi = f \equiv 1$.

Theorem 3.1 *Let f be a mapping from $\mathbb{R}^d$ into $\mathbb{R}$ with $f(X_2) \in L^2$, and $\{A_i,\ i \le d\}$ a family of Borel subsets of $\mathbb{R}$. Then, for all $h \in \mathbf{H}(X)$, and $\varphi \in \mathcal{L}$:*

$$E\left[1_A(X_1) f(X_2)\right] = \int_A E\left[H_x(X_1) f(X_2) S^h\left(\varphi(X_1 - x)\right)\right] dx\,, \tag{3.1}$$

where $H_x(y) := \prod_{i=1}^d \mathbf{1}_{\{x^i < y^i\}}$, $A := A_1 \times \ldots \times A_d$, and $S^h = S^h_{(1,\dots,d)}$.

The proof of the above Theorem will be provided at the end of this section. The representation (3.1) can be understood formally as a consequence of d successive integrations by parts, integrating up the Dirac measure to the Heaviside function H_x. The main difficulty is due to the fact that the random variable $H_x(X_1)$ is not Malliavin-differentiable, see [12], Remark 2, p. 31. We therefore adapt the argument of the proof of (1.2) in [12].

Remark 3.1 By the same argument (see Proposition 2.1.1 and Exercise 3.1 in [12], and [9]), we also obtain an alternative representation of the density p_{X_1} of X_1. This is only a re-writing of Theorem 3.1 with $f \equiv 1$:

$$p_{X_1}(x) = E\left[H_x(X_1)E\left[S^h\left(\varphi(X_1-x)\right) \mid \mathcal{F}_1\right]\right]$$
$$= E\left[H_x(X_1)S^{\bar{h}}\left(\varphi(X_1-x)\right)\right] \quad \text{where } \bar{h} := h\mathit{1}_{[0,1]} \, ,$$

and the last equality follows from [12] Lemma 3.2.1. This means that, for the problem of density estimation, we can consider processes in $\mathbf{H}(X)$ which vanish on the time interval $(1,2]$.

Since the distribution of X_1 has no atoms, we obtain the following family of representations of the regression function $r(x)$, as a direct consequence of Theorem 3.1.

Corollary 3.1 *Let g be a mapping from $\mathbb{R}^d$ into $\mathbb{R}$ with $g(X_2) \in L^2$. Then, for all $h \in \mathbf{H}(X)$ and $\varphi \in \mathcal{L}$:*

$$r(x) \;:=\; E\left[g(X_2) \mid X_1 = x\right] \;=\; \frac{q[g](x)}{q[1](x)} \text{ where } q[f](x) \;:=\; E\left[Q^{h,\varphi}[f](x)\right] \;,$$

and

$$Q^{h,\varphi}[f](x) := H_x(X_1)f(X_2)S^h\left(\varphi(X_1-x)\right) \; .$$

Remark 3.2 Variance reduction I : Optimal localization As in [9,6] and [10], we introduce a *localizing function* φ *in* $\mathcal{L}$ in order to catch the idea that the relevant information, for the computation of $r(x)$, is located in the neighborhood of x. The practical importance of this issue is highlighted in [10]. The problem of selecting an "optimal" *localizing function* will be considered in the next section. The one-dimensional case was discussed heuristically by [9].

Remark 3.3 Variance reduction II : Control variates This is a direct extension of [9] who dealt with the one-dimensional case with $g \equiv 1$. Under the conditions of Theorem 3.1, it follows from Lemma 2.1 (i) that

$$E\left[H_x(X_1)g(X_2)S^h\left(\varphi(X_1-x)\right)\right] = E\left[\left(H_x(X_1)-c\right)g(X_2)S^h\left(\varphi(X_1-x)\right)\right] \;,$$

for all $c \in \mathbb{R}$. This suggests to apply a control variate technique, i.e. choose c in order to reduce the variance of the Monte Carlo estimator of the expectation (if g is not identically equal to 0). Clearly, the variance is minimized for

$$\hat{c}(x) := \frac{E\left[H_x(X_1)g(X_2)^2S^h\left(\varphi(X_1-x)\right)^2\right]}{E\left[g(X_2)^2S^h\left(\varphi(X_1-x)\right)^2\right]} \; .$$

Remark 3.4 For later use, we observe that, by using repeatedly Lemma 2.1 (ii), the Skorohod integral on the right-hand side of (3.1) can be developed in

$$S^h\left(\varphi(X_1-x)\right) \;=\; \sum_{k=0}^{d}(-1)^k \sum_{I\in\mathcal{J}_k} \partial_I\varphi(X_1-x)S^h_{-I}(1) \; . \tag{3.2}$$

Remark 3.5 Assume that

$$\sum_{k=0}^{d} \sum_{I \in \mathcal{J}_k} E\left[\left(f(X_2) S^h_{-I}(1)\right)^2\right] < \infty , \tag{3.3}$$

then it follows from the above Remark that Theorem 3.1 (and therefore Corollary 3.1) holds for all $\varphi \in C^0(\mathbb{R}^d)$ with $\varphi(0) = 1$, $\partial_I \varphi$ exists in the distribution sense and

$$\sum_{k=0}^{d} \sum_{I \in \mathcal{J}_k} E\left[\left(\partial_I \varphi(X_1 - x)\right)^2\right] < \infty .$$

Remark 3.6 In Sect. 4.2, we shall need to extend further the class of localizing function by only requiring that

$$\partial_I \varphi \in L^2(\mathbb{R}^d) \text{ for all } k = 0, \ldots , d \text{ and } I \in \mathcal{J}_k . \tag{3.4}$$

We shall see in Proposition 4.1 that the set of functions satisfying (3.4) can be imbedded in $C^0(\mathbb{R}^d_+)$, thus providing a sense to the constraint $\varphi(0) = 1$.

Proof of Theorem 3.1 We shall prove the required representation result by using repeatedly an identity to be derived in the second part of this proof. Let us first introduce the following additional notation

$$\pi_i(x) := (0, \ldots , 0, x^{i+1}, \ldots , x^d)^* \quad \text{for } i = 0, \ldots , d-1, \text{ and } \pi_d(x) = 0 ,$$

for $x \in \mathbb{R}^d$, and :

$$I_i = (i+1, \ldots , d) \in \mathcal{J}_{d-i} \text{ for } i = 0, \ldots , d-1 \text{ and } I_d := \emptyset .$$

1. By a classical density argument, it is sufficient to prove the result for f smooth and $A^i = [a^i, b^i]$ with $a^i < b^i$.

2. In preparation of the induction argument below, we start by proving that, for all $i = 1, \ldots , d$,

$$\begin{aligned} &E\left[1_{A_i}(X_1^i)\phi^i(X_1)f(X_2)S^h_{I_i}\left(\varphi \circ \pi_i(X_1 - x)\right)\right] \\ &= \int_{A_i} E\left[H_{x^i}(X_1^i)\phi^i(X_1)f(X_2)S^h_{I_{i-1}}\left(\varphi \circ \pi_{i-1}(X_1 - x)\right)\right] dx^i , \end{aligned} \tag{3.5}$$

for any $\varphi \in \mathcal{L}$, f and $\phi^i \in C^1_b$ with $\phi^i(x)$ independent of the $i-$th component x^i.

To see this, define the r.v.

$$F_i := \int_{-\infty}^{X_1^i} 1_{A_i}(x^i)\phi^i(X_1)f(X_2)S^h_{I_i}\left(\varphi \circ \pi_{i-1}(X_1 - x)\right) dx^i .$$

Since f, ϕ^i are smooth, $A^i = [a^i, b^i]$ and $S^h_{I_i}\left(\varphi \circ \pi_{i-1}(X_1 - x)\right) \in \mathbb{D}^{1,2}$, F_i is Malliavin-differentiable. By direct computation, it follows that

$$D_t F_i = 1_{A_i}(X_1^i)\phi^i(X_1)f(X_2)S^h_{I_i}\left(\varphi \circ \pi_i(X_1 - x)\right) D_t X_1^i$$

$$+\int_{-\infty}^{X_1^i} dx^i\, 1_{A_i}(x^i) D_t\left\{\phi^i(X_1) f(X_2) S_{I_i}^h\left(\varphi\circ\pi_{i-1}(X_1-x)\right)\right\} \quad (3.6)$$

