A variance reduction technique using a quantized Brownian motion as a control variate

Antoine Lejay, Victor Reutenauer

A VARIANCE REDUCTION TECHNIQUE USING A QUANTIZED BROWNIAN MOTION AS A CONTROL VARIATE

A. LEJAY AND V. REUTENAUER

Abstract. This article exposes a new variance reduction technique for diffusion processes where a control variate is constructed using a quantization of the coefficients of the Karhunen-Loève decomposition of the underlying Brownian motion. This method may indeed be used for other Gaussian processes.

1. Introduction

This article aims at exposing a new variance reduction technique for Monte Carlo simulations for computing expectations of quantities related to Brownian motion or solutions of stochastic differential equations (SDE). Monte Carlo methods are sometimes the only available tool to perform simulations. It is customary to use SDE to model the evolution of stocks prices and rates in mathematical finance, since analytical or semi-analytical formulae are not available except for simple options. Yet they are in general naturally and simply implemented. The prices of derivatives are obtained by numerically averaging over a large number of realizations of stock prices or rates. The precision of a Monte Carlo method depends on the variance of the random quantity whose expectation is sought. A large number of methods, called *variance reduction techniques*, have been proposed to obtain better estimators and to reduce the computational time. There is no need to underline the practical importance of such techniques (see for example the books [9, 12]) which are still an active subject of research.

In order to simulate efficiently SDEs, we propose in this article a way to construct a control variate using quantization of Gaussian random variables. This scheme may be combined with standard techniques to simulate SDEs (Euler, Milstein, ...). We do not discuss here the problem of the simulation of SDE *per se* on which a large literature exists.

Date: November 8, 2010.

1991 *Mathematics Subject Classification.* Primary 60H35 ; Secondary 60H10, 65C30.

Key words and phrases. Monte Carlo simulation for diffusion processes, variance reduction, control variate, Karhunen-Loève decomposition, least squares method.

Quantization techniques consist in replacing a random variable X by a discrete one $\widehat{X}$ which takes its values in a finite set of points $\{x_1, \dots, x_Q\}$ and is chosen to be optimal. For example the points $\{x_i\}_{i=1}^Q$ are such that $\mathbb{E}[|X - \widehat{X}|^2]$ is minimal. With $p_i = \mathbb{P}[\widehat{X} = x_i]$, $\mathbb{E}[\phi(\widehat{X})] = \sum_{i=1}^Q p_i \phi(x_i)$ is an approximation of $\mathbb{E}[\phi(X)]$. Taking advantage of the way $\widehat{X}$ is chosen, Q could be much smaller than the number of realizations required to perform a Monte Carlo method to compute $\mathbb{E}[\phi(X)]$ with the same precision. Initially, quantization techniques have been developed in the field of signal theory. Recently, several applications to efficient simulation of diffusion processes and Monte Carlo methods have been proposed (see for example [1, 2, 5, 7, 22, 28] and the survey articles [25, 29]). A mathematical treatment of this theory can be found in [14].

Let X be the solution to some SDE living in $\mathbb{R}^{n'}$ and Υ be a functional on the space $\mathcal{C}([0,T];\mathbb{R}^{n'})$ of continuous functions. We are interested in computing the quantity

$$J = \mathbb{E}[\Upsilon(X)]. \tag{1}$$

Under suitable hypotheses, provided that the SDE

$$X_t = x + \int_0^t \sigma(X_s)\,\mathrm{d}B_s + \int_0^t b(X_s)\,\mathrm{d}s \tag{2}$$

solved by X has a unique strong solution for a n-dimensional Brownian motion, the Yamada-Watanabe theorem [15] states that there exists a measurable map Ψ on $\mathcal{C}([0,T];\mathbb{R}^n)$ such that $X = \Psi(B)$. Hence,

$$J = \mathbb{E}[\Upsilon \circ \Psi(B)].$$

Although no simple expression holds for this map Ψ in general, various schemes (Euler scheme, Milstein scheme, ...) allow one to construct an approximation X^{app} of X. With a scheme which depends on the underlying Brownian motion B, we write $X^{\mathrm{app}} = \Psi^{\mathrm{app}}(B)$. In general, Ψ^{app} is not a function of the whole path B but only of a part of it. For example, the Euler scheme (or the Milstein scheme when $n = 1$) depends only on the Gaussian vector $(B_{t_1^n} - B_0, \dots, B_{t_n^n} - B_{t_{n-1}^n})$ with $t_k^n = kT/n$.

The SDE (2) may be written under its Stratonovich form

$$X_t = x + \int_0^t \sigma(X_s) \circ \mathrm{d}B_s + \int_0^t c(X_s)\,\mathrm{d}s$$

with, for $\sigma = (\sigma_{i,j})_{i,j=1,\dots,n}$,

$$c_i(x) = b_i(x) - \sum_{\substack{j=1,\dots,n' \\ k=1,\dots,d}} \frac{1}{2} \frac{\partial \sigma_{i,k}(x)}{\partial x_j} \sigma_{j,k}(x).$$

For a piecewise smooth path U, one can also consider the ordinary differential equation (ODE)

$$Y_t = x + \int_0^t \sigma(Y_s)\mathrm{d}U_s + \int_0^t c(Y_s)\,\mathrm{d}s. \tag{3}$$

If σ is a function of class $\mathcal{C}^2$ and c a function of class $\mathcal{C}^1$ which have at most a linear growth, then the ODE has a unique solution Y and there exists a measurable map Θ on the set $\mathcal{C}^1_{\mathrm{p}}([0,T];\mathbb{R}^n)$ of continuous functions that are piecewise $\mathcal{C}^1$ such that $Y = \Theta(U)$. For a family $(U^n)_{n\in\mathbb{N}}$ of paths in $\mathcal{C}^1_{\mathrm{p}}([0,T];\mathbb{R}^n)$ that converges uniformly to a realization $B(\omega)$ of the Brownian motion, it is not always true that $\Theta(U^n)$ converges to the realization $X(\omega) = \Psi(B)(\omega)$. Wong-Zakai's theorem [15] provides some cases where the convergence holds. In dimension $n = 1$, this will be essentially true due to the approach of H. Doss [8] and H. Sussmann [31]. When $n > 1$, the recent theory of rough paths helps us to understand when the convergence holds, under stronger conditions on the regularity of the coefficients [10, 11, 16, 17, 23, 24]. The family of paths $(U^n)_{n\in\mathbb{N}}$ should then be carefully chosen.

Throughout this article, we assume that the coefficient σ is of class $\mathcal{C}^2$ with a bounded first order derivative and a second order derivative which is ϵ-Hölder continuous for some $\epsilon > 0$. The coefficient c is assumed to have a bounded first order derivative which is moreover ϵ-Hölder continuous. Note that these hypotheses may be too strong if the dimension of the underlying Brownian motion is equal to $n = 1$, where there is no need of the Hölder continuity of the derivatives of the coefficients. In addition, in the case of the Heston model, this condition is not true. However, this is not a real issue when we consider paths for the volatility that stay away from zero. We simply need these regularity conditions in order to apply the theory of rough paths that gives a pathwise interpretation of SDEs. Of course, one may take advantage of smoother coefficients in order to use sophisticated schemes (Milstein, Runge-Kutta, ...) to solve numerically the SDE X and the ODE Y, while we only use the Euler scheme in our numerical examples.

The algorithm we propose relies on the following steps:

- Do N times
  - Simulate an approximation $X^{\mathrm{app}} = \Psi^{\mathrm{app}}(B)$ of X with a numerical scheme (Euler scheme, ...). Although we are free to choose the scheme, one shall record the realization of the vector of the marginals $\boldsymbol{\beta} = (B_{t_1}, \ldots, B_{t_d})$ of the underlying Brownian motion at a given set of times. This is our only constraint on the scheme.
  - With the help of a quantization technique, compute from $\boldsymbol{\beta}$ a smooth approximation $\widehat{B}$ of the Brownian motion and compute an approximation $\Upsilon(\Theta^{\mathrm{app}}(\widehat{B}))$ of $\Upsilon(\Theta(\widehat{B}))$ by solving numerically the ODE (3) with $U = \widehat{B}$.

  - Use $\Upsilon(\Theta^{\mathrm{app}}(\widehat{B}))$ as a control variate for $\Upsilon(X^{\mathrm{app}})$. This last quantity is itself an approximation of $\Upsilon(X) = \Upsilon(\Psi(B))$.
- Return the mean of the quantities $\Upsilon(X^{\mathrm{app}}) - \Upsilon(\Theta^{\mathrm{app}}(\widehat{B})) + \mathbb{E}[\Upsilon(\Theta^{\mathrm{app}}(\widehat{B}))]$.

Written in some abstract way, the classical Monte Carlo estimator J^N for J in (1) is

$$J^N = \frac{1}{N}\sum_{i=1}^{N} \Upsilon \circ \Psi^{\mathrm{app}}(B^{(i)}),$$

for N independent copies $B^{(1)}, \dots, B^{(N)}$ of the underlying Brownian motion. Note that this estimator may be biased due to the use of the approximation scheme Ψ^{app}, but we do not discuss this classical issue here. The variance of J^N is equal to $\mathrm{Var}(\Upsilon \circ \Psi^{\mathrm{app}}(B))/N$, and depends on the functions Ψ^{app} and Υ.

