Functional quantization-based stratified sampling methods

Sylvain CORLAY* †   Gilles PAGÈS †

September 30, 2014

Abstract

In this article, we propose several quantization-based stratified sampling methods to reduce the variance of a Monte Carlo simulation.

Theoretical aspects of stratification lead to a strong link between optimal quadratic quantization and the variance reduction that can be achieved with stratified sampling. We first put the emphasis on the consistency of quantization for partitioning the state space in stratified sampling methods in both finite and infinite dimensional cases. We show that the proposed quantization-based strata design has uniform efficiency among the class of Lipschitz continuous functionals.

Then a stratified sampling algorithm based on product functional quantization is proposed for path-dependent functionals of multi-factor diffusions. The method is also available for other Gaussian processes such as Brownian bridge or Ornstein-Uhlenbeck processes. We derive in detail the case of Ornstein-Uhlenbeck processes.

We also study the balance between the algorithmic complexity of the simulation and the variance reduction factor.

Keywords: functional quantization, vector quantization, stratification, variance reduction, Monte Carlo simulation, Karhunen-Loève, Gaussian process, Brownian motion, Brownian bridge, Ornstein-Uhlenbeck process, Ornstein-Uhlenbeck bridge, principal component analysis, numerical integration, option pricing, Voronoi diagram, product quantizer, path-dependent option.

Introduction

The quantization of a random variable X consists of its approximation by a random variable Y taking finitely many values. This problem has been initially investigated for its applications to signal transmission and for compression issues [8]. In this context, quantization is a method of signal discretization. The aim is to choose the random variable Y so as to minimize the resulting error for a fixed quantization level N.

More recently, quantization was introduced in numerical probability to devise numerical integration methods [24] and to solve multidimensional stochastic control problems such as the pricing of American options [1] and swing options [2]. Optimal quantization has many other applications and extensions in various fields such as automatic clustering (quantization of empirical measures) and pattern recognition.

Since the early 2000's, the infinite-dimensional setting has been extensively investigated from both theoretical and numerical viewpoints with a special attention paid to functional quantization [18, 25]. Bi-measurable stochastic processes are viewed as random variables valued in functional spaces.

Still Monte Carlo simulations remain the most common numerical method in the field of numerical probability. One reason is that it is easy to implement in an industrial configuration. In the industry of derivatives, banks implement generic Monte Carlo frameworks for pricing and hedging their positions with a wide variety of financial products and models. Besides, Monte Carlo simulations are easily parallelized.

---

*Bloomberg L.P. Quantitative Research, 731 Lexington avenue, New York, NY 10022, USA. scorlay@bloomberg.net.

†Laboratoire de Probabilités et Modèles Aléatoires, UMR 7599, Université Paris 6, case 188, 4, pl. Jussieu, F-75252 Paris Cedex 5, France. gilles.pages@upmc.fr

Variance reduction methods can be used to dramatically reduce the computation time of a Monte Carlo simulation, or to increase its accuracy. The main variance reduction methods are (adaptive) control variate, pre-conditioning, importance sampling and stratification [9, 17]. The problem is that these methods may strongly depend on the payoff or the model and require significant changes in the practical implementation of the Monte Carlo simulation. Therefore, most practitioners do not use the most sophisticated methods except for marginal cases.

In this article, we point out theoretical aspects of quantization that idraw a strong link between the problem of optimal quadratic quantization of a random variable and the variance reduction that can be achieved by stratification. We emphasize the consistency of quantization for designing strata in stratified sampling methods in both the finite and infinite dimensional settings. Then we devise a stratified sampling algorithm based on product functional quantization for path-dependent functionals of multi-factor Brownian diffusions. We show that this strata design has uniform efficiency among the class of Lipschitz continuous functionals of Brownian motion. The simulation cost of the conditional path is $O(n)$ where n is the number of discretization dates as in the naive unconditioned Monte Carlo simulations. In this context, the proposed approach can be considered as a guided Monte Carlo simulation (see Figure 5). The method is applicable with any Gaussian process as soon as we can derive its Karhunen-Loève expansion. This is the case for Brownian bridge and Ornstein-Uhlenbeck processes. The special case of Ornstein-Uhlenbeck processes is detailed in Appendix A. The case of the Ornstein-Uhlenbeck bridge is presented in [6].

A very common situation is the case of Monte Carlo simulations of multi-factor Brownian diffusions approximated with an Euler discretization or another time-discretization scheme. The presented method is particularly well suited for this case, regardless of how the Brownian paths are used in the model, to drive the dynamics of the stock price, a volatility process or a drift term. Functional stratification can be used as a generic variance reduction method which does not require a reimplementation of the whole framework but only the way it is input with Brownian paths.

The article is organized as follows. Section 1 presents some necessary background on optimal quantization. The emphasis is on the functional quantization of Gaussian processes. Section 2 briefly covers the first functional quantization-based variance reduction method that was proposed in [25, 16]. Section 3 outlines the links between quantization and stratification with an emphasis on the Gaussian case. The method is further detailed in the infinite-dimensional case for Gaussian processes in Section 4. We present a simulation method for the case of Brownian motion and other examples of Gaussian processes (such as Brownian bridge and Ornstein-Uhlenbeck processes) that preserves the $O(n)$ simulation complexity where n is the number of time steps. In Section 5, we provide numerical experiments of the method with option pricing problems arising in mathematical finance. Appendix A presents the computation of the Karhunen-Loève expansion of Ornstein-Uhlenbeck processes. Appendix B presents the derivation of closed-form expressions of some regression matrices needed for our stratified sampling algorithm, in the cases of Brownian motion, Brownian bridge and Ornstein-Uhlenbeck processes.