Now recall that the function ϕ^i does not depend on its $i-$th variable. Then, it follows from (2.1) that $\int_{\mathbf{T}} D_t\{\phi^i(X_1)\} h_t^i dt = 0$. Also, we know from Lemma 2.1 (i) that $\int_{\mathbf{T}} D_t\{f(X_2)\} h_t^i dt = 0$. Therefore, it follows from (3.6) that :

$$\int_{\mathbf{T}} D_t F_i h_t^i dt = 1_{A_i}(X_1^i)\phi^i(X_1) f(X_2) S_{I_i}^h\left(\varphi\circ\pi_i(X_1-x)\right) \quad (3.7)$$
$$+\int_{-\infty}^{X_1^i} dx^i\, 1_{A_i}(x^i)\phi^i(X_1) f(X_2)\int_{\mathbf{T}} D_t S_{I_i}^h\left(\varphi\circ\pi_{i-1}(X_1-x)\right) h_t^i dt\ ,$$

where we used the fact that $\int_{\mathbf{T}} D_t X_1^i h_t^i dt = 1$ by (2.1). We now observe that :

$$E\left[\int_{\mathbf{T}} D_t F_i h_t^i dt\right] = E\left[F_i \int_{\mathbf{T}} (h_t^i)^* dW_t\right]$$
$$= \int_{A_i} dx^i E\left[H_{x^i}(X_1^i)\phi^i(X_1) f(X_2) S_{I_i}^h\left(\varphi\circ\pi_{i-1}(X_1-x)\right)\int_0^2 (h_t^i)^* dW_t\right]\ ,$$

where we used the Malliavin integration by parts formula. Then, taking expectations in (3.7), we see that :

$$E\left[1_{A_i}(X_1^i)\phi^i(X_1) f(X_2) S_{I_i}^h\left(\varphi\circ\pi_i(X_1-x)\right)\right]$$
$$= \int_{A_i} dx^i E\left[H_{x^i}(X_1^i)\phi^i(X_1) f(X_2)\left\{S_{I_i}^h\left(\varphi\circ\pi_{i-1}(X_1-x)\right)\int_{\mathbf{T}} (h_t^i)^* dW_t - \int_{\mathbf{T}} D_t\{S_{I_i}^h\left(\varphi\circ\pi_{i-1}(X_1-x)\right)\} h_t^i dt\right\}\right]$$
$$= \int_{A_i} dx^i E\left[H_{x^i}(X_1^i)\phi^i(X_1) f(X_2) S_{I_{i-1}}^h\left(\varphi\circ\pi_{i-1}(X_1-x)\right)\right]$$

3. We now use repeatedly identity (3.5). First notice that, by density, (3.5) holds also for bounded $\phi^i(x)$. Set $\phi^d(x) := \prod_{i=1}^{d-1} 1_{A_i}(x^i)$. Since $\varphi(0) = 1$, we see that :

$$E\left[1_A(X_1) f(X_2)\right] = E\left[1_A(X_1) f(X_2) S_{I_d}^h\left(\varphi\circ\pi_d(X_1-x)\right)\right]$$
$$= \int_{A_d} dx^d E\left[H_{x^d}(X_1^d)\phi^d(X_1) f(X_2) S_{I_{d-1}}^h\left(\varphi\circ\pi_{d-1}(X_1-x)\right)\right]\ .$$

We next concentrate on the integrand on the right hand-side of the last equation. We set $\phi^{d-1}(y) := H_{x^d}(y^d)\prod_{i=1}^{d-2} 1_{A_i}(y^i)$, and we use again (3.5) to see that :

$$E\left[1_A(X_1) f(X_2)\right]$$
$$= \int_{A_d} dx^d \int_{A_{d-1}} dx^{d-1} E\left[H_{x^{d-1}}(X_1^{d-1})\phi^{d-1}(X_1) f(X_2) S_{I_{d-2}}^h\left(\varphi\circ\pi_{d-2}(X_1-x)\right)\right]\ .$$

Iterating this procedure, we obtain the representation result announced in the theorem. □

Remark 3.7 Let the conditions of Theorem 3.1 hold. For later use, observe that, by similar arguments,

$$E\left[H_{x^i}(X_1^i)1_{A^{-i}}(X_1^{-i})f(X_2)\right] = \int_{A^{-i}} E\left[H_x(X_1)f(X_2)S^h_{(-i)}\left(\varphi(X_1-x)\right)\right]dx^{-i}$$

where $A^{-i} = A_1 \times \ldots \times A_{i-1} \times A_{i+1} \times \ldots \times A_d$ and $dx^{-i} = \prod_{j\neq i} dx^j$.

4 Variance reduction by localization

Given a localizing function $\varphi \in \mathcal{L}$ (or φ in some convenient relaxation of $\mathcal{L}$, see Remark 3.6), and $h \in \mathbf{H}(X)$, the representation result of Corollary 3.1 suggests to estimate the regression coefficient $r(x)$ by the Monte Carlo estimator :

$$\hat{r}_n(x) \;:=\; \frac{\hat{q}_n[g](x)}{\hat{q}_n[1](x)} \text{ where } \hat{q}_n[f](x) \;:=\; \frac{1}{n}\sum_{k=1}^{n} Q^{h,\varphi}[f](x)^{(k)} \tag{4.1}$$

and

$$Q^{h,\varphi}[f](x)^{(k)} \;:=\; H_x\left(X_1^{(k)}\right) f\left(X_2^{(k)}\right) S^{h^{(k)}}\left(\varphi(X_1^{(k)}-x)\right) . \tag{4.2}$$

Here, $\left(X^{(k)}, h^{(k)}\right)$ are independent copies with the same distribution as (X,h). By direct computation, we have

$$\text{Var}\left[\hat{q}_n[f](x)\right] = \frac{1}{n}\left\{E\left[H_x(X_1)f(X_2)^2 S^h\left(\varphi(X_1-x)\right)^2\right] - q[f](x)^2\right\} .$$

In this section, we consider the problem of minimizing the mean square error (mse, hereafter)

$$I^h[f](\varphi) := \int_{\mathbb{R}^d} E\left[H_x(X_1)f(X_2)^2 S^h\left(\varphi(X_1-x)\right)^2\right]dx$$

within the class of localizing functions. This criterion has been introduced by [9] in the one dimensional case with $f \equiv 1$.

In order to ensure that this optimization problem is well-defined, we assume that :

$$\sum_{k=0}^{d}\sum_{I\in\mathcal{J}_k} E\left[f(X_2)^2 S^h_I(1)^2\right] \;<\; \infty \;\text{ and }\; E|f(X_2)| > 0 \tag{4.3}$$

(see Remark 2.1 and Remark 3.6).

Notice that only the restriction of φ to $\mathbb{R}^d_+$ is involved in $I^h[f](\varphi)$. We then consider the set $\mathcal{L}_+ \subset \mathcal{L}$ of functions of the form $\varphi_{|_{\mathbb{R}^d_+}}$. On this set, the functional $I^h[f]$ is convex, by linearity of the Skorohod integral.

4.1 Optimal separable localization

We first consider the subset $\mathcal{L}_+^s$ of localizing functions φ of $\mathcal{L}_+$ of the form

$$\varphi(x) = \prod_{i=1}^{d} \varphi_i(x^i) ,$$

and we study the integrated mse minimization problem within the class of such localizing functions :

$$v_s^h[f] := \inf_{\varphi \in \mathcal{L}_+^s} I^h[f](\varphi) . \tag{4.4}$$

Theorem 4.1 *Let $h \in \mathbf{H}(X)$ be fixed, and f a mapping from $\mathbb{R}^d$ into $\mathbb{R}$ satisfying Condition* (4.3). *Then, there exists a unique solution $\hat{\varphi}$ to the integrated mse minimization problem* (4.4) *given by :*

$$\hat{\varphi}(x) \;=\; e^{-\hat{\eta}^* x} \;,\;\; x \in (\mathbb{R}_+)^d \;, \textit{for some } \hat{\eta} \;\in\; (0,\infty)^d \;.$$

Moreover, $\hat{\eta}$ is the unique solution of the system of non-linear equations

$$(\hat{\eta}^i)^2 = \frac{E\left[f(X_2)^2\left(\sum_{k=0}^{d-1}\sum_{I\in\mathcal{J}_k^{-i}} S^h_{-I}(1)\prod_{j\in I}\hat{\eta}^j\right)^2\right]}{E\left[f(X_2)^2\left(\sum_{k=0}^{d-1}\sum_{I\in\mathcal{J}_k^{-i}} S^h_{-(I\vee i)}(1)\prod_{j\in I}\hat{\eta}^j\right)^2\right]} \;, \tag{4.5}$$

$1 \le i \le d$, *where* $\mathcal{J}_k^{-i} = \{I \in \mathcal{J}_k \;:\; i \notin I\}$.

Observe that (4.5) is a system of (deterministic) polynomial equations.

Remark 4.1 By (3.2), one can define the integrated mse minimization problem within some convenient relaxation of the class $\mathcal{L}_+^s$ of separable localizing functions. Since C_b^0 is dense in L^2, it is clear that the relaxation suggested in Remark 3.6 does not alter the value of the minimum.