The estimator given by our algorithm is — with a minor improvement carried by the choice of good values of κ —,

$$\widehat{J}^N = \frac{1}{N}\sum_{i=1}^{N} \big(\Upsilon \circ \Psi^{\mathrm{app}}(B^{(i)}) - \kappa(\Upsilon \circ \Theta^{\mathrm{app}}(\widehat{B}^{(i)}) - \mathbb{E}[\Upsilon \circ \Theta^{\mathrm{app}}(\widehat{B})])\big),$$

where $\widehat{B}^{(i)}$ is a copy of the path $\widehat{B}$ associated to $B^{(i)}$ through its marginals at given times $(t_1, \dots, t_d)$. The constant κ may be taken equal to 1. But one can also choose the optimal constant

$$\kappa_{\mathrm{opt}} = \mathrm{Cov}(\Upsilon \circ \Psi^{\mathrm{app}}(B), \Upsilon \circ \Theta^{\mathrm{app}}(\widehat{B}))/\mathrm{Var}(\Upsilon \circ \Theta^{\mathrm{app}}(\widehat{B})), \tag{4}$$

which can be approximated using for example some schemes presented in [9, 32]... In order to simplify our presentation, we set $\kappa = 1$.

The feature of a good control variate associated to a Brownian motion B consists in finding $\widehat{B}$ such that the realizations of $\widehat{B}$ are close to to the corresponding realizations of B and $\mathbb{E}[\Upsilon \circ \Theta(\widehat{B})]$ can easily be computed.

In our case, we are willing to find a finite number of trajectories which "summarize" the behavior of the Brownian motion. A possibility consists in using a finite number of marginals $\boldsymbol{\beta} = (B_{t_0}, B_{t_1}, \dots, B_{t_d})$ of the Brownian motion, to quantize the Gaussian vector of increments $(B_{t_1} - B_{t_0}, \dots, B_{t_d} - B_{t_{d-1}})$ and then to use linear interpolations to link the points. Another possibility consists in using for $\widehat{B}$ a quantization of the truncated version of the Karhunen-Loève decomposition, that is

$$\widehat{B}_t = \sum_{i=1}^{m} \sqrt{\lambda_i}\widehat{\xi}_i f_i(t) \text{ for } t \in [0, T], \tag{5}$$

where the f_k's are d elements of a basis of $\mathrm{L}^2([0,T];\mathbb{R})$, the λ_k's are appropriate scalars, and $(\widehat{\xi}_1, \dots, \widehat{\xi}_m)$ is a quantization of a Gaussian vector $(\xi_1, \dots, \xi_m)$. This means that $(\widehat{\xi}_1, \dots, \widehat{\xi}_m)$ takes it values in a

finite set of M points that have been optimized in some way. This quantization is called a *functional quantization* and the feature of this approach is that the truncation number m can be rather small ($m \leq 10$ for example) to properly catch the global behavior of a trajectory.

Hence, quantization provides us with a map $\boldsymbol{\beta} \mapsto \widehat{B}$ such that $\Theta(\widehat{B})$ is "close" to $\Psi(B)$ (so $\Theta^{\text{app}}(\widehat{B})$ is close to $\Psi^{\text{app}}(B)$). And $\mathbb{E}[\Upsilon \circ \Theta^{\text{app}}(\widehat{B})]$ can be computed since $\widehat{B}$ takes its values in a finite set of continuous paths.

The functional quantization has been proposed and studied by G. Pagès *et al.* as a way to compute quickly an approximation of $\mathbb{E}[\Upsilon \circ \Psi(B)]$ with the expression $\mathbb{E}[\Upsilon \circ \Theta^{\text{app}}(\widehat{B})]$ (see [22, 30]). He also proposed to use functional quantization as a control variate by simulating B with a truncated Karhunen-Loève decomposition. In [5], quantization is also proposed as a way to perform variance reduction through stratification.

Here, our approach is a bit different since we are not bound to replace the Brownian motion by its truncated Karhunen-Loève decomposition. Using a least square estimators, we construct $\widehat{B}$ from $(B_{t_1}, \dots, B_{t_d})$, where $(t_1, \dots, t_d)$ are fixed times.

This is important, especially when diffusion processes are simulated, since one can choose the most convenient way to do so and only record the vector of marginals of the underlying Brownian motion.

Note that in our algorithm, the Brownian motion may be replaced by any other Gaussian process, such as Ornstein-Uhlenbeck processes or Brownian bridge.

2. The algorithm

2.1. **The idea.** To simplify the presentation, we assume that the dimension of the underlying Brownian motion is $n = 1$. To deal with the multi-dimensional case, it is sufficient to apply the algorithm to each component of the Brownian motion.

The idea behind the algorithm is the following, which we summarize in Figure 1: we fix some integers d and $m \leq d$, some times $0 \leq t_1 < \cdots < t_d \leq T$, as well as an orthonormal basis $\{f_i\}$ of $\mathrm{L}^2([0,T];\mathbb{R})$. We compute $\lambda_i = \operatorname{Var}\left(\int_0^T f_i(s)B(s)\,\mathrm{d}s\right)$. Given a realization $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)$ of $(B_{t_1}, \dots, B_{t_d})$, we define $\boldsymbol{\xi} = (\xi_1, \dots, \xi_m)$ as

$$\boldsymbol{\xi} = \operatorname{argmin}_{(\zeta_1,\dots,\zeta_m)\in\mathbb{R}^m} \sum_{i=1}^{d} \left(\sum_{j=1}^{m} \zeta_j \sqrt{\lambda_j} f_j(t_i) - \beta_{t_i} \right)^2, \tag{6}$$

and choose a quantization $\widehat{\boldsymbol{\xi}} = (\xi_1, \dots, \xi_m)$ of this vector with at most Q points. Note that there are several ways to do so: One can compute the optimal quantization of the vector $\boldsymbol{\xi}$, or a product quantization (*i.e.* each component is quantized independently), or any mixing of these two approaches [27]. Besides, the quantization is chosen by optimizing

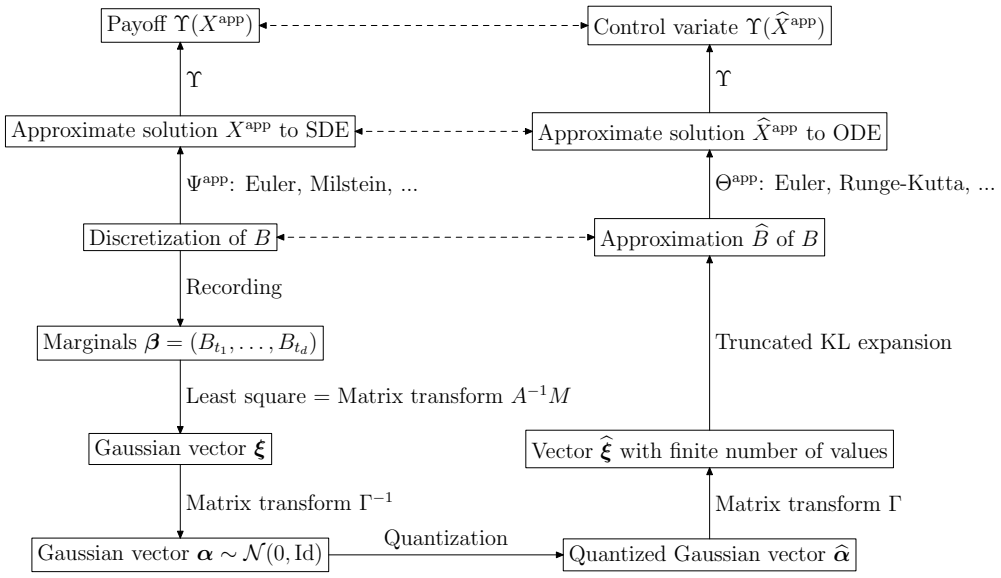

FIGURE 1. The scheme

the norm $\mathbb{E}[|\boldsymbol{\xi} - \widehat{\boldsymbol{\xi}}|^2]$ among all the possible random variables $\widehat{\boldsymbol{\xi}}$ with at most Q point. The optimal choice of $\widehat{\boldsymbol{\xi}}$ may indeed take Q' values with Q' smaller than Q.

We then use $\widehat{B}$ defined by (5) as a control variate for B. This way, $\widehat{B}$ takes its values in a set $\{\widehat{B}^{[i]}\}_{i=1,\dots,Q'}$ of Q' trajectories in $\mathcal{C}([0,T],\mathbb{R})$. Each trajectory $\widehat{B}^{[i]}$ is associated with a value of $\widehat{\boldsymbol{\xi}}$. Since all the possible values of $\widehat{\boldsymbol{\xi}}$ are known, $\Upsilon \circ \Theta^{\text{app}}(\widehat{B}^{[i]})$ can be computed prior to any simulation. For $p^{[i]} = \mathbb{P}[\widehat{\boldsymbol{\xi}} = (x_1^{[i]}, \dots, x_m^{[i]})]$ for the possible points $(x_1^{[i]}, \dots, x_m^{[i]})$ of the quantized vector, then

$$\mathbb{E}[\Upsilon \circ \Theta^{\text{app}}(\widehat{B})] = \sum_{i=1}^{Q} p^{[i]} \Upsilon \circ \Theta^{\text{app}}(\widehat{B}^{[i]})$$

and this quantity may also be precomputed.