1 Vector and functional quantization

1.1 Introduction to quantization of random variables

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $(E, |\cdot|)$ a reflexive separable Banach space. The principle of the quantization of a random variable X taking its values in E is to approximate X by a random variable Y taking a finite number N of values in E. The discrete random variable Y is a quantizer of X of level N. The resulting discretization error to be minimized is the L^p norm of $|X - Y|$.

$$\min \left\{ \|X - Y\|_p, Y : \Omega \to E \text{ measurable, } \operatorname{card}(Y(\Omega)) \leq N \right\}. \tag{1}$$

Definition 1 (Voronoi partition). *Consider $N \in \mathbb{N}^*$, $\Gamma = \{\gamma_1, \ldots, \gamma_N\} \subset E$ and let $C = \{C_1, \ldots, C_N\}$ be a Borel partition of E. C is a Voronoi partition associated with Γ if $\forall i \in \{1, \ldots, N\}$, $C_i \subset \{\xi \in E, |\xi - \gamma_i| = \min_{j \in \{1, \ldots, N\}} |\xi - \gamma_j|\}$. C_i is called Voronoi cell associated with γ_i in C.*

.

Proposition 1.1. *Let X and Y be two random variables valued in E, where Y takes its values in the fixed set of knots $\Gamma = \{\gamma_1, \ldots, \gamma_N\} \subset E$ for $N \in \mathbb{N}^*$. We define $\widehat{X}^\Gamma := \mathrm{Proj}_\Gamma(X)$ where $\mathrm{Proj}_\Gamma = \sum_{i=1}^{N} \gamma_i \mathbf{1}_{C_i}$ is a nearest neighbor projection onto Γ. Then we have $\left| X - \widehat{X}^\Gamma \right| \leq |X - Y|$ a.s. and thus $\left\| X - \widehat{X}^\Gamma \right\|_p \leq \|X - Y\|_p$.*

A consequence is that solving (1) amounts to solving the simpler problem

$$\min \left\{ \|X - \mathrm{Proj}_\Gamma(X)\|_p, \ \Gamma \subset E, \ \mathrm{card}(\Gamma) \leq N \right\}.$$

The quantity $\|X - \mathrm{Proj}_\Gamma(X)\|_p$ is called the mean L^p quantization error. The problem of the existence of a minimum is addressed in [22, 10] for the finite-dimensional case.

- For every $N \geq 1$, the mean L^p quantization error is Lipschitz continuous and reaches a minimum. An N-tuple that achieves the minimum has pairwise distinct components, as soon as $\mathrm{card}(\mathrm{supp}(\mathbb{P}_X)) \geq N$.

This result stands in the general case of a random variable valued in a reflexive separable Banach space [18].

- If the support of $\mathbb{P}_X$ has infinite cardinal, the optimal quantization error decreases, and converges to 0 as the quantization level N goes to infinity. In the finite-dimensional case, and for distributions that are absolutely continuous with respect to the Lebesgue measure, the rate of convergence is ruled by Theorem 1.2.

Theorem 1.2 (Zador, Luschgy, Pagès).   • (Sharp rate) *Let $r > 0$ and $X : \Omega \to \mathbb{R}^d \in L^{r+\eta}(\mathbb{P})$ for some $\eta > 0$. Let $\mathbb{P}_X(d\xi) = \phi(\xi)d\xi + \mu(d\xi)$ be the canonical decomposition of $\mathbb{P}_X$ (μ and the Lebesgue measure are singular). Then, if $\phi \not\equiv 0$, the L^r quantization error at level N, $\mathcal{E}_{N,r}$ satisfies*

$$\mathcal{E}_{N,r}(X, \mathbb{R}^d) \underset{N \to \infty}{\sim} \widetilde{J}_{r,d} \times \left(\int_{\mathbb{R}^d} \phi^{\frac{d}{d+r}}(u) du \right)^{\frac{1}{d} + \frac{1}{r}} \times N^{-\frac{1}{d}}, \quad \text{where} \quad \widetilde{J}_{r,d} \in (0, \infty). \tag{2}$$

- (Non-asymptotic upper bound) *Let $d \geq 1$. There exists $C_{d,r,\eta} \in (0, \infty)$ such that, for every $\mathbb{R}^d$-valued random vector X,*

$$\forall N \geq 1, \quad \mathcal{E}_{N,r}(X, \mathbb{R}^d) \leq C_{d,r,\eta} \|X\|_{r+\eta} N^{-\frac{1}{d}}. \tag{3}$$

The first claim was stated for the case of distributions with compact support by Zador in [28]. The extension to general probability distributions in $\mathbb{R}^d$ was developed in [4]. The first mathematically rigorous proof can be found in [10]. The non-asymptotic error bound (3) is proved in [21].

In Figure 1, we display the Voronoi partition of a random N-quantizer and an optimal quadratic quantizer of level N for the bivariate normal distribution $\mathcal{N}(0, I_2)$.

1.2   Self-consistency of optimal quantizers

We now assume that E is a separable Hilbert space $(H, \langle \cdot, \cdot \rangle_H)$. We denote by $\mathcal{C}_N(X)$ the set of L^2 optimal quantizers of X of level N, and by $\mathcal{E}_N(X)$ the minimal quadratic distortion that can be achieved when approximating X by a quantizer of level N.

Definition 2 (Stationarity). *A quantizer Y of X is stationary (or self-consistent) if*

$$Y = \mathbb{E}[X|Y]. \tag{4}$$

Proposition 1.3 (Stationarity of L^2 optimal quantizers). *A quadratic optimal quantizer is stationary.*