We split the proof of the above Theorem in several Lemmas. The conditions of Theorem 4.1 are implicitly assumed in the rest of this section. We shall use the following additional notations :

$$H_x^{-i} \;:=\; \prod_{j\neq i} H_{x^j} \text{ and } \varphi^{-i}(x) \;:=\; \prod_{j\neq i}\varphi_j(x^j) \;\text{ for } \varphi \in \mathcal{L}_+^s \;.$$

Lemma 4.1 *Let h be an arbitrary element in $\mathbf{H}(X)$. Then, for all $\varphi \in \mathcal{L}_+^s$:*

$$S^h\left(\varphi(X_1)\right) = \varphi_i(X_1^i) S^h\left(\varphi^{-i}(X_1)\right) - \varphi_i'(X_1^i) S^h_{-(i)}\left(\varphi^{-i}(X_1)\right) \;.$$

Proof By Lemma 2.1 (ii), we directly compute that :

$$\begin{aligned}
S^h\left(\varphi(X_1)\right) &= S^h_{-I_i}\left(\varphi_i(X_1^i) S^h_{I_i}\left(\varphi^{-i}(X_1)\right)\right) \\
&= S^h_{-I_{i-1}}\left(\varphi_i(X_1^i) S^h_{I_{i-1}}\left(\varphi^{-i}(X_1)\right) - (\varphi_i)'(X_1^i) S^h_{I_i}\left(\varphi^{-i}(X_1)\right)\right) \\
&= \varphi_i(X_1^i) S^h\left(\varphi^{-i}(X_1)\right) - (\varphi_i)'(X_1^i) S^h_{-(i)}\left(\varphi^{-i}(X_1)\right) \;.
\end{aligned}$$

□

Remark 4.2 Let φ be an arbitrary separable localizing function in $\mathcal{L}_+^s$. Under (4.3), φ is in the effective domain of $I^h[f]$ (see Remark 3.4). Given a function $\psi : \mathbb{R}_+ \longrightarrow \mathbb{R}$ with $\psi(0) = 0$, define the mapping from $\mathbb{R}_+^d$ into $\mathbb{R}$:

$$\phi_i[\varphi, \psi](x) := \varphi^{-i}(x)\psi(x^i) \ = \ \psi(x^i) \prod_{j \neq i} \varphi_j(x^j) \ .$$

Then, if ψ is C^1 and has compact support, we have $\phi_i[\varphi, \varphi_i + \psi] \in \mathcal{L}_+^s$ and, by (4.3), is in the effective domain of $I^h[f]$.

Lemma 4.2 *Let φ be an arbitrary smooth separable localizing function. For all integer $i \leq d$, we denote by $\Psi_i(\varphi)$ the collection of all maps $\psi : \mathbb{R}_+ \longrightarrow \mathbb{R}$ such that $\phi_i[\varphi, \psi] \in \mathcal{L}_+^s$. Then, the minimization problem*

$$\min_{\psi \in \Psi_i(\varphi)} I^h[f](\phi_i[\varphi, \psi])$$

has a unique solution $\hat{\psi}(y) := e^{-\hat{\eta}^i y}$ for some $\hat{\eta}^i > 0$ defined by

$$(\hat{\eta}^i)^2 = \frac{\int_{\mathbb{R}^{d-1}} E\left[H_x^{-i}(X_1^{-i}) f(X_2)^2 S^h(\varphi^{-i}(X_1 - x))^2\right] dx^{-i}}{\int_{\mathbb{R}^{d-1}} E\left[H_x^{-i}(X_1^{-i}) f(X_2)^2 S^h_{(-i)}(\varphi^{-i}(X_1 - x))^2\right] dx^{-i}} \ . \qquad (4.6)$$

Proof 1. Assume that $\varphi_i \in \Psi_i(\varphi)$ is optimal (since $\phi_i[\varphi, \psi]$ does not depend on the i-th component φ_i of φ, we can use this notation to indicate the optimum) and consider some function $\psi : \mathbb{R}_+ \longrightarrow \mathbb{R}$ with compact support and $\psi(0) = 0$. By Remark 4.2, $\phi_i[\varphi, \varphi_i + \varepsilon\psi] \in \mathcal{L}_+^s$ for all $\varepsilon \in \mathbb{R}$. Then for all ε

$$I^h[f](\varphi) \leq I^h[f](\phi_i[\varphi, \varphi_i + \varepsilon\psi]) \ .$$

By Linearity of the Skorohod integral,

$$I^h[f](\varphi) \leq I^h[f](\varphi) + \varepsilon^2 I^h[f](\varphi^{-i}\psi) + 2\varepsilon \int_{\mathbb{R}^d} E\left[H_x(X_1) f(X_2)^2 S^h\left(\varphi(X_1 - x)\right) S^h\left(\varphi^{-i}\psi(X_1 - x)\right)\right] dx \ .$$

For $\varepsilon > 0$, we divide the above inequality by ε and we let ε go to 0. This implies that :

$$\int_{\mathbb{R}^d} E\left[H_x(X_1) f(X_2)^2 S^h\left(\varphi(X_1 - x)\right) S^h\left(\varphi^{-i}\psi(X_1 - x)\right)\right] dx \geq 0 \ .$$

Applying the same argument with $\varepsilon < 0$, it follows that equality holds in the above inequality. By Fubini's theorem, this provides

$$\int_{\mathbb{R}^{d-1}} E\left[H_x^{-i}(X_1^{-i}) f(X_2)^2 \int_{-\infty}^{X_1^i} S^h\left(\varphi(X_1 - x)\right) S^h\left(\varphi^{-i}\psi(X_1 - x)\right) dx^i\right] dx^{-i} = 0 \ . \qquad (4.7)$$

2. Now, using Lemma 4.1, and performing the change of variable $y = X_1^i - x^i$ (ω by ω),

$$\int_{-\infty}^{X_1^i} S^h\left(\varphi(X_1-x)\right) S^h\left(\varphi^{-i}\psi(X_1-x)\right) dx^i$$
$$= \int_{-\infty}^{X_1^i} \left(\varphi_i(X_1^i-x^i)F_i - \varphi_i'(X_1^i-x^i)G_i\right)\left(\psi(X_1^i-x^i)F_i - \psi'(X_1^i-x^i)G_i\right) dx^i$$
$$= \int_0^\infty \left(\varphi_i(y)F_i - \varphi_i'(y)G_i\right)\left(\psi(y)F_i - \psi'(y)G_i\right) dy\ ,$$

where we used the notations $F_i := S^h\left(\varphi^{-i}(X_1-x)\right)$ and $G_i := S^h_{-(i)}\left(\varphi^{-i}(X_1-x)\right)$. Recall that φ^{-i} does not depend on the $i-$th component. Integrating by parts, and recalling that ψ has compact support and $\psi(0) = 0$, this provides :

$$\int_{-\infty}^{X_1^i} S^h\left(\varphi(X_1-x)\right) S^h\left(\varphi^{-i}\psi(X_1-x)\right) dx^i = \int_0^\infty \psi(y)\left[\varphi_i(y)F_i^2 - \varphi_i''(y)G_i^2\right] dy\ .$$

Plug this equality in (4.7) and use again Fubini's theorem to see that :

$$0 = \int_0^\infty \psi(y)\left[\alpha_i\varphi_i(y) - \beta_i\varphi_i''(y)\right] dy$$

where α_i and β_i are the non-negative parameters defined by

$$\alpha_i := \int_{\mathbb{R}^{d-1}} E\left[H_x^{-i}(X_1^{-i}) f(X_2)^2 F_i^2\right] dx^{-i}$$
$$\text{and}\quad \beta_i := \int_{\mathbb{R}^{d-1}} E\left[H_x^{-i}(X_1^{-i}) f(X_2)^2 G_i^2\right] dx^{-i}\ .$$

By the arbitrariness of the perturbation function ψ, this implies that φ_i satisfies the ordinary differential equation

$$\alpha_i\varphi_i(y) - \beta_i\varphi_i''(y) \;=\; 0 \text{ for all } y \geq 0\ ,$$

together with the boundary condition $\varphi_i(0) = 1$. We shall prove in the next step that $\beta_i > 0$. Recalling that φ has to be bounded, as an element of $\mathcal{L}_+^s$, this provides the unique solution $\varphi_i(y) = e^{-\eta^i y}$ where $\eta^i = (\alpha_i/\beta_i)^{1/2}$.
3. To see that $\beta_i > 0$, take an arbitrary $x^i \in \mathbb{R}$, and use the trivial inequality $H_x^{-i} \geq H_x^{-i} 1_{(B_n)^{d-1}} 1_{(x^i,\infty)} = H_x 1_{(B_n)^{d-1}}$ to get :

$$\beta_i \geq \int_{(B_n)^{d-1}} E\left[H_x(X_1) f(X_2)^2 G_i^2\right] dx^{-i}\ .$$

Here n is an arbitrary positive integer, and $B_n := [-n, n]$. By Jensen's inequality, this provides :

$$\beta_i \geq (2n)^{1-d}\left\{\int_{(B_n)^{d-1}} E\left[H_x(X_1)|f(X_2)|G_i\right] dx^{-i}\right\}^2$$

$$= (2n)^{1-d} \left\{ E\left[H_{x^i}(X_1^i) 1_{(B_n)^{d-1}}(X_1^{-i}) |f(X_2)| \right] \right\}^2 ,$$

where we used Remark 3.7 together with the definition of G_i. Since this inequality holds for all $x^i \in \mathbb{R}$, we may send x^i to $-\infty$ and use Fatou's lemma to get :

$$\beta_i \geq (2n)^{1-d} \left\{ E\left[1_{(B_n)^{d-1}}(X_1^{-i}) |f(X_2)| \right] \right\}^2 .$$

Since $E|f(X_2)| > 0$, this proves that $\beta_i > 0$ by choosing a sufficiently large n.
4. Conversely, let φ_i be defined as in the statement of the Lemma. Then $I^h[f](\varphi) \leq I^h[f](\varphi^{-i}(\varphi_i + \psi))$ for all function ψ with compact support such that $\psi(0) = 0$. Using (4.3), we see by using classical density arguments that $I^h[f](\varphi) \leq I^h[f](\varphi^{-i}\psi)$ for all functions ψ such that $\psi\varphi^{-i} \in \mathcal{L}_+^s$. □

The last lemma suggests to introduce the subset $\mathcal{L}_+^{\exp}$ of $\mathcal{L}_+^s$ consisting of all separable localizing functions

$$\varphi_\eta(x) := \exp\left(-\eta^* x\right) , \ x \in \mathbb{R}_+^d ,$$

for some $\eta \in (0, \infty)^d$. For ease of notation we set :

$$J^h[f](\eta) := I^h[f](\varphi_\eta) \text{ and } w^h[f] := \inf_{\eta^1, \ldots, \eta^d > 0} J^h[f](\eta) . \tag{4.8}$$

Lemma 4.3 *Consider an arbitrary constant $K > w^h[f]$. Then*

$$\mathrm{cl}\left(\{\eta \in (0,\infty)^d \ : \ J^h[f](\eta) \leq K\} \right) \textit{ is a compact subset of } (0,\infty)^d .$$

Proof Fix $K > w^h[f]$, and let $\eta \in (0, \infty)^d$ be such that $J^h[f](\eta) \leq K$. We need to prove that all η^i's are bounded and bounded away from zero.