2.2. **Preliminary computations.** Let M be the $d \times m$-matrix defined by $M_{i,j} = \sqrt{\lambda_j} f_j(t_i)$, $i = 1, \dots, d$, $j = 1, \dots, m$. We assume:

(H) The rank of M is m.

Solving the linear least squares is then equivalent to solving $M^T \boldsymbol{\xi} = \boldsymbol{\beta}$. As M is not necessarily a square matrix, one can indeed equivalently solve

$$A\boldsymbol{\xi} = M\boldsymbol{\beta} \text{ with } A = MM^T.$$

The matrix A is an invertible $m \times m$-symmetric matrix and then $\boldsymbol{\xi} = A^{-1}M\boldsymbol{\beta}$.

Numerical remark 1. Numerically, one should be careful when inverting the matrix A to use a robust enough algorithm, such as the Cholesky decomposition or the QR method [13].

The vector of marginals $\boldsymbol{\beta} = (B_{t_1}, \ldots, B_{t_n})$ is a Gaussian vector and it is easily seen that $\boldsymbol{\beta} = L\Delta\boldsymbol{\alpha}$ where $\boldsymbol{\alpha} \sim \mathcal{N}(0, \mathrm{Id}_{d\times d})$, L is the lower triangular $d \times d$-matrix defined by $L_{i,j} = 1$ if $i \leq j$ and $L_{i,j} = 0$ otherwise, and Δ is the diagonal matrix with $\sqrt{t_1}, \sqrt{t_2 - t_1}, \ldots, \sqrt{t_d - t_{d-1}}$ as diagonal terms. Since, $\boldsymbol{\xi} = A^{-1}M\boldsymbol{\beta}$ and $\mathrm{Cov}(\boldsymbol{\alpha}) = \mathrm{Id}$,

$$\mathrm{Cov}(\boldsymbol{\xi}) = (A^{-1}ML\Delta)(A^{-1}ML\Delta)^T.$$

In addition, there exists a unique, invertible (under Hypothesis (H)), lower triangular $d\times d$-matrix Γ such that $\mathrm{Cov}(\boldsymbol{\xi}) = \Gamma\Gamma^T$, and then $\boldsymbol{\alpha} = \Gamma^{-1}\boldsymbol{\xi} \sim \mathcal{N}(0, \mathrm{Id}_{m\times m})$ is a m-vector of independent Gaussian random variables.

Numerical remark 2. From the numerical point of view, the matrix Γ will be computed with a Cholesky decomposition [13, Theorem 5.2.3, p. 88].

The matrix Γ is important for practical purposes, since the quantization of $\boldsymbol{\xi}$ will be deduced from the quantization of $\boldsymbol{\alpha} \sim \mathcal{N}(0, \mathrm{Id}_{m\times m})$, which depends only on our choice of m and Q.

2.3. **The algorithm.** The algorithm is then the following.

(1) Choose a number d of marginals to use, a number $m \leq d$ of coefficients and the times $0 < t_1 < \ldots < t_d \leq T$.
(2) Choose an orthonormal basis $\{f_j\}_{j\geq 1}$ of $\mathrm{L}^2([0,T];\mathbb{R})$ and compute the corresponding weights $\{\lambda_j\}_{j\geq 1}$ in the Karhunen-Loève decomposition of the Brownian motion. Compute the matrices Γ, Γ^{-1} and $R = \Gamma^{-1}A^{-1}M$.
(3) Choose a quantization of m-dimensional vectors $\boldsymbol{\alpha} \sim \mathcal{N}(0, \mathrm{Id}_{m\times m})$, that is a discrete set $\{\widehat{\boldsymbol{\alpha}}^{[i]}\}_{i=1,\ldots,Q'}$ called the *code book* and a "projection" π which associates to a realization of $\boldsymbol{\alpha}$ an element of the code book. Compute for $i = 1, \ldots, Q'$, $p^{[i]} = \mathbb{P}[\pi(\boldsymbol{\alpha}) = \widehat{\boldsymbol{\alpha}}^{[i]}]$, $\phi^{[i]} = \Upsilon \circ \Theta^{\mathrm{app}}(\widehat{B}^{[i]})$ where

$$\widehat{B}_t^{[i]} = \sum_{j=1}^{m} \widehat{\xi}_j^{[i]} \sqrt{\lambda_j} f_j(t) \text{ with } \widehat{\boldsymbol{\xi}}^{[i]} = \Gamma\widehat{\boldsymbol{\alpha}}^{(i)},$$

and $\widehat{J} = \sum_{i=1}^{Q'} \phi^{[i]} p^{[i]}$.
(4) Set $J \leftarrow 0$.
(5) For i from 1 to N do

(a) Simulate a realization $B(\omega)$ of the Brownian motion B, record $\boldsymbol{\beta}(\omega) = (B_{t_1}(\omega), \dots, B_{t_d}(\omega))$ and compute $\phi = \Upsilon \circ \Psi^{\text{app}}(B(\omega))$.
(b) Compute $\boldsymbol{\alpha}(\omega) = R\boldsymbol{\beta}(\omega)$ which is a realization of $\boldsymbol{\xi} \sim \mathcal{N}(0, \mathrm{Id}_{m\times m})$.
(c) Find the index i such that $\widehat{\boldsymbol{\alpha}}^{[i]} = \pi(\boldsymbol{\alpha})$.
(d) Set $J \leftarrow J + \phi - \phi^{[i]}$.

(6) Return $N^{-1}J + \widehat{J}$, which is an estimator of $\mathbb{E}[\Upsilon \circ \Psi(B)]$.

Numerical remark 3. As explained in the Numerical remarks 1 and 2, we use in order to compute A^{-1}, Γ and Γ^{-1} a Cholesky decomposition. The feature of our algorithm is that it uses matrices with reasonably small size (of order 10×10).

Remark 1. For the Brownian motion, there is a natural choice of a basis $\{f_j\}_{j\geq 1}$, which is the set of eigenfunctions of the covariance operator $f \in \mathcal{C}([0,T];\mathbb{R}) \mapsto \left(t \mapsto \int_0^{+\infty} (s \wedge t) f(s)\,\mathrm{d}s\right)$ of the Brownian motion. In this case,

$$\lambda_j = \left(\frac{T}{\pi\left(j-\frac{1}{2}\right)}\right)^2 \text{ and } f_j(t) = \sqrt{\frac{2}{T}} \sin\left(\pi\left(j-\frac{1}{2}\right)\frac{t}{T}\right) \tag{7}$$

for $j \geq 1$. Let us note that the algorithm may be applied to any Gaussian process. However, it is not always so easy to get the eigenfunctions of the corresponding covariance operator (see [28] for example).

One may also use another natural basis, the Schauder basis, although this one is not orthonormal. Numerical experiments on the Brownian motion give results that are not as good as the ones obtained with the basis above.

Remark 2. In this algorithm, we have chosen to use the quantization of Gaussian random vectors with covariance matrix equal to identity. With slight changes in the algorithm, we can either quantize the vector $\Delta^{-1}L^{-1}\boldsymbol{\beta} \sim \mathcal{N}(0, \mathrm{Id}_{d\times d})$ of size $d \geq m$ or the vector $\boldsymbol{\xi} \sim \mathcal{N}(0, (A^{-1}ML\Delta)(A^{-1}ML\Delta)^T)$, which have covariances matrices different from identity. Our choice is justified by the numerical cost of the quantization procedure and the need for vectors to have the smallest size possible. Quantizations of random m-vectors with distribution $\mathcal{N}(0, \mathrm{Id}_{m\times m})$ may be prepared prior to any simulation [27] and is then suitable for any SDE. In addition, tabulated values can be found in the WEB site [26], thanks to the works of G. Pagès and J. Printems.

3. Numerical tests

We have performed numerical tests with 3 classical models in finance: the Black & Scholes model

$$X_t = X_0 + \int_0^t \sigma X_s \, \mathrm{d}B_s + \int_0^t \nu X_s \, \mathrm{d}s = X_0 \exp\left(\sigma B_t + \left(\nu \frac{1}{2}\sigma^2\right) t\right),$$

the Cox-Ingersoll-Ross (CIR) model

$$X_t = X_0 + \int_0^t k(a - X_s)ds + \sigma \int_0^t \sqrt{X_s} dB_s$$

and the Heston model

$$\begin{cases} V_t = V_0 + \int_0^t k(a - V_s)ds + \sigma \int_0^t \sqrt{V_s} dB_s^1, \\ X_t = X_0 + \int_0^t \nu X_s ds + \int_0^t \sqrt{V_s} X_s dB_s^2, \\ \langle B^1, B^2 \rangle_t = \rho t, \end{cases}$$

where B^1 and B^2 are two Brownian motions. The marginals $(B_{t_1}, \ldots, B_{t_d})$ are taken at times $t_i = iT/10$, $i = 1, \ldots, 10$.