We refer to [10] for the proof in the finite-dimensional setting and to [18] for the more general case of separable Hilbert spaces. Stationarity is a particularity of the quadratic case ($p = 2$). In the general L^p setting, a similar property involving the notion of p-center holds [11].
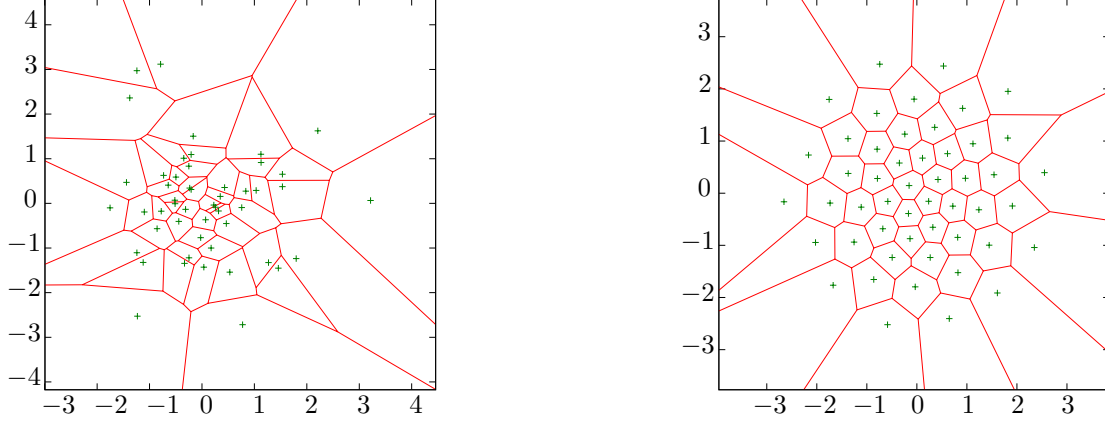
3

Figure 1: Voronoi partition of a random quantizer (left) and an optimal quantizer (right) of level $N = 48$ of the $\mathcal{N}(0, I_2)$ distribution.

Proposition 1.4. *Let X be an H-valued L^2 random variable. Let us denote by D_N^X the squared quadratic quantization error associated with a codebook of size N with respect to X.*

$$D_N^X : \quad \begin{matrix} H^N & \to & \mathbb{R}_+ \\ \Gamma = (\gamma_1, \ldots, \gamma_N) & \mapsto & \mathbb{E}\big[\min_{1 \le i \le N} |X - \gamma_i|_H^2 \big]. \end{matrix}$$

The distortion D_N^X is $|\cdot|_H$-differentiable at N-quantizers $\Gamma \in H^N$ with pairwise distinct components and such that boundaries of Voronoi cells are $\mathbb{P}_X$-negligible

$$\nabla D_N^X(\Gamma) = 2 \Big(\int_{C_i(\Gamma)} (\gamma_i - \xi) \mathbb{P}_X(d\xi) \Big)_{1 \le i \le N} = 2 \Big(\mathbb{E}\big[\big(\widehat{X}^\Gamma - X \big) \mathbf{1}_{\{\widehat{X}^\Gamma = \gamma_i\}} \big] \Big)_{1 \le i \le N}. \tag{5}$$

Hence any Voronoi quantizer associated with a critical point of D_N^X is a stationary quantizer.

We refer to [23] for a detailed proof.

Definition 3 (Centroidal projection). *Let $C = \{C_1, \ldots, C_N\}$ be a Borel partition of H. For $1 \le i \le N$, we define $G_i := \begin{cases} \mathbb{E}[X | X \in C_i] & \text{if } \mathbb{P}[X \in C_i] \ne 0, \\ 0 & \text{otherwise}, \end{cases}$ the centroids associated with X and C.*

The centroidal projection associated with C and X is the application $\text{Proj}_{C,X} : x \mapsto \sum_{i=1}^N G_i \mathbf{1}_{C_i}(x)$.

1.3 Optimal quantization and principal component analysis

For any finite-dimensional subspace U of H, we denote by Π_U the orthogonal projection onto U.

Proposition 1.5. *Let U be a finite-dimensional linear subspace of H. Then*

$$\mathcal{E}_N(X)^2 \le \mathbb{E}\Big[|X - \Pi_U(X)|^2 \Big] + \mathcal{E}_N(\Pi_U(X))^2. \tag{6}$$

Moreover, if an optimal quantizer of X of level N lies in U, we have equality in (6).

We refer to [18] for a detailed proof. This allows us to define the quantization dimension of X of level N by $d_N(X) := \min\big\{ \dim \text{span}(\Gamma), \Gamma \in \mathcal{C}_N(X) \big\}$. It follows from Proposition 1.5 that

$$\mathcal{E}_N^2(X) = \min \left\{ \mathbb{E}[\|X - \Pi_V(X)\|^2] + \mathcal{E}_N^2(\Pi_V(X)), \begin{array}{l} V \subset H \text{ linear subspace} \\ \text{such that } \dim V \ge d_N(X) \end{array} \right\}.$$

4

1.3.1 Covariance operator of a Gaussian measure

Definition 4. *Let X be a centered H-valued L^2 Gaussian random variable. Its covariance operator $C_X : H \to H$ is defined by $C_X y = \mathbb{E}[\langle y, X \rangle X]$.*

1. *If X is $\mathbb{R}^d$-valued, the matrix of C_X in the canonical basis is the covariance matrix of X.*

2. *If $X = (X_t)_{t \in [0,T]}$ is a bi-measurable centered process of covariance function $\Gamma_X(s,t) := \mathbb{E}[X_s X_t]$ satisfying $\int_{[0,T]} \Gamma_X(s,s)ds < +\infty$, then X can be seen as a random variable valued in $L^2([0,T],dt)$ satisfying $\mathbb{E}\left[|X|^2\right] < \infty$, and*

$$C_X y = \int_{[0,T]} y(s)\Gamma_X(s,\cdot)ds, \quad y \in L^2([0,T],dt).$$

In [18], it is proved that linear subspaces U of H spanned by n-stationary quantizers of Gaussian measures correspond to principal subspaces of X. In other words, they are spanned by the eigenvectors of C_X corresponding to the largest eigenvalues.