Let $1 \leq i \leq d$ be a fixed integer, and set $\varphi := \varphi_\eta$. By Lemma 4.1,

$$\begin{aligned} &\int_{\mathbb{R}^d} H_x(X_1) f(X_2)^2 S^h \left(\varphi(X_1 - x)\right)^2 dx \\ &= \int_{\mathbb{R}^{d-1}} H_{x^{-i}}(X_1^{-i}) f(X_2)^2 \int_{-\infty}^{X_1^i} \left(\varphi_i(X_1^i - x^i) F_i - \varphi_i'(X_1^i - x^i) G_i\right)^2 dx^i dx^{-i} , \end{aligned}$$

where we used the notations of the previous proof $F_i := S^h\left(\varphi^{-i}(X_1 - x)\right)$ and $G_i := S^h_{-(i)}\left(\varphi^{-i}(X_1 - x)\right)$. Using the fact that $(\partial\varphi/\partial x^i) = -\eta^i\varphi$, it follows from a trivial change of variable that

$$\begin{aligned} &\int_{\mathbb{R}^d} H_x(X_1) f(X_2)^2 S^h \left(\varphi(X_1 - x)\right)^2 dx \\ &= \int_{\mathbb{R}^{d-1}} H_{x^{-i}}(X_1^{-i}) f(X_2)^2 (F_i + \eta^i G_i)^2 dx^{-i} \int_0^\infty \varphi_i(y)^2 dy \\ &= (2\eta^i)^{-1} \int_{\mathbb{R}^{d-1}} H_{x^{-i}}(X_1^{-i}) f(X_2)^2 (F_i + \eta^i G_i)^2 dx^{-i} . \end{aligned}$$

We therefore have :

$$K \geq J^h[f](\eta) := \int_{\mathbb{R}^d} E\left[H_x(X_1) f(X_2)^2 S^h \left(\varphi(X_1 - x)\right)^2 \right] dx$$

$$= (2\eta^i)^{-1} \int_{\mathbb{R}^{d-1}} E\left[H_{x^{-i}}(X_1^{-i}) f(X_2)^2 (F_i + \eta^i G_i)^2\right] dx^{-i}$$
$$\geq (2\eta^i)^{-1} \int_{B_n^{d-1}} E\left[H_{x^{-i}}(X_1^{-i}) f(X_2)^2 (F_i + \eta^i G_i)^2\right] dx^{-i} \ ,$$

where we use the notation $B_n := [-n, n]$ for some arbitrary integer n. Observing that $1 \geq H_{x^i}$ for all $x^i \in \mathbb{R}$, we obtain after integrating the variable x^i over the domain B_n :

$$2nK \geq (2\eta^i)^{-1} \int_{B_n^d} E\left[H_x(X_1) f(X_2)^2 (F_i + \eta^i G_i)^2\right] dx \ .$$

By Jensen's inequality, this provides :

$$2nK \geq (2\eta^i)^{-1}(2n)^{-d} \left\{ \int_{B_n^d} E\left[H_x(X_1) |f(X_2)| (F_i + \eta^i G_i)\right] dx \right\}^2 \ . \quad (4.9)$$

We now use Theorem 3.1 and Remark 3.7 to see that :

$$\int_{B_n^d} E\left[H_x(X_1)|f(X_2)|F_i\right] dx = E\left[1_{B_n^d}(X_1)|f(X_2)|\right]$$
$$\int_{B_n^d} E\left[H_x(X_1)|f(X_2)|G_i\right] dx = \int_{B_n} E\left[1_{B_n^{d-1}}(X_1^{-i}) H_{x^i}(X_1^i)|f(X_2)|\right] dx^i$$

are both strictly positive for sufficiently large n. This provides the required bound for $(\eta^i)^{-1}$ and η^i out of inequality (4.9). □

Lemma 4.4 *There exists a unique solution $\hat{\eta} \in (0, \infty)^d$ to the optimization problem w of* (4.8)*, i.e.,*

$$w^h[f] = J^h[f](\hat{\eta}) = I^h[f](\varphi_{\hat{\eta}}) < I^h[f](\varphi_\eta) \text{ for all } \eta^1, \ldots, \eta^d > 0 \text{ with } \eta \neq \hat{\eta} \ .$$

Proof Observe that the mapping $\eta \mapsto J^h[f](\eta)$ is stricly convex and lower semi-continuous. Then, existence and uniqueness of a solution $\hat{\eta}$ follow immediately from Lemma 4.3. □

Proof of Theorem 4.1 Let (φ_n) be a minimizing sequence of (4.4). Using repeatedly Lemma 4.2, we can define a minimizing sequence (φ_{η_n}) in $\mathcal{L}_+^{\exp}$. Then, existence of a solution follows from Lemma 4.4. The uniqueness and the characterization of the optimal solution follow from Lemma 4.2. The system of nonlinear equations (4.5) is obtained from (4.6) by developing the Skorohod integral on the right hand-side and then performing the integration as in the above proof, see Remark 3.4. □

4.2 Variance reduction with general localization

We now consider the integrated mse minimization problem with the class of all localizing functions. In contrast with the separable case, we cannot work directly with smooth localizing functions.

4.2.1 Existence

By Remark 3.4, after a change of variable inside the expectation (ω by ω), the objective function can be written in

$$\begin{aligned}
I^h[f](\varphi) &= E\left[f(X_2)^2 \int_{-\infty}^{X_1^1} \cdots \int_{-\infty}^{X_1^d} \left(\sum_{k=0}^{d} (-1)^k \sum_{I \in \mathcal{J}_k} \partial_I \varphi(X_1 - \xi) S_{-I}^h(1) \right)^2 d\xi \right] \\
&= \int_{\mathbb{R}_+^d} E\left[f(X_2)^2 \left(\sum_{k=0}^{d} (-1)^k \sum_{I \in \mathcal{J}_k} \partial_I \varphi(\xi) S_{-I}^h(1) \right)^2 \right] d\xi \ , \\
&= \int_{\mathbb{R}_+^d} E\left[(\partial\varphi(\xi)^* Q_h)^2 \right] d\xi \ , \\
&= \int_{\mathbb{R}_+^d} \partial\varphi(\xi)^* E[Q_h Q_h^*] \partial\varphi(\xi) d\xi \ ,
\end{aligned}$$

where we have introduced the column vectors

$$\partial\varphi := (\ \partial_I \varphi\ )_{I \in \mathcal{J}_k,\ k=0,\dots,d} \quad \text{and} \quad Q_h := \left(\ (-1)^k f(X_2) S_{-I}^h(1)\ \right)_{I \in \mathcal{J}_k,\ k=0,\dots,d} .$$

Notice that the matrix

$$\Gamma_h := E[Q_h Q_h^*]$$

is symmetric and non-negative. We shall assume later that it is indeed positive definite (see Theorem 4.2 below).

The above discussion leads us to consider the following *Bounded Cross Derivatives* Sobolev space. Consider the space $\mathrm{BCD}_0(\mathbb{R}_+^{\mathrm{d}})$ of functions $\varphi\ : \mathbb{R}_+^d \to \mathbb{R}$ such that all partial derivatives $\partial_I \varphi, I \in \mathcal{I}_k, k = 0, ..., d$, exist and are continuous on the interior of $\mathbb{R}_+^d$ and can be extended continuously to the boundary. Endow it with the inner product:

$$< \varphi, \psi >_{\mathrm{BCD}_0} := \int_{\mathbb{R}_+^d} \partial\varphi^* \partial\psi dx$$

which is clearly positive definite. Then $\mathrm{BCD}_0(\mathbb{R}_+^{\mathrm{d}})$ is a pre-Hilbert space, and its completion is a Hilbert space, which we denote by $\mathrm{BCD}(\mathbb{R}_+^{\mathrm{d}})$, and which is endowed with the scalar product:

$$< u, v >_{\mathrm{BCD}} := \int_{\mathbb{R}_+^d} \partial u^* \partial v dx$$

and the corresponding norm $\|u\|_{\mathrm{BCD}} =< u, u >_{\mathrm{BCD}}^{1/2}$.