To simulate the Black & Scholes model, we used the exponential representation, so that $X = \Psi(B)$ where $\Psi(B) = (x \exp(\sigma B_t + \mu t - \sigma^2 t/2))_{t \geq 0}$. We also use $\Theta = \Psi$.

To simulate the CIR and the Heston models, we used the modification of the Euler scheme proposed by Berkaoui, Bossy and Diop [4]. Any of the schemes studied in [21] could have been used, as long as one record the values of $(B_{t_1}, \ldots, B_{t_d})$ while performing the simulations (of course, one may use much more marginals of B).

With these schemes, we produce N independent realizations $X^{(1)}, \ldots, X^{(N)}$ (of approximations) of the path X. In all the cases, the Brownian motion is discretized using a time step $\delta t = 0.01$. From the realizations $(B_{t_1}^{(i)}, \ldots, B_{t_d}^{(i)})$, $i = 1, \ldots, N$, of the marginals, we obtain through our quantization procedure quantized paths $\widehat{B}^{(i)}$, $i = 1, \ldots, N$ that are taken in the finite set $\{\widehat{B}^{[j]}\}_{j=1,\ldots,Q'}$. We denote by $\mathfrak{q}(i)$ the integer such that $\widehat{B}^{(i)} = \widehat{B}^{[\mathfrak{q}(i)]}$. To compute $X^{[j]} = \Theta^{\mathrm{app}}(\widehat{B}^{[j]})$ along the quantized paths, we use the Euler scheme for ODE after having transformed the SDEs into Stratonovich SDEs, and then replace the Brownian motion by $\widehat{B}^{[j]}$. For the sake of simplicity, we have also used an explicit Euler scheme with the same timestep as the Euler scheme for SDEs. Let us note that in our case, although a wide range of techniques are available (Runge-Kutta, implicit schemes, ...), we prefer keep a simple scheme. Indeed, the precision for the resolution of the ODE is not a real issue. A bias induced by the scheme has a very little impact on the Monte Carlo estimator: whatever the scheme, the control variate has a zero mean.

We compare three estimators v, v' and v'_κ of the empirical variance of $\Upsilon(X_T)$ associated to the three empirical means

$$\begin{aligned}
\mu &= \frac{1}{N}\sum_{i=1}^{N} \Upsilon(X^{(i)}),\\
\mu' &= \frac{1}{N}\sum_{i=1}^{N} \big(\Upsilon(X^{(i)}) - (\Upsilon(\widehat{X}^{[\mathfrak{q}(i)]}) - \mathbb{E}[\Upsilon(\widehat{X})])\big),\\
\text{and } \mu'_\kappa &= \frac{1}{N}\sum_{i=1}^{N} \big(\Upsilon(X^{(i)}) - \kappa^{(i)}(\Upsilon(\widehat{X}^{[\mathfrak{q}(i)]}) - \mathbb{E}[\Upsilon(\widehat{X})])\big),
\end{aligned}$$

where $\kappa^{(i)}$ is constructed from $X^{(1)}, \dots, X^{(i)}$ to approximate κ_{opt} in (4).

Remark 3. Unless one knows how to simulate exactly X from B, these estimators are biased. Yet the bias is only induced by the choice of the scheme used to compute X. The control variate does not induce any supplementary bias. We no longer discuss this issue which is out of the scope of this article.

The results are summarized in Tables 2, 3 and 4. The mean of the estimator is denoted by μ (in all the cases, the relative error between the different empirical estimator is very small), which is then the price of the option.

The efficiency of a variance reduction technique shall take the computation time into account. In Tables 2, 3 and 4, T_{s} is the time spent to compute the empirical mean μ without the control variate, while T'_{s} is the time spent to compute the empirical mean μ with the control variate (time θ_{simul} per simulation). With our control variate, the extra computational time is due to (a) the initialization of the algorithm (time T_{init}), which depends on the number of points of quantization, and (b) the research of the nearest neighbor of the Gaussian coefficients among the quantized points (time θ_{search} per simulation). To research the vector's nearest neighbor, we have used a method called *kd-tree* [3] which appears to be efficient, while for a scalar, we used a search by dichotomy.

The method has been applied to several call options, for three values of K: $K = 0.5$ (In the money, ITM), $K = 1.0$ (At the money, ATM)

and $K = 1.3$ (Out of the money, OTM). The options are

$$\begin{aligned}
\text{European: } & \Upsilon(X) = (X_T - K)_+, \\
\text{Asian: } & \Upsilon(X) = \left(\frac{1}{T} \int_0^T X_s \, \mathrm{d}s - K \right)_+, \\
\text{Asian binary: } & \Upsilon(X) = \mathbf{1}_{\frac{1}{T}\int_0^T X_s \, \mathrm{d}s > K}, \\
\text{Ratchet: } & F_0 = 0, \ F_i = (K + F_{i-1} + \rho_i)_+ \text{ and } \rho_i = \frac{X_{t_i} - X_{t_{i-1}}}{(t_i - t_{i-1}) X_{t_i}}, \\
\text{Lookback fixed: } & \Upsilon(X) = X_T - \min_{t \in [0,T]} X_t, \\
\text{Lookback floating: } & \Upsilon(X) = (\max_{t \in [0,T]} X_t - K)_+, \\
\text{Cliquet: } & \max \left\{ F_0, \min \left\{ C_0, \sum_{i=1}^N \max\{F_i, \min\{C_i, R_i\}\} \right\} \right\}
\end{aligned}$$

where the floors $F_i = 0$ and the caps $C_i = K$, and $R_i = X_{t_i} - X_{t_{i-1}}$. For the cliquet and ratchet options, we use $t_i = iT/12$.

Thus, with N simulations and Q quantized path, the ratio of the times $T'_{\mathrm{s}}/T_{\mathrm{s}}$ is given by

$$\frac{T'_{\mathrm{s}}}{T_{\mathrm{s}}} = \frac{T_{\mathrm{init}}(Q) + N\theta_{\mathrm{search}} + N\theta_{\mathrm{simul}}}{N\theta_{\mathrm{simul}}}.$$

Let us note that the initialization phase implies the resolution of Q ordinary differential equations, so that $T_{\mathrm{init}} \approx Q\theta_{\mathrm{simul}}$.

Thus, in Tables 2, 3 and 4, we give the *efficiency ratio* for 100,000 and 1,000,000 simulations. We call the *efficiency ratio* the quantities

$$e = \frac{T_{\mathrm{s}}}{T'_{\mathrm{s}}} \times \frac{v}{v'} \text{ and } e_\kappa = \frac{T_{\mathrm{s}}}{T'_{\mathrm{s}}} \times \frac{v}{v'_\kappa}.$$

When these quantities are close to 1, then the method provides no gain. It is then simpler to increase the size of the sample without any correction to obtain the same result. When these quantities are greater than 1, then for a given precision, it corresponds to the gain factor in execution time. If these quantities are smaller than 1, then the methods shall not be used.

In Table 1, we also give the prices obtained by computing the options prices by computing $\sum_{j=1}^{Q'} p^{[j]} \Upsilon(X^{[j]})$, that is the quantized price. The result of this table should be compared with the values given in the column μ (the values of the Monte Carlo estimator) of Tables 2, 3 and 4. It is not surprising to see that the quantized prices can sometimes be far from the estimated prices for the Ratchet, Lookback and Cliquet options. Although the quantized price is computed quickly, this estimation is not always reliable.

For most of the simulations, the variance reduction technique is efficient or very efficient: The efficiency ratios above 5 have been written

Option	B&S	CIR	Heston
European OTM	0.015	0.024	0.063
European ATM	0.106	0.195	0.140
European ITM	0.549	0.697	0.462
Asian OTM	0	1.0×10^{-4}	0.017
Asian ATM	0.053	0.099	0.081
Asian ITM	0.513	0.590	0.466
Asian Binary OTM	0	0.029	0.017
Asian Binary ATM	0.520	0.809	0.081
Asian Binary ITM	1	1	0.467
Ratchet OTM	72.7	81.16	63.8
Ratchet ATM	56.2	64.67	48.7
Ratchet ITM	29.3	37.35	26.7
Lookback fixed OTM	0.018	0.029	0.090
Lookback fixed ATM	0.143	0.228	0.222
Lookback fixed ITM	0.643	0.728	0.723
Lookback floating ITM	0.135	0.218	0.188
Cliquet OTM	0.165	0.236	0.265
Cliquet ATM	0.165	0.235	0.223
Cliquet ITM	0.084	0.097	0.008

TABLE 1. Table of the prices computed using the same quantization of the Brownian motion as the one of Table 2, 3 and 4.

in bold. For very small values of prices, the methods may be inefficient (the figures are then emphasized). In this case, using the estimator μ'_κ improves the quality of the estimator, but does not provide significant gains with respect to μ. Most of the time, using μ'_κ instead of μ' does not significantly improve the results. In addition, the method may also be inefficient when the payoff is discontinuous, as one may expect.

4. Conclusion

We have presented a new variance reduction technique which belongs to the class of hybrid quantization-Monte Carlo method related to the simulation of SDE.

This method relies on functional quantization and may be combined with other variance reduction techniques applied to the Brownian motion.