Theorem 1.6. *Let Γ be an optimal codebook for the Gaussian random variable X, $U = \mathrm{span}(\Gamma)$ and $m = \dim U$. Then $C_X(U) = U$ and $\mathbb{E}\left[|X - \Pi_U(X)|^2\right] = \sum_{j \geq m+1} \lambda_j^X$, where $\lambda_1^X \geq \lambda_2^X \geq \cdots > 0$ are the ordered non-zero eigenvalues of C_X (repeated as many times as their multiplicity). We have*

$$\sum_{j \geq m+1} \lambda_j^X = \inf\left\{\mathbb{E}\left[|X - \Pi_V(X)|^2\right], V \subset H \text{ linear subspace, } \dim V = m\right\}.$$

The minimal quadratic distortion $\mathcal{E}_N(X)$ is given by

$$\mathcal{E}_N(X)^2 = \sum_{j \geq m+1} \lambda_j^X + \mathcal{E}_N\left(\bigotimes_{j=1}^m \mathcal{N}\left(0, \lambda_j^X\right)\right)^2 \quad \text{for } m \geq d_N(X), \tag{7}$$

A proof is available in [18]. This shows that the optimal quantization of a Gaussian process X boils down to a finite-dimensional quantization problem, if the Karhunen-Loève eigensystem $(e_n^X, \lambda_n^X)_{n \in \mathbb{N}^*}$ is known.

1.4 Product quantization

Let $(e_n)_{n \in \mathbb{N}^*}$ be a Hilbert basis of H, and $(N_n)_{n \geq 1}$ an integer sequence such that $\prod_{n \geq 1} N_n < \infty$ (so that $N_n = 1$ for large enough n). For every $n \in \mathbb{N}^*$, we consider a codebook of size N_n, $\Gamma^n := \left\{\gamma_1^n, \ldots, \gamma_{N_n}^n\right\} \subset \mathbb{R}$.

The codebook Γ is defined as the set of knots in H whose coordinates in the base $(e_n)_{n \in \mathbb{N}^*}$ are the Cartesian product of the one-dimensional codebooks Γ^n.

Proposition 1.7 (Case of independent marginals). *With the same notation, if we assume that the marginals of X, $(\langle X, e_1 \rangle, \langle X, e_2 \rangle, \ldots)$ are independent, and that for each $k \in \mathbb{N}^*$, $Y^k := \mathrm{Proj}_{\Gamma^k}(\langle X, e_k \rangle)$ is a stationary quantizer of $\langle X, e_k \rangle$, then $Y = \mathrm{Proj}_\Gamma(X)$ is a stationary quantizer of X.*

In the case of independent marginals, optimal product quantization remains stationary and the simple shape of Voronoi cells simplifies the nearest neighbor search.

1.5 Numerical optimal quantization

Various algorithms have been developed to compute optimal N-grids in the finite-dimensional setting. A review of these methods is available in [23]. Let us mention Lloyd's algorithm for the quadratic case. Another approach is the stochastic gradient method which is suggested by the fact that the quadratic distortion function has an integral representation and is differentiable at any N-tuple having pairwise distinct components and a $\mathbb{P}_X$-negligible Voronoi tessellation boundary [24].

Equation (5) shows that any Voronoi quantizer associated with a critical point of D_N^X is a stationary quantizer. In the case of one-dimensional distributions, such as the Gaussian distribution, the (tridiagonal) Hessian of the distortion has a closed-form expression. Hence, a Newton-Raphson method can be easily implemented. It is thoroughly studied in [24] in the Gaussian case and remains the fastest way to compute L^2 optimal quantizers of one-dimensional Gaussian variables.

1.6 Quantization of Gaussian processes

1.6.1 Optimal quantization

From now on, we will assume that X is a bi-measurable Gaussian process and has a continuous covariance function Γ^X and satisfies $\mathbb{E}\left[|X|_{L_T^2}^2\right] = \int_0^T \mathbb{E}[X_s^2]ds < \infty$.

We have seen in Section 1.3 that in this setting, the L^2 optimal quantization X amounts to the quantization of a finite-dimensional Gaussian vector $\bigotimes_{j=1}^m \mathcal{N}\left(0, \lambda_j^X\right)$ for some positive integer m, the quantization dimension.

Several usual Gaussian processes have explicit Karhunen-Loève expansions, such as Brownian motion, Brownian bridge and Ornstein-Uhlenbeck processes and bridges. (The case of a stationary Ornstein-Uhlenbeck process is derived for normalized parameters in the stationary case in [12, p.195].) In Appendix A, we derive the Karhunen-Loève expansion of the Ornstein-Uhlenbeck process in the general case (for any value of the parameters and the initial variance). The K-L expansion of the Ornstein-Uhlenbeck bridge is derived in [6]. To the best of our knowedge, no closed-form expression is available for fractional Brownian motion. In the article, numerical examples will be presented for the following cases.

1. **Brownian motion on** $[0,T]$:

$$e_n^W(t) := \sqrt{\frac{2}{T}} \sin\left(\pi(n-1/2)\frac{t}{T}\right), \qquad \lambda_n^W := \left(\frac{T}{\pi(n-1/2)}\right)^2, \quad n \geq 1. \tag{8}$$

2. **Brownian bridge on** $[0,T]$:

$$e_n^B(t) := \sqrt{\frac{2}{T}} \sin\left(\pi n\frac{t}{T}\right), \qquad \lambda_n^B := \left(\frac{T}{\pi n}\right)^2, \quad n \geq 1. \tag{9}$$

3. **The Ornstein-Uhlenbeck process on** $[0,T]$, **starting from** 0, and defined by the SDE

$$dX_t = -\theta X_t dt + \sigma dW_t, \tag{10}$$

with $\sigma \geq 0$, $\theta > 0$ and W a standard Brownian motion on $[0,T]$:

$$e_n^{OU}(t) := \frac{1}{\sqrt{\frac{T}{2} - \frac{\sin(2\omega_n T)}{4\omega_n}}} \sin(\omega_n t), \qquad \lambda_n^{OU} := \frac{\sigma^2}{\omega_n^2 + \theta^2}, \quad n \geq 1, \tag{11}$$

where $(\omega_n)_{n\geq 1}$ are the increasingly sorted positive solutions of $\theta \sin(\omega_n T) + \omega_n \cos(\omega_n T) = 0$ (see Appendix A).

4. **The stationary Ornstein-Uhlenbeck process on** $[0,T]$ (see Appendix A).