The main purpose of this section is to prove an existence result for the integrated mse minimization problem when the localizing functions are relaxed to the space $\mathrm{BCD}(\mathbb{R}_+^{\mathrm{d}})$. To do this we need to incorporate the constraint $\varphi(0) = 1$ which has to be satisfied by any localizing function. Since the functions of $\mathrm{BCD}(\mathbb{R}_+^{\mathrm{d}})$ are only defined almost everywhere, this requires some preparation.

Denote by $C^{\infty}\left(\mathbb{R}_{+}^{d}\right)$ the space of all functions $\varphi:\mathbb{R}_{+}^{d}\rightarrow\mathbb{R}$, indefinitely differentiable on the interior of $\mathbb{R}_{+}^{d}$, and such that all derivatives can be extended continuously to the boundary. Denote by $C_{0}^{\infty}\left(\mathbb{R}_{+}^{d}\right)$ the space of functions in $C^{\infty}\left(\mathbb{R}_{+}^{d}\right)$ which have bounded support.

Lemma 4.5 (Localization) *Take some* $\varphi\in C_{0}^{\infty}\left(\mathbb{R}_{+}^{d}\right)$. *If* $u\in\mathrm{BCD}(\mathbb{R}_{+}^{\mathrm{d}})$, *then* $\varphi u\in\mathrm{BCD}(\mathbb{R}_{+}^{\mathrm{d}})$, *and the map* $u\rightarrow\varphi u$ *from* $\mathrm{BCD}(\mathbb{R}_{+}^{\mathrm{d}})$ *into itself is continuous.*

Proof Fix $I=(i_1,...,i_k)\in I_k$ for some $k=0,...,d$. Since all the i_j are different, Leibniz's formula takes a particularly simple form, namely:

$$\partial_I\left(\varphi u\right)=\frac{\partial^k}{\partial x^{i_1}...\partial x^{i_k}}\left(\varphi u\right)\;=\sum_{(I_1,I_2)\in A}\left(\partial_{I_1}\varphi\right)\left(\partial_{I_2}u\right)$$

where A is the set of all partitions of I in disjoint subsets I_1 and I_2. It follows from the assumption on φ that the $\partial_{I_1}\varphi\left(x\right)$ are uniformly bounded on $\mathbb{R}_{+}^{d}$, so that:

$$\left\|\partial_I\left(\varphi u\right)\right\|_{L^2}\leq C_I\left\|\partial u\right\|_{L^2}$$

for some constant C_I, and the result follows. □

Denote by $C_b^0\left(\mathbb{R}_{+}^{d}\right)$ the space of all bounded continuous functions on $\mathbb{R}_{+}^{d}$, endowed with the topology of uniform convergence.

Proposition 4.1 *There is a linear continuous map* $i:\mathrm{BCD}(\mathbb{R}_{+}^{\mathrm{d}})\rightarrow\mathrm{C}_{\mathrm{b}}^{0}\left(\mathbb{R}_{+}^{\mathrm{d}}\right)$ *such that* $u=i\left(u\right)$ *almost everywhere. Moreover,* $\lim_{\|x\|\rightarrow\infty}i(u)(x)=0$ *for all* $u\in\mathrm{BCD}(\mathbb{R}_{+}^{\mathrm{d}})$.

Proof 1 Pick some number $M>0$, and a function $\varphi_M\in C_{0}^{\infty}\left(\mathbb{R}_{+}^{d}\right)$ such that:

$$\begin{aligned}\varphi_M\left(x\right)&=1\;\text{ for }\;\|x\|\leq M/2\\ \varphi_M\left(x\right)&=0\;\text{ for }\;\|x\|\geq M\;.\end{aligned}$$

For any $u\in\mathrm{BCD}(\mathbb{R}_{+}^{\mathrm{d}})$, and $x\in\mathbb{R}_{+}^{d}$ set:

$$\begin{aligned}i_M\left(u\right)\left(x\right)&:=\left(-1\right)^d\int H_x\partial_{I_d}\left(\varphi_M u\right)dy\\ &=\left(-1\right)^d\int\left(1_M H_x\right)\partial_{I_d}\left(\varphi_M u\right)dy\end{aligned}\tag{4.10}$$

where $1_M(y)=1$ if $\|y\|<M$ and 0 if $\|y\|\geq M$. The right-hand side of this formula clearly is a continuous function of x, so $i_M\left(u\right)$ is continuous. In addition, we have:

$$\begin{aligned}\left|i_M(u)(x)-i_M(v)(x)\right|&\leq\left\|\varphi_M(u-v)\right\|_{\mathrm{BCD}}\left\|1_M H_x\right\|_{L^2}\\ &\leq C_M\left\|u-v\right\|_{\mathrm{BCD}}\left\|1_M H_x\right\|_{L^2}\end{aligned}$$

for some constant C_M, according to the localization lemma. If v converges to u in $\mathrm{BCD}(\mathbb{R}_{+}^{\mathrm{d}})$, then $i_M\left(v\right)$ converges uniformly to $i_M\left(u\right)$.

We next rewrite the right-hand side of (4.10) :

$$i_M(u)(x) = (-1)^d \int_{x+\mathbb{R}^d_+} \frac{\partial^d(\varphi_M u)}{\partial y^1 ... \partial y^d} dy \ .$$

For $v \in C^\infty\left(\mathbb{R}^d_+\right)$, we can apply Stokes' formula:

$$\begin{aligned} i_M(v)(x) &= (-1)^d \int_{x+\mathbb{R}^d_+} \frac{\partial}{\partial y^1}\left[\frac{\partial^{d-1}(\varphi_M v)}{\partial y^2 ... \partial y^d}\right] dy \\ &= (-1)^{d-1} \int_{x^{-1}+\mathbb{R}^{d-1}_+} \frac{\partial^{d-1}(\varphi_M v)}{\partial y^2 ... \partial y^d}\, dy^{-1} \ , \end{aligned}$$

which corresponds, in this context, to the partial integration with respect to the y^1 variable. Iterating this argument, we see that :

$$i_M(v)(x) \ = \ (\varphi_M v)(x) \text{ for all } x \in \mathbb{R}^d_+ \ .$$

If $u \in \mathrm{BCD}(\mathbb{R}^{\mathrm{d}}_+)$, we can find a sequence $v_n \in C^\infty\left(\mathbb{R}^d_+\right)$ converging to u in $\mathrm{BCD}(\mathbb{R}^{\mathrm{d}}_+)$. Then $i_M(v_n)$ converges to $i_M(u)$ uniformly and therefore $i_M(v_n) \longrightarrow i_M(u)$ in L^2. On the other hand, it follows from Lemma 4.5 that $\varphi_M v_n$ converges to $\varphi_M u$ in $\mathrm{BCD}(\mathbb{R}^{\mathrm{d}}_+)$. We can then identify the limits and conclude that :

$$i_M(u) \ = \ \varphi_M u \text{ almost everywhere, for all } u \in \mathrm{BCD}(\mathbb{R}^{\mathrm{d}}_+) \ .$$

But $\varphi_M u = u$ on the set $\|x\| \leq M/2$. It follows that $i_M(u) = u$ on the set $\|x\| \leq M/2$. Since $i_M(u) = i_{M'}(u)$ almost everywhere on the set $\|x\| \leq M \wedge M'/2$, we can define the function i by a classical localization argument :

$$i(u)(x) \ := \ i_M(u)(x) \text{ with } M \ := \ 2\|x\| \ .$$

Proof 2 We next prove that

$$\lim_{|x|\to\infty} i(u)(x) \ = \ 0 \text{ for all } u \in \mathrm{BCD}(\mathbb{R}^{\mathrm{d}}_+) \ . \tag{4.11}$$

To see this, observe that for all $u \in \mathrm{BCD}(\mathbb{R}^{\mathrm{d}}_+)$ and $x \in x_0 + \mathbb{R}^d_+$:

$$i(u)(x)^2 = i(u)(x_0)^2 - (-1)^d \int H_{x_0}(y) H_y(x) \partial_{I_d}(u^2)(y) dy \ , \tag{4.12}$$

where $\int H_{x_0}(y) H_y(x) \partial_{I_d}(u^2)(y) dy$ is well-defined as a sum of L^2-scalar products of elements of L^2. Then

$$\begin{aligned} \liminf_{\substack{|x|\to\infty \\ x\in x_0+\mathbb{R}^d_+}} i(u)(x)^2 &= \limsup_{\substack{|x|\to\infty \\ x\in x_0+\mathbb{R}^d_+}} i(u)(x)^2 \\ &= i(u)(x_0)^2 - (-1)^d \int H_{x_0} \partial_{I_d}(u^2) dy \ < \ \infty \ , \end{aligned}$$

and (4.11) follows from the fact that $u \in L^2(\mathbb{R}_+)$.