Numerical tests show the robustness and the efficiency of the method which has good gain factors — up to order 10 — even for complex payoffs. However, this method is not suitable for SDEs with a large number of underlying Brownian motions — as the number of quantized paths grows very quickly — and to estimate small quantities or prices.

As the end-user is free to fix a lot of parameters (number of quantization points, way to decompose the vectors, ...), the numerical results exposed here may be improved by tuning the parameters to get the best balance between a small ratio of variances and the extra computational cost. Note that however, this tuning should be model dependent.

Appendix A. Approximation of the solution of the SDE using approximation of the Brownian motion

A.1. **Ordinary differential equation: the map Θ.** For a coefficient σ of class $\mathcal{C}^2$ and a coefficient c of class $\mathcal{C}^1$ with at most a linear growth, we have introduced the map Θ which transforms a path $U \in \mathcal{C}^1_{\mathrm{p}}([0,T];\mathbb{R}^n)$ into the unique solution Y to (3).

Let $\dot{U}$ be the derivative of U with respect to the time and set $\|U\|_{1\text{-var}} = \|\dot{U}\|_\infty$. From the Gronwall Lemma, it is standard that $\|\Theta(U)\|_\infty \leq K(1+\|U\|_{1\text{-var}})$, where K depends on the Lipschitz constant of σ and b, as well as the starting point x and the horizon T.

Then for two paths U and V, again from the Gronwall lemma,

$$\|\Theta(U)-\Theta(V)\|_\infty \leq K'(1+\|U\|_\infty)\|U-V\|_{1\text{-var}},$$

where K depends on the Lipschitz constant of σ and b, as well as the starting point x and the horizon T.

Note that if Θ^{app} is defined using the Euler scheme for ODE, then a similar computation with the discrete Gronwall inequality leads to

$$\|\Theta^{\mathrm{app}}(U)-\Theta^{\mathrm{app}}(V)\|_\infty \leq K''(1+\|U\|_\infty)\|U-V\|_{1\text{-var}},$$

for some constant K'' that depends on the Lipschitz constant of σ and b, as well as the starting point x and the horizon T.

If $n=1$, then we have a better estimate: it follows from the approach proposed by H. Doss [8] and H. Sussmann [31] (See also [15] for example) that Θ is locally Lipschitz continuous with respect to the uniform norm when σ has bounded first and second derivatives.

For an orthonormal basis $\{f_i\}_{i=1,2,\dots}$ of $\mathrm{L}^2([0,T];\mathbb{R})$, we consider the subset $\mathfrak{K}$ of $\mathcal{C}^1_{\mathrm{p}}([0,T];\mathbb{R}^n)$ of paths of type

$$X^j_t = \sum_{i=1}^m \xi^j_i f_i(t), \text{ for } j=1,\dots,n,$$

for some integer $m>0$ and a family $\{\xi^j_i\}_{i=1,\dots,m,j=1,\dots,n}$ of scalars.

If the f_i belong to $\mathcal{C}^1_{\mathrm{p}}([0,T];\mathbb{R})$, for $U=\{\sum_{i=1}^m \xi^j_i f_i(t)\}_{j=1}^m$ and $V=\{\sum_{i=1}^m \zeta^j_i f_i(t)\}_{j=1}^m$ in $\mathfrak{K}$, then

$$\|\Theta(U)-\Theta(V)\|_\infty \leq C\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|, \tag{8}$$

with C a constant which depends only on m, n, T and the coefficients σ and b, where $\boldsymbol{\xi}=\{\xi^j_i\}_{i=1,\dots,m,j=1,\dots,n}$, $\boldsymbol{\zeta}=\{\zeta^j_i\}_{i=1,\dots,m,j=1,\dots,n}$, and $\|\cdot\|$ is a matrix norm.

Again, if one uses the Euler scheme for ODE to construct Θ^{app}, then

$$\|\Theta^{\text{app}}(U) - \Theta^{\text{app}}(V)\|_\infty \leq C' \|\boldsymbol{\xi} - \boldsymbol{\zeta}\|,$$

for some constant C' which depends only on m, n, T and the coefficients σ and b.

A.2. **Is our control variate a good approximation of the diffusion?** As explained in the introduction, a good control variate for $\Upsilon(X)$ is given by a random variable $\Upsilon(Y)$ such that $\mathbb{E}[\Upsilon(Y)]$ may be computed and Y is strongly correlated and close to X.

Assuming that Υ is continuous, one may wonder if the process $\widehat{X}$ constructed with our scheme is close to X, the solution to the SDE consider.

Let us denote by $\boldsymbol{\xi}^{KL}$ the first m coefficients of the truncated Karhunen-Loève decomposition B^m of the Brownian motion B:

$$B_t^m = \sum_{i=1}^m \xi_i^{KL} \sqrt{\lambda_i} f_i(t).$$

Let us also consider

$$\boldsymbol{\xi}(d) = \text{argmin}_{\boldsymbol{\zeta}(d) \in \mathbb{R}^m} \sum_{i=1}^d \left(B_{t_i} - \sum_{j=1}^m \zeta_j(d) \sqrt{\lambda_j} f_j(t_i) \right)^2. \tag{9}$$

We then set

$$\widetilde{B}_t^m = \sum_{i=1}^m \xi_i(d) \sqrt{\lambda_i} f_i(t).$$

Using the above notation, the Gaussian vector $\boldsymbol{\xi}(d)$ may be expressed as $\Gamma(d)\boldsymbol{\alpha}(d)$ for some invertible matrix $\Gamma(d)$ such that $\Gamma(d)\Gamma(d)^T = \text{Cov}(\boldsymbol{\xi}(d))$, so that $\boldsymbol{\alpha}(d) \sim \mathcal{N}(0, \text{Id})$. The approximation $\widehat{B}^m$ of the Brownian motion is

$$\widehat{B}_t^m = \sum_{i=1}^m \sqrt{\lambda_i} \Gamma^{-1}(d) \widehat{\alpha}_i f_i(t), \tag{10}$$

where $\widehat{\boldsymbol{\alpha}}(d)$ is the quantization of $\boldsymbol{\alpha}(d)$.

Obviously, the distance between $\widehat{X}$ and X is given by $\|\Psi(X) - \Theta(\widehat{X})\|_\infty$ and may be controlled by the sum of three terms:

- The distance between $\Psi(B)$ and $\Theta(B^m)$, where B^m is the truncated Karhunen-Loève decomposition of the Brownian motion.
- The distance between $\Theta(B^m)$ and $\Theta(\widetilde{B}^m)$.
- The distance between $\Theta(\widetilde{B}^m)$ and $\Theta(\widehat{B}^m)$.

Obviously with (8), the distances between $\Theta(B^m)$ and $\Theta(\widetilde{B}^m)$. and between $\Theta(\widetilde{B}^m)$ and $\Theta(\widehat{B}^m)$ depend only on the distance between the coefficients in the Karhunen-Loève type decompositions.

A.3. **Convergence of the solution to the least squares problem to the coefficients of the Karhunen-Loève decomposition.** The Gaussian coefficients of the Karhunen-Loève decomposition are unique provided that they are independent.

Let us denote by $\Gamma(d)$ a $d\times d$-matrix such that $\mathrm{Cov}(\boldsymbol{\xi}(d)) = \Gamma(d)\Gamma(d)^T$, where $\boldsymbol{\xi}(d)$ is obtained by solving the least squares problem (9). Thus, this matrix depends on the number of coefficients m as well as the times $(t_1,\dots,t_d)$ of the marginals of the Brownian motion.

We now consider that m is fixed and that d increases to $+\infty$.

Lemma 1. *If $t_i^d = iT/d$ for $i = 1,\dots,d$, then the $d\times d$-matrix $\Gamma(d)$ with $\Gamma(d)\Gamma(d)^T = \mathrm{Cov}(\boldsymbol{\xi}(d))$ converges to the identity as $d\to\infty$, $\boldsymbol{\xi}(d)$ converges almost surely to $\boldsymbol{\xi}^{KL}(d)$ and $\widetilde{B}_t^m(d)$ converges almost surely uniformly to the Karhunen-Loève decomposition B^m of B truncated at order the m.*

Proof. Let $M(d)$ be the matrix defined by $M_{i,j}(d) = \sqrt{\lambda_j} f_j(t_i^d)$ for $t_i^d = iT/d$, $i = 1,\dots,d$, $j = 1,\dots,m$, as in Section 2.2. Set $\boldsymbol{\beta}(d) = (B_{t_1^d},\dots,B_{t_d^d})$. Thus as $d\to\infty$,

$$\frac{T}{d}M(d)\boldsymbol{\beta}(d) = \begin{bmatrix} \frac{T}{d}\sum_{i=1}^d \sqrt{\lambda_1} f_1(t_i^d) B_{t_i^d} \\ \vdots \\ \frac{T}{d}\sum_{i=1}^d \sqrt{\lambda_m} f_m(t_i^d) B_{t_i^d} \end{bmatrix} \xrightarrow{d\to\infty} \begin{bmatrix} \sqrt{\lambda_1}\int_0^T f_1(s)B(s)\,\mathrm{d}s \\ \vdots \\ \sqrt{\lambda_m}\int_0^T f_m(s)B(s)\,\mathrm{d}s \end{bmatrix}.$$