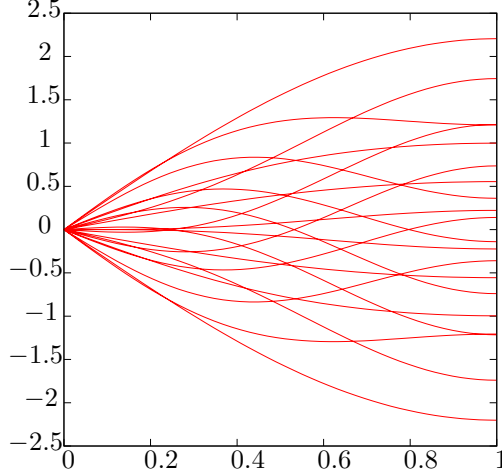In Figure 2, we display a quadratic N-optimal quantizer of Brownian motion.

6

Figure 2: Optimal quantization of Brownian motion on $[0,1]$.

1.6.2 Product quantization

Thanks to Equation (7), the product quantization of the finite-dimensional distribution $\xi \overset{\mathcal{L}}{\sim} \overset{m}{\underset{j=1}{\bigotimes}} \mathcal{N}\left(0, \lambda_j^X\right)$ yields a stationary quantizer $\widehat{X}$ of X of the form $\widehat{X} = \sum_{n \geq 1} \sqrt{\lambda_n^X} \widehat{\xi}_n e_n^X$, where $\widehat{\xi}_n$ is an optimal N_n-quantizer of ξ_n and $\prod_{n \geq 1} N_n \leq N$, $N_n \geq 1$ (so that for large enough n, $N_n = 1$ and $\widehat{\xi}_n \equiv 0$.) The paths corresponding to a multi-index $\underline{i} = \{i_1, \ldots, i_n, \ldots\}$ are of the form $\chi_{\underline{i}} = \sum_{n \geq 1} \sqrt{\lambda_n^X} \gamma_{i_n}^{(N_n)} e_n^X$.

Such a functional quantizer $\widehat{X}$ is called a K-L product quantizer. Furthermore, we denote by $\mathcal{O}_{pq}(X, N)$ the set of K-L product quantizers of size at most N of X. In the case of product quantization, the counterpart of Equation (7) is

$$\mathbb{E}\left[\min_{\underline{i}} \left|X - \chi_{\underline{i}}\right|^2\right] = \sum_{n=1}^{m} \lambda_n^X \mathbb{E}\left[\min_{1 \leq i_n \leq N_n} \left|\xi_n - \gamma_{i_n}^{(N_n)}\right|^2\right] + \sum_{n \geq m+1} \lambda_n^X$$
$$= \sum_{n=1}^{m} \lambda_n^X \left(\mathbb{E}\left[\min_{1 \leq i_n \leq N_n} \left|\xi_n - \gamma_{i_n}^{(N_n)}\right|^2\right] - 1\right) + \mathbb{E}\left[|X|_{L_T^2}^2\right], \quad (12)$$

where m is the quantization dimension.

1.6.3 Product decomposition and blind optimization

The minimal quadratic error for a K-L product quantizer of level N is the solution of the minimization problem

$$\mathcal{E}_N^{pq} := \min\left\{\mathcal{E}(\chi), \ \chi \in \mathcal{O}_{pq}(X, N)\right\}, \quad (13)$$

where $\mathcal{E}(\chi)$ is the quadratic distortion of the product quantizer χ. Thanks to (12), this comes to

$$\min\left\{\sum_{n=1}^{d} \lambda_n^X \mathcal{E}_{N_n}\left(\mathcal{N}(0,1)\right)^2 + \sum_{n \geq d+1} \lambda_n^X, \ N_1 \times \cdots \times N_d \leq N, \ d \geq 1\right\}. \quad (14)$$

A solution of (13) is called an optimal K-L product quantizer.

7

The blind optimization procedure consists of computing the criterion for every possible decomposition $N_1 \times \cdots \times N_d \leq N$, $d \geq 1$ and $N_1 \geq N_2 \geq \cdots$. For a given Gaussian process X, results can be kept off-line for a future use. The method is more thoroughly described in [25]. Optimal decompositions for a wide range of values of N for both Brownian bridge and Brownian motion are available on the web site `www.quantize.maths-fi.com` [26] for download. In the case of Ornstein-Uhlenbeck processes, the optimal decomposition depends on the diffusion parameters (σ and θ in (10)) and the maturity.

Some optimal decompositions for the stationary Ornstein-Uhlenbeck process are given in Table 1.

N	N_{rec}	Squared L^2 quantization Error	Product decomposition
1	1	1.5	1
10	10	0.65318	5 × 2
100	96	0.40929	6 × 4 × 2 × 2
1000	960	0.29618	10 × 6 × 4 × 2 × 2
10000	9984	0.23150	13 × 8 × 4 × 3 × 2 × 2 × 2

Table 1: Record of optimal product decompositions of the stationary centered Ornstein-Uhlenbeck process solution of the SDE $dX_t = -X_t dt + dW_t$ on $[0,3]$.

In the following, we will face similar cases (other criteria than the quadratic distortion) where the blind optimization procedure applies.

In Figure 3, we display optimal product quantizers of Brownian motion and Brownian bridge on $[0,1]$. In Figure 4, we display optimal product quantizers of the centered Ornstein-Uhlenbeck process starting from $X_0 = 0$ and a stationary Ornstein-Uhlenbeck on $[0,3]$.
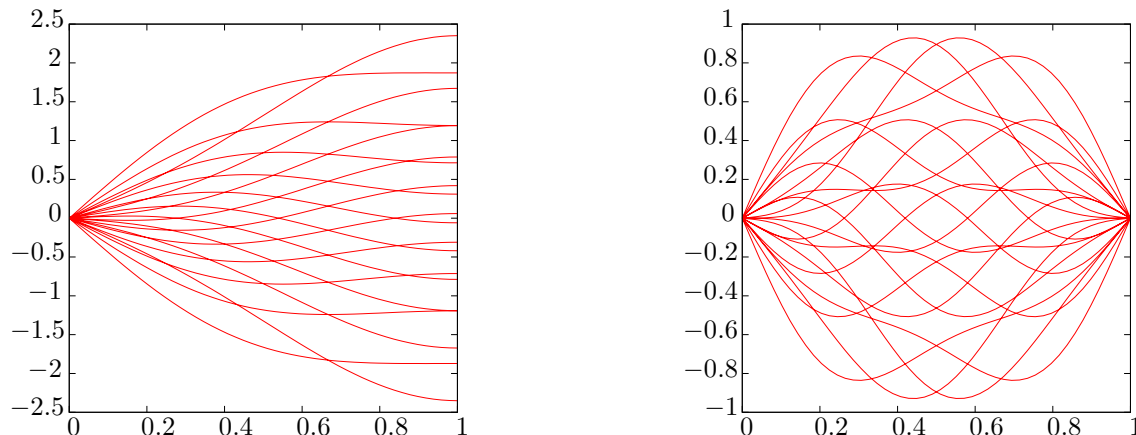