Proof 3 Using again (4.12), we directly estimate that :

$$i(u)(x_0)^2 = i(u)(x)^2 + (-1)^d \sum_{k=0}^{d} \sum_{I \in \mathcal{J}_k} \int H_x(y) H_y(x_0) \partial_I u(y) \partial_{\bar{I}} u(y) dy$$
$$\leq i(u)(x)^2 + \sum_{k=0}^{d} \sum_{I \in \mathcal{J}_k} \|\partial_I u\|_{L^2} \|\partial_{\bar{I}} u\|_{L^2}$$
$$\leq i(u)(x)^2 + C_d \|u\|^2_{\mathrm{BCD}} ,$$

where C_d is a constant which only depends on d. By sending $|x|$ to infinity and using (4.11), this provides $|i(u)(x_0)| \leq C_d \|u\|^2_{\mathrm{BCD}}$. Since $i(u-v) = i(u) - i(v)$, this shows that :

$$\sup_{x_0 \in \mathbb{R}^d_+} |i(u)(x_0) - i(v)(x_0)| \leq C_d \|u - v\|_{\mathrm{BCD}} ,$$

for all u, $v \in \mathrm{BCD}(\mathbb{R}^{\mathrm{d}}_+)$. Hence, i is a linear 1 continuous map. □

Although the functions of the space $\mathrm{BCD}(\mathbb{R}^{\mathrm{d}}_+)$ are defined almost everywhere, the evaluation function is well defined from the previous proposition by $u(x) := i(u)(x)$ for all $x \in \mathbb{R}^d_+$. We can then consider the following relaxation of the integrated mse minimization problem :

$$v^h[f] := \inf_{\varphi \in \bar{\mathcal{L}}^{\mathrm{BCD}}} I^h[f](\varphi) \tag{4.13}$$
$$\text{where} \quad \bar{\mathcal{L}}^{\mathrm{BCD}} := \left\{ \varphi \in \mathrm{BCD}(\mathbb{R}^{\mathrm{d}}_+) \ : \ \varphi(0) = 1 \right\} .$$

Observe that $I^h[f](\varphi)$ is well-defined by (3.2) and (4.3). We are now ready for the main result of this section.

Theorem 4.2 *Let $h \in \mathbf{H}(X)$ be fixed, f a mapping from $\mathbb{R}^d$ into $\mathbb{R}$ satisfying Condition* (4.3) *and such that Γ_h is positive definite. Then, there exists a unique solution $\hat{\varphi}$ to the integrated mse minimization problem* (4.13).

Proof Clearly, $I^h[f]$ is strictly convex and continuous on $\mathrm{BCD}(\mathbb{R}^{\mathrm{d}}_+)$. Since Γ_h is positive definite, it is also coercive for the norm $\|\cdot\|_{\mathrm{BCD}}$. Identifying u with its continuous version $i(u)$, it follows from Proposition 4.1 that the set $\bar{\mathcal{L}}^{\mathrm{BCD}}$ is closed. Hence, the existence result follows by classical arguments. □

4.2.2 PDE characterization

We continue our discussion by concentrating on the two dimensional case $d = 2$. Set

$$a \ := \ E[f(X_2)^2] \ , \ b \ := \ E\left[f(X_2)^2 \left(S^h(1) \right)^2 \right]$$

and introduce the vector and the symmetric non-negative matrix

$$Q \ := \ \begin{pmatrix} S^h_2(1) \\ S^h_1(1) \end{pmatrix} c \ := \ E\left[f(X_2)^2 S^h(1) Q \right] \ , q \ := \ E[f(X_2)^2 Q Q^*] \, .$$

Observing that $E[f(X_2)^2 S^h(1)] = E[f(X_2)^2 S_i^h(1)] = 0$, it follows that the objective function can be written in

$$I^h[f](\varphi) = \int_{\mathbb{R}_+^2} \left(b\varphi^2 - \varphi 2c^* \nabla\varphi + \nabla\varphi^* q \nabla\varphi + a\varphi_{12}^2\right) dx \; . \qquad (4.14)$$

Combining standard techniques of calculus of variation with Theorem 4.2, the above representation leads to the following characterization of the optimal localizing function.

Theorem 4.3 *Let $d = 2$. Then, the exists a unique continuous function in V satisfying $\varphi(0) = 1$ and*

$$b\varphi - \mathrm{Tr}[qD^2\varphi] + a\varphi_{1122} \;=\; 0 \text{ on } \mathbb{R}_+^2 \;, \qquad (4.15)$$

$$-c^1\varphi + q^{12}\varphi_1 + q^{11}\varphi_2 - au_{112} \text{ on } \mathbb{R}_+ \times \{0\} \;, \qquad (4.16)$$

$$-c^2\varphi + q^{12}\varphi_2 + q^{22}\varphi_1 - au_{122} \text{ on } \{0\} \times \mathbb{R}_+ \; . \qquad (4.17)$$

This function is the unique solution to the integrated mse minimization problem (4.13).

Remark 4.3 In general, no separable localizing function is optimal for the problem of integrated mse minimization within the class $\bar{\mathcal{L}}^{\mathrm{BCD}}$ of all localizing functions. We shall verify this claim in the two-dimensional case. Clearly it is sufficient to prove that the exponential localizing function $\varphi_{\hat{\eta}}$ is not optimal for the problem v^h. Indeed, it follows from (4.5) that $(x, y) := \hat{\eta}$ is characterized by :

$$0 = b + 2yc^1 + y^2 q^{11} - ax^2y^2 - q^{22}x^2$$
$$0 = b + 2xc^2 + x^2 q^{22} - ax^2y^2 - q^{11}y^2 \; .$$

Suppose to the contrary that the $\varphi_{\hat{\eta}}$ solves the problem v^h. Then, it follows from (4.15)-(4.16)-(4.17) that (x, y) has to satisfy the additional requirements :

$$0 = -c^1 - q^{12}x - q^{11}y + ax^2y$$
$$0 = -c^2 - q^{12}y - q^{22}x + axy^2$$
$$0 = b - q^{22}x^2 - q^{11}y^2 - 2q^{12}xy + ax^2y^2 \; .$$

One then easily checks that, except for the special case $q^{12} = 0$, the above system has no solution.

5 Numerical experiments

In this section, we consider the process X defined by the dynamics :

$$dX_t = \mathrm{diag}[X_t]\sigma dW_t \;\;,\;\; X_0^1 = X_0^2 = X_0^3 = 1 \; .$$

where

$$\sigma = \begin{bmatrix} 0.2 & 0 & 0 \\ 0.08 & 0.4 & 0 \\ 0.03 & -0.15 & 0.32 \end{bmatrix} \; .$$

The assumptions of Sect. 2 are satisfied when considering the logarithm of X.

Our aim is to estimate the density function p_{X_1} of X_1 and the regression function :

$$r(x) = 100 * E\left[\left(\frac{X_2^1 + X_2^2}{2} - X_2^3\right)^+ \mid X_1 = x\right] \tag{5.1}$$

on a grid of points $x = (x^1, x^2, x^3)$. By direct computation, we see that

$$(D_t X_2)^{-1} := \begin{bmatrix} \frac{1}{\sigma^{11} X_1^1} & 0 & 0 \\ -\frac{\sigma^{2,1}}{\sigma^{11}\sigma^{22} X_1^1} & \frac{1}{\sigma^{22} X_1^2} & 0 \\ \frac{\sigma^{21}\sigma^{32} - \sigma^{22}\sigma^{31}}{\sigma^{11}\sigma^{22}\sigma^{33} X_1^1} & \frac{-\sigma^{32}}{\sigma^{22}\sigma^{33} X_1^2} & \frac{1}{\sigma^{33} X_1^3} \end{bmatrix},$$

so that, with $\hat{h} := (D_t X_2)^{-1}\left(1_{[0,1)} - 1_{[1,2)}\right)$, all Skorohod integrals $S_I^{\hat{h}}\left(\varphi_{\hat{\eta}}(X_1 - x)\right)$ are computed explicitly.

We first estimate the optimal separable localizing function. The computation of the optimal coefficients $\hat{\eta}^i$ requires to solve numerically the system of non-linear equations (4.5); this turns out to be feasible by a simple iterative procedure, and is by no means time-consuming. Next, for each point x of our grid, we estimate the control variate function $\hat{c}(x)$ of Remark 3.3. The estimation of $\hat{\eta}$ and $\hat{c}$ is based on $100,000$ simulated paths.

The simulated paths of X and W are obtained by a very standard random numbers generator. In order to isolate the performance of the variance reduction technique studied in this paper, we do not introduce any other variance reduction method.

5.1 Density estimation

We start by estimating the density function p_{X_1} of X_1 at different points $x = (x^1, x^2, x^3)$. Each estimation is based on $20,000$ simulations of X_1. We provide the empirical mean and standard deviation (in brackets) of $1,000$ different estimators.

The density estimators are computed by using Remark 3.1, i.e. we replace $\hat{h}$ by $\bar{h} = \hat{h} 1_{t \le 1}$ in the representation of Corollary 3.1.

The first results concerning the density estimation suggest that the most important part of the variance reduction is obtained by the localizing procedure. The introduction of the function $\hat{c}(x)$ does not significantly improve the variance. This may be explained by the fact that the estimation of $\hat{c}(x)$ is rather difficult since it involves the Heaviside function H_x.

5.2 Regression function estimation

We next turn to the estimation of the regression function (5.1). Each estimation is based on $50,000$ simulations. We provide the empirical mean and standard deviation (in brackets) of $1,000$ different estimators.