Let $\widetilde{B}_t^m := \sum_{j=1}^m \xi_j\sqrt{\lambda_j} f_j(t)$, $t\in[0,T]$, be the truncated Karhunen-Loève decomposition of B. Since $\{f_i\}_{i\geq 0}$ is an orthonormal basis of $\mathrm{L}^2([0,T])$, one gets easily that $\lambda_i\xi_i = \sqrt{\lambda_i}\int_0^T f_i(s)B(s)ds$ for $i = 1,\dots,m$, and then that $Td^{-1}M(d)\boldsymbol{\beta}(d)$ converges to the vector $D\boldsymbol{\xi}^{KL}$ with $\boldsymbol{\xi}^{KL} = (\xi_1^{KL},\dots,\xi_m^{KL})$, where D is the diagonal matrix with $(\lambda_1,\dots,\lambda_m)$ as diagonal elements. On the other hand, one easily gets from the same argument that

$$\frac{T}{d}M(d)M(d)^T \xrightarrow{d\to\infty} \left(\sqrt{\lambda_i\lambda_j}\int_0^T f_i(s)f_j(s)ds\right)_{i,j=1,\dots,d} = D.$$

Since the λ_i are positive, $Q(d) = Td^{-1}M(d)M(d)^T$ is invertible if d is large enough and $Q(d)^{-1}$ converges to D^{-1}. It follows that

$$\boldsymbol{\xi}(d) = Q(d)^{-1}Td^{-1}M(d)\boldsymbol{\beta}(d) \xrightarrow[d\to\infty]{} D^{-1}D\boldsymbol{\xi}^{KL} = \boldsymbol{\xi}^{KL}.$$

The uniform convergence of $\widetilde{B}^m$ to B^m follows from the convergence of $\boldsymbol{\xi}(d)$ to the coefficients $\boldsymbol{\xi}^{KL}$ of the Karhunen-Loève decomposition of B, and the convergence of $\Gamma(d)$ to Id follows from the fact that $\mathrm{Cov}(\boldsymbol{\xi}(d))$ converges to $\mathrm{Cov}(\beta) = \mathrm{Id}$ and then $\Gamma(d)$ also converges to Id (see [13, Theorem 5.2.3] for example, the convergence follows from the explicit construction of the Cholesky decomposition). □

Remark 4. From a practical point of view, numerical experiments show that the covariance matrix $\mathrm{Cov}(\boldsymbol{\xi}(d))$ is already close to Id even for small values of d when the basis given in Remark 1 is used. Thus, one can take for d values of order 10 when $m \leq 10$ to get from the least squares approach a good approximation of the truncated Karhunen-Loève decomposition of B.

A.4. **Convergence of the quantized coefficients.** With (8) and (10),

$$\|\Theta(\widehat{B}^m) - \Theta(\widetilde{B}^m)\|_\infty \leq C\|\boldsymbol{\alpha}(d) - \widehat{\boldsymbol{\alpha}}(d)\|,$$

where C depends on the coefficients of Γ, m, d and the f_i.

The distance between $\Theta(\widehat{B}^m)$ and $\Theta(\widetilde{B}^m)$ therefore depends on the distance between $\widehat{\boldsymbol{\alpha}}(d)$ and $\boldsymbol{\alpha}(d)$. There are various ways to choose an optimal quantized (block quantized, product quantizer, ...). For an optimal quantizer with at most Q points (which means that the whole vector $\boldsymbol{\alpha}(d)$ is quantized), the Zador theorem [14] implies that the $\mathrm{L}^2(\mathbb{P})$-distance between $\boldsymbol{\alpha}(d)$ and $\widehat{\boldsymbol{\alpha}}(d)$ is equivalent for large Q to $C(d)/Q^{1/d}$.

Remark 5. In [22, 29, 30], G. Pagès *et al.* have studied the rate of convergence of the approximation of the Brownian motion by a truncation of the Karhunen-Loève decomposition. In their approach, they choose carefully the truncation order m as a function of Q to get the optimal rate. When constructing a control variate, we are not necessarily looking for the best approximation. In addition, it is good for efficiency to keep m fixed and rather small (in all our simulations, $m \leq 10$). Besides, we do not quantize the true coefficient of the Karhunen-Loève decomposition but only an approximation of them.

A.5. **Convergence of the SDEs.** A good control variate shall stay close to the computed quantities. Our algorithm provides for a realization of the Brownian motion B a path $\widehat{B}^m$. If B^m is the truncated Karhunen-Loève decomposition of B, we have seen that when the number of marginals d we use and the number of points of quantization Q tend to infinity, then $\Theta(\widehat{B}^m)$ converges to $\Theta(B^m)$.

Yet let us note that in practice, d and Q shall remain rather small is order to avoid a high computational cost which annihilate the whole interest of the method.

However, one may wonder how close $\Theta(B^m)$ is to $\Psi(B)$? Using the results of [6, 10, 30] based on the theory of rough paths [10, 11, 16, 17, 23, 24], $\Theta(B^m)$ converges to $\Psi(B)$ when m tends to infinity. However, the rate is not known. As said above in Remark 5, we do not aim at letting m tend to infinity. However, this result justifies the effectiveness of our approach.

Let us note however that this convergence holds due to the convergence of B^m to B in the γ-Hölder norm, $\gamma < 1/2$. The uniform convergence is not sufficient (See for example [17, 20] for explicit counter-examples). Besides, it shall be assumed that the second derivative of σ and the first derivative of c are ϵ-Hölder continuous for some $\epsilon > 0$. The distance between $\Theta(B^m)$ and $\Psi(B)$ depends on the γ-Hölder norm of $B - B^m$, but also on the γ-Hölder norms of B and B^m (see [17, 19]). For the CIR model, this result could only be invoked when the solution stay away from 0, because of the degeneracy of the diffusion coefficient at 0. The distance also depends on the minimum of the paths $\Theta(B^m)$ and $\Psi(B)$.

If the dimension of the space is equal to 1, then one may use the Doss-Sussman theory [8, 31] which implies that $\Theta(B^m)$ converges to $\Psi(B)$, since B^m converges uniformly to B. In addition, $\|\Psi(B) - \Theta(B^m)\|_\infty \leq K\|B - B^m\|_\infty$ for a constant K which depends on the maximal and minimal values B_T and B_T^m. The $\mathrm{L}^2(\mathbb{P})$-norm of the uniform distance between B and B^m is of order $\sqrt{\log n/n}$ (See [29]).

However, the gain in terms of variance reduction seems here pretty difficult to evaluate here. One does not know in general the variance of $\mathrm{Var}(\Upsilon(\Psi(B)))$ and an evaluation of the distance between $\Psi(B)$ and $\Theta(B^m)$, which is out of the scope of this article.

In addition, several simulations on the same class of model, *e.g.,* Black & Scholes, have shown no clear correlation between the gain in variance and the values of the parameters. This means that the gain seems difficult to predict because the map Ψ is highly non-linear.

Acknowledgement. The authors are indebted to Aymen Bergaoui, Madalina Deaconu, Zied Ghazai, Samuel Herrmann, Christophe Michel, Gilles Pagès, Yiqing Wang for interesting discussions about this method or problems related to quantization.

References

[1] V. Bally, G. Pagès, and J. Printems, *A quantization tree method for pricing and hedging multidimensional American options*, Math. Finance **15** (2005), no. 1, 119–168.

[2] O. Bardou, S. Bouthemy, and G. Pagès, *Optimal quantization for the pricing of swing options*, Appl. Math. Finance (2009), no. 1–2, 183–217.

[3] J.L. Bently, *Multidimensional divide-and-conquer*, Communications of the ACM **23** (1980), no. 4, 214–229.

[4] A. Berkaoui, M. Bossy, and A. Diop, *Euler scheme for SDEs with non-Lipschitz diffusion coefficient: strong convergence*, ESAIM Probab. Stat. **12** (2008), 1–11, DOI 10.1051/ps:2007030.

[5] S. Corlay and G. Pagès, *Functional quantization based stratified sampling methods* (2010). Preprint.

[6] L. Coutin and N. Victoir, *Enhanced Gaussian Processes and Applications*, ESAIM Probab. Stat. **13** (2009), 247–260, DOI 10.1051/ps:2008007.

[7] F. Delarue and S. Menozzi, *An interpolated stochastic algorithm for quasi-linear PDEs*, Math. Comp. **77** (2008), no. 261, 125–158.