Figure 3: Optimal product quantization of Brownian motion (left) and Brownian bridge (right) on $[0,1]$.

1.6.4 Rate of decay of the quantization error

The rate of decay of the quadratic functional quantization error of Gaussian processes was first investigated in [18] and more precise results were then established in [19]. These results rely on assumptions on the asymptotic behavior of the Karhunen-Loève eigenvalues of the considered process.

Let X be a bi-measurable centered Gaussian process on $[0,T]$ of continuous covariance function Γ^X and such that $\int_0^T \mathbb{E}[X_s^2]ds < \infty$. Its Karhunen-Loève eigensystem is denoted by $\left(e_n^X, \lambda_n^X\right)_{n \geq 1}$.

Theorem 1.8 (Quadratic quantization error asymptotics). *Assume that $\lambda_n^X \sim \phi(n)$ as $n \to \infty$, where $\phi : (s, \infty) \to (0, \infty)$ is a decreasing function such that $\lim_{x \to \infty} \frac{\phi(tx)}{\phi(x)} = t^{-b}$ for $b > 1$ and $s > 0$. Set $\psi(x) := \frac{1}{x\phi(x)}$.*
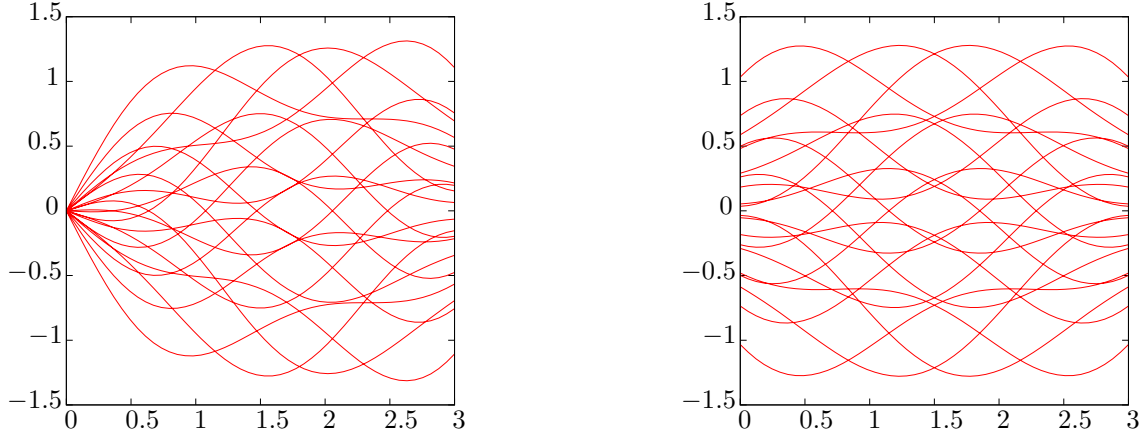
8

Figure 4: Optimal product quantization of a centered Ornstein-Uhlenbeck process, starting from $X_0 = 0$ (left) and stationary (right) solution of the SDE $dX_t = -X_t dt + dW_t$, on $[0, 3]$.

Then

$$\mathcal{E}_N(X) \sim \left(\left(\frac{b}{2} \right)^{b-1} \frac{b}{b-1} \right)^{1/2} \psi(\log(N))^{-1/2} \quad \text{as } N \to \infty.$$

Moreover, the optimal product quantization dimension $m^X(N)$ verifies $m^X(N) \sim \frac{2}{b} \log(N)$ as $N \to \infty$, and the optimal product quantization error $\mathcal{E}_N^{pq}(X)$ of level N satisfies

$$\mathcal{E}_N^{pq}(X) \lesssim \left(\left(\frac{b}{2} \right)^{b-1} \frac{b}{b-1} + C(1) \right)^{1/2} \psi(\log(N))^{-1/2} \quad \text{as } N \to \infty,$$

where $C(1)$ is a universal positive constant.

A proof is available in [19]. Despite of the fact that optimal product quantization is not asymptotically optimal, it provides a rate-optimal sequences of quantizers. Typical rates are $\underset{N \to \infty}{\sim} \log(N)^{-\alpha}$ for $\alpha > 0$. For Brownian motion, Brownian bridge and Ornstein-Uhlenbeck processes, we have $\alpha = \frac{1}{2}$.

2 A first attempt to quantization-based variance reduction: quantization as a control variate

This method has been originally proposed in [25]. Let X be an E-value L^2 random variable, consider $N \in \mathbb{N}^*$ and let $\Gamma = \{y_1, \ldots, y_N\}$ be an N-codebook. We define a quantizer Y of E by $Y := \text{Proj}(X) = \sum_{i=1}^{N} y_i \mathbf{1}_{C_i}(X)$ where $C = \{C_1, \ldots, C_N\}$ is a partition of E. At this stage, we do not need Proj to be a nearest neighbor projection onto Γ.