Density estimation
Reduction by φ : Optimal localization,
Reduction by c : Control variate,

$x^1 = 0.7$

$x^3 \backslash x^2$		0.7	1.0	1.3
	True value	1.08	0.93	0.45
	Reduction by φ, c	1.09 [0.11]	0.94 [0.08]	0.45 [0.03]
0.7	Reduction by φ	1.09 [0.16]	0.94 [0.09]	0.45 [0.04]
	Reduction by c	1.07 [0.23]	0.93 [0.24]	0.46 [0.25]
	No Reduction	1.07 [0.26]	0.93 [0.26]	0.47 [0.28]
	True value	1.13	0.62	0.21
	Reduction by φ, c	1.13 [0.08]	0.62 [0.04]	0.21 [0.02]
1.0	Reduction by φ	1.14 [0.09]	0.62 [0.04]	0.21 [0.02]
	Reduction by c	1.11 [0.26]	0.61 [0.26]	0.21 [0.27]
	No Reduction	1.12 [0.29]	0.61 [0.29]	0.22 [0.31]
	True value	0.53	0.21	0.05
	Reduction by φ, c	0.53 [0.04]	0.21 [0.02]	0.05 [0.01]
1.3	Reduction by φ	0.53 [0.04]	0.21 [0.02]	0.05 [0.01]
	Reduction by c	0.51 [0.26]	0.19 [0.25]	0.06 [0.26]
	No Reduction	0.51 [0.29]	0.20 [0.29]	0.06 [0.31]

$x^1 = 1.0$

$x^3 \backslash x^2$		0.7	1.0	1.3
	True value	1.78	2.44	1.65
	Reduction by φ, c	1.80 [0.10]	2.44 [0.07]	1.65 [0.04]
0.7	Reduction by φ	1.80 [0.11]	2.44 [0.08]	1.65 [0.04]
	Reduction by c	1.78 [0.26]	2.45 [0.26]	1.67 [0.27]
	No Reduction	1.79 [0.30]	2.45 [0.31]	1.68 [0.32]
	True value	2.72	2.33	1.12
	Reduction by φ, c	2.73 [0.07]	2.34 [0.04]	1.12 [0.02]
1.0	Reduction by φ	2.73 [0.08]	2.34 [0.04]	1.12 [0.02]
	Reduction by c	2.73 [0.27]	2.35 [0.27]	1.15 [0.29]
	No Reduction	2.74 [0.34]	2.36 [0.35]	1.16 [0.37]
	True value	1.68	1.02	0.38
	Reduction by φ, c	1.69 [0.03]	1.02 [0.01]	0.38 [0.01]
1.3	Reduction by φ	1.69 [0.03]	1.02 [0.01]	0.38 [0.01]
	Reduction by c	1.69 [0.27]	1.05 [0.27]	0.41 [0.28]
	No Reduction	1.70 [0.35]	1.06 [0.37]	0.43 [0.39]

$x^1 = 1.3$

$x^3 \backslash x^2$		0.7	1.0	1.3
	True value	0.29	0.56	0.48
	Reduction by φ, c	0.29 [0.03]	0.56 [0.02]	0.48 [0.01]
0.7	Reduction by φ	0.30 [0.03]	0.56 [0.02]	0.48 [0.01]
	Reduction by c	0.28 [0.30]	0.56 [0.31]	0.50 [0.30]
	No Reduction	0.30 [0.41]	0.57 [0.43]	0.51 [0.44]
	True value	0.59	0.70	0.43
	Reduction by φ, c	0.59 [0.02]	0.70 [0.01]	0.43 [0.01]
1.0	Reduction by φ	0.59 [0.02]	0.70 [0.01]	0.45 [0.27]
	Reduction by c	0.58 [0.31]	0.70 [0.29]	0.45 [0.29]
	No Reduction	0.60 [0.47]	0.72 [0.48]	0.47 [0.49]
	True value	0.44	0.38	0.18
	Reduction by φ, c	0.44 [0.01]	0.38 [0.00]	0.18 [0.00]
1.3	Reduction by φ	0.44 [0.01]	0.38 [0.00]	0.18 [0.00]
	Reduction by c	0.44 [0.30]	0.38 [0.28]	0.19 [0.27]
	No Reduction	0.45 [0.48]	0.40 [0.49]	0.22 [0.51]

Regression function estimation
Reduction by φ : Optimal localization,

$x^1 = 0.9$

$x^3 \backslash x^2$		0.9	1.0	1.1
	True value	17.26	20.56	24.05
0.9	Reduction by φ	17.28 [1.12]	20.49 [1.19]	24.06 [1.17]
	No Reduction	16.88 [5.01]	20.62 [7.14]	25.04 [11.52]
	True value	13.70	16.59	19.61
1.0	Reduction by φ	13.72 [0.82]	16.59 [0.92]	19.73 [1.00]
	No Reduction	12.97 [6.20]	16.08 [10.41]	21.35 [25.48]
	True value	10.88	13.39	16.11
1.1	Reduction by φ	10.94 [0.85]	13.48 [0.90]	16.32 [1.05]
	No Reduction	10.58 [17.54]	13.81 [28.01]	13.19 [166.22]

$x^1 = 1.0$

$x^3 \backslash x^2$		0.9	1.0	1.1
	True value	20.08	23.58	27.24
0.9	Reduction by φ	19.93 [1.01]	23.40 [1.06]	27.08 [1.16]
	No Reduction	20.59 [8.80]	23.94 [32.24]	30.95 [63.28]
	True value	16.08	19.18	22.47
1.0	Reduction by φ	15.94 [0.85]	19.00 [0.92]]	22.27 [0.95]
	No Reduction	16.04 [11.48]	20.25 [32.05]	23.48 [68.62]
	True value	12.87	15.58	18.50
1.1	Reduction by φ	12.77 [0.76]	15.57 [0.85]	18.50 [0.96]
	No Reduction	11.26 [55.83]	14.11 [30.39]	25.46 [325.15]

$x^1 = 1.1$

$x^3 \backslash x^2$		0.9	1.0	1.1
	True value	23.13	26.81	30.64
0.9	Reduction by φ	23.12 [1.08]	26.68 [1.10]	30.58 [1.24]
	No Reduction	24.64 [28.23]	27.94 [33.39]	30.20 [116.46]
	True value	18.69	21.98	25.45
1.0	Reduction by φ	18.63 [0.94]	21.95 [0.91]	25.47 [1.01]
	No Reduction	19.54 [26.10]	23.63 [34.24]	27.82 [110.05]
	True value	15.07	17.99	21.10
1.1	Reduction by φ	15.04 [0.83]	17.96 [0.78]	21.17 [0.96]
	No Reduction	13.98 [30.29]	22.43 [623.93]	17.37 [180.44]

In view of the poor performance of the control variate technique (which involves the time-consuming computation of $\hat{c}(x)$), we concentrate on the use of the optimal localizing function. The results reported below prove the efficiency of this variance reduction technique, as the variance is significantly improved when the optimal localizing function is incorporated.

6 Optimal portfolio selection and option pricing

The representation of the conditional expectation presented in this paper has already proved to be powerful for the pricing of American options (see [10]). The algorithm developed to estimate the early exercise value is based on the dynamic programming equation which leads to a backward induction algorithm that requires the computation of a conditional expectation at each step (see also [1,5] and [11] for similar approaches).

Following [3], we propose to use the same approach to solve stochastic control problems written in a standard form.

6.1 Problem formulation

Consider the following simple optimal portfolio selection problem. The financial market consists in a non-risky asset, with price process normalized to unity, and two risky assets, one of which is non-tradable. We focus on the problem of valuation of a contingent claim, with payoff $G = g(X_T)$ written on the non-tradable asset X. We are then in an incomplete market framework, where G can be partially hedged by trading on the (tradable) risky asset Z whose price process is correlated to X.

More precisely, we assume that the dynamics of the pair process (X, Z) is given by

$$\begin{aligned} dX_t &= X_t \left(\mu^1 dt + \sigma^{11} dW_t^1\right) \\ dZ_t &= Z_t \left(\mu^2 dt + \sigma^{21} dW_t^1 + \sigma^{22} dW_t^2\right) \ , \end{aligned}$$

where W is a standard Brownian motion in $\mathbb{R}^2$, and μ^1, μ^2, $\sigma^{11} > 0$, $\sigma^{22} > 0$, σ^{21} are some given constants.

An admissible strategy is a U-valued predictable process (U is some compact subset of $\mathbb{R}$). We denote by $\mathcal{U}$ the set of such processes. Given a strategy $\nu \in \mathcal{U}$, the corresponding wealth process Y^ν is defined by

$$Y_t^\nu = Y_0 + \int_0^t \nu_r (Z_r)^{-1} dZ_r \ = \ Y_0 + \int_0^t \nu_r [\mu^2 dr + \sigma^{21} dW_r^1 + \sigma^{22} dW_r^2] \ .$$

Since the contingent claim can not be perfectly hedged, we consider the valuation rule induced by the utility indifference principle. Further simplification of the problem is obtained by assuming an exponential utility function with risk aversion parameter $a > 0$. In the presence of the liability G, the agent solves the following utility maximization problem

$$v^G(0, x, y) := \sup_{\nu \in \mathcal{U}} E\left[-e^{-a(Y_T^\nu - g(X_T))} \middle| (X_0, Y_0) = (x, y) \right] \ ,$$

where $T > 0$ is a given time horizon. Observe that the above value function does not depend on the state variable Z. The comparison to the maximal expected utility v^0, in the absence of any liability, leads to the so-called utility indifference valuation rule (see e.g. [8]) :

$$p(G, x, y) := \inf \left\{ \pi \in \mathbb{R} \ : \ v^G(0, x, y + \pi) \geq v^0(0, x, y) \right\} \ .$$

Observing that

$$v^G(0, x, y) = e^{-ay} v^G(0, x, 0) \ , \tag{6.1}$$

we see that

$$p(G,x,y) \;=\; p(G,x) \;=\; \frac{1}{a}\ln\left(\frac{w^G(0,x)}{w^0(0,x)}\right) \text{ where } w^G(0,x) \;:=\; v^G(0,x,0)\;.$$

Hence, the computation of the valuation rule p is reduced to the computation of the value functions $w^G(0,x)$ and $w^0(0,x)$.