[8] H. Doss, *Liens entre équations différentielles stochastiques et ordinaires*, Ann. Inst. H. Poincaré Sect. B (N.S.) **13** (1977), no. 2, 99–125.
[9] G. S. Fishman, *Monte Carlo*, Springer Series in Operations Research, Springer-Verlag, New York, 1996. Concepts, algorithms, and applications.
[10] P. Friz and N. Victoir, *Differential Equations Driven by Gaussian Signals II* (2007), available at `arxiv:0711.0668`. Preprint.
[11] ———, *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications*, Cambridge University Press, 2010.
[12] P. Glasserman, *Monte Carlo methods in financial engineering*, Applications of Mathematics (New York), vol. 53, Springer-Verlag, New York, 2004. Stochastic Modelling and Applied Probability.
[13] G. H. Golub and C. F. Van Loan, *Matrix computations*, 3rd ed., Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, 1996.
[14] S. Graf and H. Luschgy, *Foundations of quantization for probability distributions*, Lecture Notes in Mathematics, vol. 1730, Springer-Verlag, Berlin, 2000.
[15] I. Karatzas and S. E. Shreve, *Brownian motion and stochastic calculus*, 2nd ed., Graduate Texts in Mathematics, vol. 113, Springer-Verlag, New York, 1991.
[16] A. Lejay, *An introduction to rough paths*, Séminaire de probabilités XXXVII, Lecture Notes in Mathematics, vol. 1832, Springer-Verlag, 2003, pp. 1–59.
[17] ———, *Yet another introduction to rough paths*, Séminaire de probabilités XLII, Lecture Notes in Mathematics, vol. 1979, Springer-Verlag, 2009, pp. 1–101.
[18] ———, *On rough differential equations*, Electron. J. Probab. **14** (2009), no. 12, 341–364.
[19] ———, *Global solutions to rough differential equations with unbounded vector fields* (2010). Preprint.
[20] A. Lejay and T. Lyons, *On the importance of the Lévy area for systems controlled by converging stochastic processes. Application to homogenization*, New Trends in Potential Theory, Conference Proceedings, Bucharest, September 2002 and 2003 (D. Bakry, L. Beznea, Gh. Bucur, and M. Röckner, eds.), The Theta Foundation, 2006, pp. 63–84.
[21] R. Lord, R. Koekkoek, and D. van Dijk, *A Comparison of Biased Simulation Schemes for Stochastic Volatility Models*, Quant. Finance **10** (2010), no. 2, 177-194.
[22] H. Luschgy and G. Pagès, *Functional quantization of a class of Brownian diffusions: a constructive approach*, Stochastic Process. Appl. **116** (2006), no. 2, 310–336.
[23] Terry J. Lyons, *Differential equations driven by rough signals*, Rev. Mat. Iberoamericana **14** (1998), no. 2, 215–310.
[24] T. Lyons and Z. Qian, *System control and rough paths*, Oxford Mathematical Monographs, Oxford University Press, Oxford, 2002. Oxford Science Publications.
[25] G. Pagès, H. Pham, and J. Printems, *Optimal quantization methods and applications to numerical problems in finance*, Handbook of computational and numerical methods in finance, Birkhäuser Boston, 2004, pp. 253–297.
[26] G. Pagès and J. Printems, *Web site devoted to vector and functional optimal quantization: `www.quantize.maths-fi.com`*, 2005.
[27] G. Pagès and J. Printems, *Optimal quadratic quantization for numerics: the Gaussian case*, Monte Carlo Methods Appl. **9** (2003), no. 2, 135–165.
[28] ———, *Functional quantization for numerics with an application to option pricing*, Monte Carlo Methods Appl. **11** (2005), no. 4, 407–446.

[29] ———, *Optimal quantization for Finance: from random vectors to stochastic processes* (2008). Preprint, Université Paris 6.
[30] G. Pagès and A. Sellami, *Convergence of multi-dimensional quantized SDE's* (2008), available at `arxiv:0801.0726`.
[31] H.J. Sussmann, *On the gap between deterministic and stochastic ordinary differential equations*, Ann. Probability **6** (1978), no. 1, 19–41.
[32] R. Szechtman, *Control variate techniques for Monte Carlo simulation*, Proceedings of the 2003 Winter Simulation Conference (S. Chick, P.J. Sánchez, D. Ferrin, and D.J. Morrice, eds.), 2003, available at `www.informs-cs.org/wsc03papers/017.pdf`.

A. Lejay, Projet Tosca, Institut Elie Cartan Nancy, Nancy-Université, CNRS, INRIA, Boulevard des Aiguillettes B.P. 239 F-54506 Vandœuvre lès Nancy cedex, France.
E-mail address: `Antoine.Lejay@iecn.u-nancy.fr`
URL: `http://www.iecn.u-nancy.fr/~lejay`

V. Reutenauer, Credit Agricole Corporate and Investment Bank, Interest Rates and Hybrid Quantitative Research, 9 quai du Président Paul Doumer, 92920 Paris-La Défense cedex, France
E-mail address: `victor.reutenauer@ca-cib.fr`

Option	v'/v	v'_κ/v	$\frac{T'_{\mathrm{s},N}}{T_\mathrm{s}}$	$\frac{vT_\mathrm{s}}{v'T'_{N,\mathrm{s}}}$	$\frac{vT_\mathrm{s}}{v'_\kappa T'_{N,\mathrm{s}}}$	$\frac{T'_{M,\mathrm{s}}}{T_\mathrm{s}}$	$\frac{vT_\mathrm{s}}{v'T'_{M,\mathrm{s}}}$	$\frac{vT_\mathrm{s}}{v'_\kappa T'_{\kappa,\mathrm{s}}}$	μ	σ	σ'	σ'_κ
European OTM	0.52	0.5	1.2	1.5	1.6	1.2	1.5	1.6	0.017	$4.2 \cdot 10^{-03}$	$2.2 \cdot 10^{-03}$	$2.1 \cdot 10^{-03}$
European ATM	0.29	0.28	1.2	2.8	2.9	1.2	2.8	2.9	0.11	0.024	$6.8 \cdot 10^{-03}$	$6.6 \cdot 10^{-03}$
European ITM	0.23	0.22	1.2	3.5	3.6	1.2	3.5	3.6	0.55	0.045	0.01	$9.9 \cdot 10^{-03}$
Asian OTM	1	$\star$	1.2	*0.8*	$\star$	1.2	*0.8*	$\star$	$9.3 \cdot 10^{-04}$	$1.0 \cdot 10^{-04}$	$1.0 \cdot 10^{-04}$	$\star$
Asian ATM	0.08	0.08	1.2	**10**	**10**	1.2	**10**	**10**	0.055	$6.4 \cdot 10^{-03}$	$5.1 \cdot 10^{-04}$	$5.1 \cdot 10^{-04}$
Asian ITM	0.056	0.056	1.2	**14**	**14**	1.2	**14**	**14**	0.52	0.014	$7.8 \cdot 10^{-04}$	$7.9 \cdot 10^{-04}$
Asian Binary OTM	1	$\star$	1.2	*0.8*	$\star$	1.2	*0.8*	$\star$	0.016	0.016	0.016	$\star$
Asian Binary ATM	0.31	0.28	1.2	2.6	2.8	1.2	2.6	2.8	0.52	0.25	0.077	0.071
Asian Binary ITM	$\star$	$\star$	1.2	$\star$	$\star$	1.2	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
Ratchet OTM	0.069	0.069	1.2	**12**	**12**	1.2	**12**	**12**	72	$1.8 \cdot 10^{+02}$	12	12
Ratchet ATM	0.07	0.07	1.2	**11**	**11**	1.2	**11**	**11**	56	$1.7 \cdot 10^{+02}$	12	12
Ratchet ITM	0.084	0.081	1.2	**9.6**	**9.9**	1.2	**9.6**	**9.9**	30	$1.4 \cdot 10^{+02}$	12	11
Lookback fixed OTM	0.34	0.33	1.2	2.4	2.4	1.2	2.4	2.4	0.029	$6.6 \cdot 10^{-03}$	$2.3 \cdot 10^{-03}$	$2.2 \cdot 10^{-03}$
Lookback fixed ATM	0.16	0.16	1.2	5	5	1.2	**5**	5	0.19	0.026	$4.2 \cdot 10^{-03}$	$4.2 \cdot 10^{-03}$
Lookback fixed ITM	0.16	0.16	1.2	5	5	1.2	**5**	5	0.69	0.026	$4.2 \cdot 10^{-03}$	$4.2 \cdot 10^{-03}$
Lookback floating ITM	0.42	0.4	1.2	1.9	2	1.2	1.9	2	0.17	0.023	$9.8 \cdot 10^{-03}$	$9.3 \cdot 10^{-03}$
Cliquet OTM	0.24	0.24	1.2	3.3	3.3	1.2	3.3	3.3	0.26	0.019	$4.6 \cdot 10^{-03}$	$4.6 \cdot 10^{-03}$
Cliquet ATM	0.27	0.26	1.2	3	3.1	1.2	3	3.1	0.25	0.015	$4.1 \cdot 10^{-03}$	$4.0 \cdot 10^{-03}$
Cliquet ITM	3.3	0.72	1.2	*0.24*	1.1	1.2	*0.24*	1.1	0.097	$1.6 \cdot 10^{-04}$	$5.2 \cdot 10^{-04}$	$1.1 \cdot 10^{-04}$

TABLE 2. Results for the Black & Scholes model with $X_0 = 1$, $\sigma = 0.2$, $\nu = 0.05$, $T = 1$. The first coefficient of the KL-decomposition is decomposed with 7 points and the 5 next are decomposed with 20 points (140 quantized paths). For the simulations, N = 100,000 and M = 1,000,000. The $\star$ indicates irrelevant results (the quantized price is equal to 0 and the algorithm for computing κ does not converge).