Let $F : E \to E$ be a Lipschitz continuous function. In order to compute $\mathbb{E}[F(X)]$, we use that:

$$
\begin{aligned}
\mathbb{E}[F(X)] \quad &= \mathbb{E}\left[F(\text{Proj}(X))\right] + \mathbb{E}\left[F(X) - F(\text{Proj}(X))\right] \\
&= \underbrace{\mathbb{E}\left[F(\text{Proj}(X))\right]}_{(a)} + \underbrace{\frac{1}{M} \sum_{m=1}^{M} F\left(X^{(m)}\right) - F\left(\text{Proj}\left(X^{(m)}\right)\right)}_{(b)} + R_{N,M},
\end{aligned}
\tag{15}
$$

where $X^{(m)}, 1 \leq m \leq M$ are M independent copies of X, and $R_{N,M}$ is a remainder term defined by Equation (15). Term (a) is computed by quantization-based cubature and Term (b) is computed by a Monte Carlo

9

simulation. We have

$$\|R_{N,M}\|_2 = \frac{\sigma(F(X) - F(\text{Proj}(X)))}{\sqrt{M}} \leq \frac{\|F(X) - F(\text{Proj}(X))\|_2}{\sqrt{M}} \leq [F]_{\text{Lip}} \frac{\|X - \text{Proj}(X)\|_2}{\sqrt{M}}.$$

Furthermore, $\sqrt{M} R_{N,M} \overset{\mathcal{L}}{\to} \mathcal{N}\left(0, \text{Var}\left(F(X) - F(\text{Proj}(X))\right)\right)$.

Consequently, in the d-dimensional case, if F is Lipschitz continuous and $\left(\widehat{X}^N\right)_{N \in \mathbb{N}} = (\text{Proj}^N(X))_{N \in \mathbb{N}}$ is a rate-optimal sequence of quantizers of X, then we have $\left\|F(X) - F\left(\text{Proj}^N(X)\right)\right\|_2 \leq [F]_{\text{Lip}} \frac{C_X}{N^{1/d}}$ so that

$$\|R_{N,M}\|_2 \leq [F]_{\text{Lip}} \frac{C_X}{M^{1/2} N^{1/d}}.$$

Likewise, in the case of Brownian motion, if $\left(\widehat{W}^N\right)_{N \geq 1}$ is a rate-optimal sequence of quadratic K-L product quantizers of Brownian motion, if F is a Lipschitz continuous functional, then $\left\|F(W) - F\left(\widehat{W}^N\right)\right\|_2 \leq [F]_{\text{Lip}} \frac{C_W}{\log(N)^{1/2}}$ so that

$$\|R_{N,M}\|_2 \leq [F]_{\text{Lip}} \frac{C_W}{M \log(N)^{1/2}}.$$

The bottleneck of fast nearest neighbor search

• **The complexity of the projection:** When implementing the quantization-based control variate variable method (15) , for every draw of the Monte Carlo simulation, one has to compute the projection $\text{Proj}(X^{(m)})$. As a consequence, the efficiency of the method is conditioned by the efficiency of the projection procedure. When dealing with Voronoi quantization, this is simply the nearest neighbor projection.

The problem of nearest neighbor projection, also known as the post-office problem [15], has been widely investigated in the area of computational geometry. It has been solved near optimally in the low dimensional case. Algorithms differ on their practical efficiency on real data sets. For large dimensions, most solutions have a complexity that is exponential with the dimension, or require a longer query time than the obvious brute force algorithm. In fact for dimension $d > \log N$, a brute force algorithm is usually the best choice. Still, even in low dimension, fast nearest neighbor search is a critical part of the algorithm. Let us mention [5] for a fast nearest neighbor search algorithm based on recursive vector quantization.

The speed of the projection can also be increased by relaxing the hypothesis that the projection onto the quantizer is a nearest neighbor projection or by choosing simpler partitions of the state space.

• **The functional case:** The problem of nearest neighbor search is even less tractable in the functional case, as one does not simulate the whole trajectory of the stochastic process but only its marginals at discrete dates, and therefore we can only make an assumption on the interpolation to compute the nearest neighbor.

In [16], a functional quantizer of Brownian motion is used as a control variate variable.

3 Application of quantization to stratification

3.1 Some background on stratified sampling

The main idea of stratification is to localize the Monte Carlo simulation on the elements of a measurable partition of the state space of an L^2 random variable $X : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$. Let $(A_i)_{i \in I}$ be a finite $\mathcal{E}$-measurable partition of E. The sets A_i are called *strata*. We assume that the weights $p_i = \mathbb{P}(X \in A_i)$, $i \in I$ are positive. We will make two pseudo or operating assumptions on these strata:

• $\forall i \in I$, $p_i = \mathbb{P}(X \in A_i)$ is known.

• $\forall i \in I$, the random variable $X_i \overset{\mathcal{L}}{\sim} \mathcal{L}(X | X \in A_i)$ can be simulated at a reasonable cost (say similar to that of X itself).

Tractability of simulation is a major constraint for practical implementation and it has a strong impact on the design of the strata. In practice, we can formulate the condition by assuming that $X_i = \phi_i(U)$ where U is uniformly distributed on $[0,1]^{r_i}$ and $\phi_i : [0,1]^{r_i} \to \mathbb{R}$ is an easily computable function. (We have $r_i \in \mathbb{N} \cup \{+\infty\}$, the case $r_i = +\infty$ corresponds to the acceptance-rejection method.)

Let $F : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $\mathbb{E}[|F(X)|] < +\infty$. We have

$$\mathbb{E}[F(X)] = \sum_{i \in I} \mathbb{E}[\mathbf{1}_{\{X \in A_i\}} F(X)] = \sum_{i \in I} p_i \mathbb{E}[F(X)|X \in A_i] = \sum_{i \in I} p_i \mathbb{E}[F(X_i)].$$