Changing the time origin, one defines accordingly the dynamic version of the problem and the induced value functions $v^G(t,x,y)$ and $w^G(t,x)$. The value function v^G satisfies the dynamic programming principle

$$v^G(t,x,y) = \operatorname*{ess\,sup}_{\nu\in\mathcal{U}} \; E\left[v^G\left(t+\Delta, X_{t+\Delta}, Y^\nu_{t+\Delta}\right) \;\middle|\; (X_t,Y_t)=(x,y)\right] \;,$$

for any time step $\Delta > 0$. In view of (6.1), this translates to w^G in :

$$w^G(t,x) = \operatorname*{ess\,sup}_{\nu\in\mathcal{U}} \; E\left[e^{-aY^\nu_{t+\Delta}} w^G\left(t+\Delta, X_{t+\Delta}\right) \;\middle|\; (X_t,Y_t)=(x,0)\right] \;. \quad (6.2)$$

Hence, in the context of the particular model studied in this section, the number of state variables is reduced to one, which considerably simplifies the computations. The reason for considering such simplifications is that we are mainly concerned by the dynamic application of the conditional expectation estimation presented in the first sections of this paper.

6.2 The case $G = 0$

When there is no contingent claim to be delivered, the value function v^0 does not depend on the state variable X. It follows from (6.2) that the optimal control process is constant in time. The Monte Carlo estimation procedure is then considerably simplified, as it is sufficient to perform it on a single time step.

6.3 Discretization in time for v^G

Set $t_k := n^{-1}kT$, $n\in\mathbb{N}$ and $k=0,\ldots,n$ so that the time step is $\Delta = n^{-1}T$. By restricting the control process to be constant on $[t_k,t_{k+1})$, the dynamic programming principle suggests the following (backward) approximation of w^G :

$$\hat{w}^G(t_n, X_{t_n}) = -e^{ag(X_{t_n})} \quad (6.3)$$

$$\hat{w}^G(t_k, X_{t_k}) = \operatorname*{ess\,sup}_{\nu\in U} E\left[e^{-a\nu(Z_{t_{k+1}}-Z_{t_k})/Z_{t_k}}\,\hat{w}^G\left(t_{k+1}, X_{t_{k+1}}\right)|X_{t_k}\right]\;;\; k<n\;.$$

We now can appeal to the conditional expectation representation studied in this paper. By a trivial adaptation of Corollary 3.1 to the case where the time intervals $[0,1]$ and $[1,2]$ are replaced by $[0,t_k]$ and $[t_k,t_{k+1}]$ (see (3.5)), it is easily checked that, for each $\nu\in U$,

$$E\left[e^{-a\nu(Z_{t_{k+1}}-Z_{t_k})/Z_{t_k}}\,\hat{w}^G\left(t_{k+1}, X_{t_{k+1}}\right)\middle|\, X_t = x\right]$$

$$= \frac{E\left[1_{\{X_{t_k}>x\}} e^{-a\nu(Z_{t_{k+1}}-Z_{t_k})/Z_{t_k}} \hat{w}^G\left(t_{k+1}, X_{t_{k+1}}\right) S_k\right]}{E\left[1_{\{X_{t_k}>x\}} S_k\right]} \tag{6.4}$$

where $S_k := \int_0^{t_{k+1}} \varphi(X_{t_k}^2 - x^2) h_k dW_t^2$, $\varphi(x) = e^{-\eta x}$, and

$$h_k := (\sigma^{22} X_{t_k})^{-1}\left[(t_k)^{-1} 1_{[0,t_k)} - (t_{k+1} - t_k)^{-1} 1_{[t_k,t_{k+1}]}\right] .$$

We then use a backward induction procedure to compute the value function.

6.4 Monte Carlo approximation

We start by simulating N paths, X^i, of the process X. Given the value function for $k = n$ (i.e $t_n = T$),

$$\hat{w}^G(t_n, X_{t_n}^{(i)}) = -e^{-a\left(0-g(X_{t_n}^{(i)})\right)}$$

we use the other simulated paths, X^j, $j \neq i$, in order to build the Monte Carlo approximation of (6.4) at the point $(t_{n-1}, X_{t_{n-1}}^{(i)})$. The optimization over the parameter ν is achieved by a simple Newton-Raphson algorithm. Iterating this procedure backward, we can estimate, for each k and i, $\hat{w}(t_k, X_{t_k}^{(i)})$ together with the corresponding optimal control $\hat{\nu}(t_k, X_{t_k}^{(i)})$.

6.5 Numerical experiments

We consider the contingent claim defined by the payoff function $g(x) = 5 * \min\{K, x\}$. We fix $(X_0, Z_0) = (1, 1)$, $(\mu^1, \mu^2) = (0.1, 0.1)$, $\sigma^{11} = 0.15$, $U = [0, 40]$, $T = 1$, $n = 10$, and we perform the computations for different values of K, σ^{12}, σ^{22} and a. Each estimation of the value functions v^G, and the induced price p^G, is based on 8192 simulated paths. For each experiment, we compute 200 estimators. The average and the standard deviation in percentage of the average (in brackets) are collected in the following table.

As expected, the price is increasing with the risk aversion parameter a and with K. It is decreasing with the "relative correlation" between X and Z. For $K = \infty$, $G = 5 * X_T$ which is close to $5 * Z_T$ (in distribution) when the volatility coefficients are such that $\sigma^{11} = \sigma^{21} >> \sigma^{22}$. Then, for a low risk aversion parameter, we should obtain a price of order of $5 * Z_0 = 5$. In the following table, the result is indeed close to 5 for $a = 0.25$.

$K = 1.2,\ \sigma^{21} = 0.1,\ \sigma^{22} = 0.1$				
	p^G		v^G	
$a = 0.25$	5.20	[0.46%]	−2.94	[0.55%]
$a = 1$	5.26	[1.54%]	−154.04	[14.36%]
$K = 1.2,\ \sigma^{21} = 0.05,\ \sigma^{22} = 0.2$				
	p^G		v^G	
$a = 0.25$	5.26	[0.50%]	−2.98	[0.56%]
$a = 1$	5.40	[0.31%]	−177.91	[1.66%]
$K = \infty,\ \sigma^{21} = 0.05,\ \sigma^{22} = 0.2$				
	p^G		v^G	
$a = 0.25$	5.59	[0.52%]	−3.23	[0.61%]
$a = 1$	6.29	[1.30%]	−433.25	[8.91%]
$K = \infty,\ \sigma^{21} = 0.15,\ \sigma^{22} = 0.005$				
	p^G		v^G	
$a = 0.25$	5.09	[0.22%]	−2.85	[0.39%]

References

1. Bally, V., Pagès, G.: A quantization algorithm for solving multi-dimensional optimal stopping problems. Preprint 2001
2. Bosq, D.: Non-parametric statistics for stochastic processes. Berlin Heidelberg New York: Springer 1998
3. Bouchard, B.: Contrôle stochastique appliqué à la finance. PhD thesis, Paris: Université Paris Dauphine 2000
4. Bouchard, B., Touzi, N.: Discrete-time approximation and Monte Carlo simulation of backward stochastic differential equations. Preprint 2002
5. Carrière, J.F.: Valuation of the early-exercise price for options using simulations and nonparametric regression. Insurance Math. Econ. **19**, 19–30 (1996)
6. Fournié, E., Lasry, J.-M., Lebuchoux, J., Lions, P.-L.: Applications of Malliavin calculus to Monte Carlo methods in finance II. Finance Stochast. **5**, 201–236 (2001)
7. Fournié, E., Lasry, J.-M., Lebuchoux, J., Lions, P.-L., Touzi, N.: Applications of Malliavin calculus to Monte Carlo methods in finance. Finance Stochast. **3**, 391–412 (1999)
8. Hodges, S., Neuberger, A.: Optimal replication of contingent claims under transaction costs. Rev. Futures Markets **8**, 222–239 (1989)
9. Kohatsu-Higa, A., Pettersson, R.: Variance reduction methods for simulation of densities on Wiener space. SIAM J. Num. Anal. **40**, 431–450 (2002)
10. Lions, P.-L., Regnier, H.: Calcul du prix et des sensibilités d'une option américaine par une méthode de Monte Carlo. Preprint (2001)
11. Longstaff, F.A., Schwartz, R.S.: Valuing American options by simulation : A simple least-square approach. Rev. Financial Stud. **14**, 113–147 (2001)
12. Nualart, D.: The Malliavin calculus and related topics. Berlin Heidelberg New York: Springer 1995