Option	v'/v	v'_κ/v	$\frac{T'_{\mathrm{s},N}}{T_\mathrm{s}}$	$\frac{vT_\mathrm{s}}{v'T'_{N,\mathrm{s}}}$	$\frac{vT_\mathrm{s}}{v'_\kappa T'_{N,\mathrm{s}}}$	$\frac{T'_{M,\mathrm{s}}}{T_\mathrm{s}}$	$\frac{vT_\mathrm{s}}{v'T'_{M,\mathrm{s}}}$	$\frac{vT_\mathrm{s}}{v'_\kappa T'_{\kappa,\mathrm{s}}}$	μ	σ	σ'	σ'_κ
European OTM	0.5	0.48	1.2	1.6	1.7	1.2	1.6	1.7	0.029	$4.7 \cdot 10^{-03}$	$2.4 \cdot 10^{-03}$	$2.3 \cdot 10^{-03}$
European ATM	0.33	0.32	1.2	2.4	2.6	1.2	2.4	2.6	0.21	0.024	$8.0 \cdot 10^{-03}$	$7.6 \cdot 10^{-03}$
European ITM	0.32	0.31	1.2	2.5	2.6	1.2	2.5	2.6	0.7	0.028	$9.1 \cdot 10^{-03}$	$8.7 \cdot 10^{-03}$
Asian OTM	0.94	0.88	1.2	*0.87*	*0.93*	1.2	*0.87*	*0.93*	$9.3 \cdot 10^{-04}$	$6.9 \cdot 10^{-05}$	$6.5 \cdot 10^{-05}$	$6.1 \cdot 10^{-05}$
Asian ATM	0.05	0.05	1.2	**16**	**16**	1.2	**16**	**16**	0.1	$7.4 \cdot 10^{-03}$	$3.7 \cdot 10^{-04}$	$3.7 \cdot 10^{-04}$
Asian ITM	0.05	0.05	1.2	**16**	**16**	1.2	**16**	**16**	0.6	$9.5 \cdot 10^{-03}$	$4.7 \cdot 10^{-04}$	$4.7 \cdot 10^{-04}$
Asian Binary OTM	1.1	0.74	1.2	*0.75*	1.1	1.2	*0.75*	1.1	0.022	0.022	0.024	0.016
Asian Binary ATM	0.21	0.19	1.2	3.8	4.3	1.2	3.8	4.3	0.84	0.14	0.029	0.026
Asian Binary ITM	$\star$	$\star$	1.2	$\star$	$\star$	1.2	$\star$	$\star$	1	0	$\star$	$\star$
Ratchet OTM	0.068	0.068	1.2	**12**	**12**	1.2	**12**	**12**	81	$1.1 \cdot 10^{+02}$	7.4	7.4
Ratchet ATM	0.068	0.068	1.2	**12**	**12**	1.2	**12**	**12**	65	$1.1 \cdot 10^{+02}$	7.3	7.3
Ratchet ITM	0.078	0.077	1.2	**10**	**11**	1.2	**10**	**11**	38	97	7.6	7.4
Lookback fixed OTM	0.31	0.31	1.2	2.6	2.7	1.2	2.6	2.7	0.049	$7.2 \cdot 10^{-03}$	$2.2 \cdot 10^{-03}$	$2.2 \cdot 10^{-03}$
Lookback fixed ATM	0.19	0.18	1.2	4.4	4.5	1.2	4.4	4.5	0.28	0.021	$3.8 \cdot 10^{-03}$	$3.8 \cdot 10^{-03}$
Lookback fixed ITM	0.19	0.18	1.2	4.4	4.5	1.2	4.4	4.5	0.78	0.021	$3.8 \cdot 10^{-03}$	$3.8 \cdot 10^{-03}$
Lookback floating ITM	0.47	0.44	1.2	1.7	1.9	1.2	1.7	1.9	0.26	0.021	$1.0 \cdot 10^{-02}$	$9.2 \cdot 10^{-03}$
Cliquet OTM	0.3	0.3	1.2	2.7	2.7	1.2	2.7	2.7	0.33	0.014	$4.2 \cdot 10^{-03}$	$4.2 \cdot 10^{-03}$
Cliquet ATM	0.35	0.32	1.2	2.4	2.5	1.2	2.4	2.5	0.32	0.011	$3.9 \cdot 10^{-03}$	$3.6 \cdot 10^{-03}$
Cliquet ITM	8.2	0.89	1.2	*0.1*	*0.91*	1.2	*0.1*	*0.91*	0.1	$1.0 \cdot 10^{-05}$	$8.5 \cdot 10^{-05}$	$9.3 \cdot 10^{-06}$

TABLE 3. Results for the CIR model with $X_0 = 1$, $\sigma = 0.2$, $a = 1.5$, $k = 0.5$, $T = 1$. The first coefficient of the KL-decomposition is decomposed with 7 points and the 5 next are decomposed with 20 points (140 quantized paths). For the simulations, $N = 100{,}000$ and $M = 1{,}000{,}000$. The $\star$ indicates irrelevant results (the variance is already equal to 0).

Option	v'/v	v'_κ/v	$\frac{T'_N}{T}$	$\frac{vT}{v'T'_N}$	$\frac{vT}{v'_\kappa T'_N}$	$\frac{T'_M}{T}$	$\frac{vT}{v'T'_M}$	$\frac{vT}{v'_\kappa T'_\kappa}$	μ	σ	σ'	σ'_κ
European OTM	0.3	0.28	1.5	2.2	2.3	1.3	2.6	2.9	0.096	0.086	0.026	0.024
European ATM	0.22	0.2	1.5	3	3.2	1.3	3.6	3.9	0.19	0.14	0.031	0.029
European ITM	0.15	0.15	1.5	4.2	4.4	1.3	**5.2**	**5.4**	0.53	0.24	0.036	0.034
Asian OTM	0.29	0.26	1.5	2.3	2.5	1.3	2.7	3	0.026	0.011	$3.1 \cdot 10^{-03}$	$2.8 \cdot 10^{-03}$
Asian ATM	0.13	0.11	1.5	**5.1**	**5.6**	1.3	**6.2**	**6.8**	0.1	0.035	$4.4 \cdot 10^{-03}$	$4.0 \cdot 10^{-03}$
Asian ITM	0.074	0.071	1.5	**8.7**	**9.1**	1.3	**11**	**11**	0.5	0.07	$5.2 \cdot 10^{-03}$	$5.0 \cdot 10^{-03}$
Asian Binary OTM	0.29	0.26	1.5	2.3	2.5	1.3	2.7	3	0.026	0.011	$3.1 \cdot 10^{-03}$	$2.8 \cdot 10^{-03}$
Asian Binary ATM	0.13	0.11	1.5	**5.1**	**5.7**	1.3	**6.2**	**6.8**	0.1	0.035	$4.4 \cdot 10^{-03}$	$4.0 \cdot 10^{-03}$
Asian Binary ITM	0.074	0.071	1.5	**8.8**	**9.1**	1.3	**11**	**11**	0.5	0.07	$5.2 \cdot 10^{-03}$	$5.0 \cdot 10^{-03}$
Ratchet OTM	0.086	0.083	1.5	**7.5**	**7.8**	1.3	**9.1**	**9.4**	65	$6.8 \cdot 10^{+02}$	59	57
Ratchet ATM	0.094	0.09	1.5	**6.9**	**7.2**	1.3	**8.3**	**8.7**	51	$5.9 \cdot 10^{+02}$	56	53
Ratchet ITM	0.12	0.11	1.5	**5.4**	**5.7**	1.3	**6.5**	**6.9**	30	$3.8 \cdot 10^{+02}$	46	43
Lookback fixed OTM	0.3	0.23	1.5	2.1	2.8	1.3	2.6	3.4	0.19	0.15	0.044	0.034
Lookback fixed ATM	0.21	0.17	1.5	3.1	3.8	1.3	3.7	4.6	0.39	0.2	0.042	0.034
Lookback fixed ITM	0.21	0.17	1.5	3.1	3.8	1.3	3.7	4.6	0.89	0.2	0.042	0.034
Lookback floating ITM	0.25	0.24	1.5	2.6	2.7	1.3	3.1	3.3	0.3	0.15	0.038	0.036
Cliquet OTM	0.26	0.26	1.5	2.5	2.5	1.3	3.1	3.1	0.5	0.1	0.026	0.026
Cliquet ATM	0.64	0.45	1.5	1	1.4	1.3	1.2	1.8	0.37	0.021	0.013	$9.5 \cdot 10^{-03}$
Cliquet ITM	7.7	0.86	1.5	*0.084*	*0.75*	1.3	*0.1*	*0.91*	0.098	$1.2 \cdot 10^{-04}$	$9.5 \cdot 10^{-04}$	$1.1 \cdot 10^{-04}$

TABLE 4. Results for the Heston model with $V_0 = 0.2$, $X_0 = 1$, $\sigma = 0.2$, $k = 0.5$, $a = 0.2$, $\rho = 0.3$, $\nu = 0.02$, $T = 1$. Both for the underlying Brownian motion of the volatility and the price, the first coefficient of the KL-decomposition is quantized with 7 points, while the 4 next coefficients are quantized with 20 points (19,600 quantized paths). For the simulations, $N = 100{,}000$ and $M = 1{,}000{,}000$.