The stratification concept comes into play now. Let M be the global budget allocated to the computation of $\mathbb{E}[F(X)]$ and let $M_i = q_i M$ be the budget allocated to compute $\mathbb{E}[F(X_i)]$ in each stratum (with $0 \le q_i \le 1$, $i \in I$ and $\sum_{i \in I} q_i = 1$). This leads to define the (unbiased) estimator of $\mathbb{E}[F(X)]$:

$$\overline{F(X)}_M^I := \sum_{i \in I} p_i \frac{1}{M_i} \sum_{k=1}^{M_i} F\left(X_i^k\right), \tag{16}$$

where $(X_i^k)_{1 \le k \le M_i}$ is a $\mathcal{L}(X|X \in A_i)$-distributed random sample. We have

$$\mathrm{Var}\left(\overline{F(X)}_M^I\right) = \frac{1}{M} \sum_{i \in I} \frac{p_i^2}{q_i} \sigma_{F,i}^2, \tag{17}$$

where $\sigma_{F,i}^2 = \mathrm{Var}(F(X)|X \in A_i) = \mathrm{Var}(F(X_i))$, $i \in I$. Optimizing the allocation of the number of draws to the different strata amounts to solving the following minimization problem:

$$\min_{(q_i) \in \mathcal{P}_I} \sum_{i \in I} \frac{p_i^2}{q_i} \sigma_{F,i}^2 \quad \text{where } \mathcal{P}_I := \left\{ (q_i)_{i \in I} \in \mathbb{R}_+^I \,\middle|\, \sum_{i \in I} q_i = 1 \right\}. \tag{18}$$

3.1.1 Natural stratified sampling

A natural choice is to set

$$q_i = p_i, \quad i \in I. \tag{19}$$

since the weights p_i are known. Furthermore, this always reduces the variance.

$$\sum_{i \in I} \frac{p_i^2}{q_i} \sigma_{F,i}^2 = \sum_{i \in I} p_i \sigma_{F,i}^2 = \sum_{i \in I} \mathbb{E}\left[\left(F(X) - \mathbb{E}[F(X)|X \in A_i]\right)^2 \mathbf{1}_{A_i}(X)\right]$$
$$= \|F(X) - \mathbb{E}[F(X)|\sigma(\{X \in A_i\}, \ i \in I)]\|_2^2$$
$$\le \|F(X) - \mathbb{E}[F(X)]\|_2^2 = \mathrm{Var}(F(X)).$$

3.1.2 Optimal stratified sampling

The optimal choice is the solution to the constrained minimization problem (18). Schwarz's inequality yields

$$\sum_{i \in I} p_i \sigma_{F,i} = \sum_{i \in I} \frac{p_i \sigma_{F,i}}{\sqrt{q_i}} \sqrt{q_i} \le \left(\sum_{i \in I} \frac{p_i^2 \sigma_{F,i}^2}{q_i} \right)^{1/2} \left(\sum_{i \in I} q_i \right)^{1/2}.$$

The solution corresponds to the equality case in Schwarz's inequality, that is

$$q_i^* = \frac{p_i \sigma_{F,i}}{\sum_{j \in I} p_j \sigma_{F,j}}, \quad i \in I \tag{20}$$

with a resulting minimal variance of $\left(\sum_{i \in I} p_i \sigma_{F,i} \right)^2$. At this stage, the problem is that we do not *a priori* know the local inertia $\sigma_{F,i}^2$. Still, using the fact that L^p norms are decreasing with p, we see that

$$\sigma_{F,i} \geq \mathbb{E}\Big[\left| F(X) - \mathbb{E}\left[F(X) | \{ X \in A_i \} \right] \right| \,\Big|\, \{ X \in A_i \} \Big],$$

so that

$$\left(\sum_{i \in I} p_i \sigma_{F,i} \right)^2 \geq \left\| F(X) - \mathbb{E}\left[F(X) | \sigma(\{ X \in A_i \}, \ i \in I) \right] \right\|_1^2.$$

In [30], Étoré and Jourdain proposed an algorithm which adaptively modifies the proportion of further drawings in each stratum and which converges to the optimal allocation.

In Section 3.2, we show that the problem of designing good strata, in term of variance reduction is linked with optimal quantization. Besides, with quantization-based stratified sampling, the weights p_i are already known.

3.2 Quantization and stratified sampling

The main drawback of using quantization as a control variate is the repeated computations of the projections onto the quantizer. (Nearest neighbor searches in the case of a Voronoi quantizer.) In the case of stratified sampling, *one does not have to use a projection procedure.* Instead, we must focus on the cost of the simulation of conditional distributions $\mathcal{L}(X | X \in A_i)$, $i \in I$.

Proposition 3.1 brings together previous results and highlights the relationships with quantization. It shows that stratification has uniform efficiency among the class of Lipschitz continuous functionals.

Proposition 3.1 (Universal stratification). *Let $A = (A_i)_{i \in I}$ be a partition of E and let $\mathrm{Proj}_{A,X}$ denote the centroidal projection associated with X and A, defined in Definition 3.*

1. *Considering the local inertia of X in A_i, $\sigma_i^2 = \mathbb{E}\left[|X - \mathbb{E}[X|X \in A_i]|^2 \,\Big|\, X \in A_i \right]$, we have for every Lipschitz continuous function $F : E \to E$, $\sigma_{F,i} \leq [F]_{\mathrm{Lip}} \sigma_i$ where $[F]_{\mathrm{Lip}} = \sup\limits_{x \neq y} \frac{F(x) - F(y)}{|x-y|}$, so that*

$$\sup_{[F]_{\mathrm{Lip}} \leq 1} \sigma_{F,i} = \sigma_i, \tag{21}$$

2. *In the case of natural stratified sampling (see Section 3.1.1),*

$$\sup_{[F]_{\mathrm{Lip}} \leq 1} \left(\sum_{i \in I} p_i \sigma_{F,i}^2 \right) = \sum_{i \in I} p_i \sigma_i^2 = \left\| X - \mathbb{E}[X | \sigma(\{ X \in A_i \}, \ i \in I)] \right\|_2^2 = \left\| X - \mathrm{Proj}_{A,X}(X) \right\|_2^2. \tag{22}$$

3. *In the case of the optimal choice (see Section 3.1.2),*

$$\sup_{[F]_{\mathrm{Lip}} \leq 1} \left(\sum_{i \in I} p_i \sigma_{F,i} \right)^2 = \left(\sum_{i \in I} p_i \sigma_i \right)^2, \tag{23}$$

and

$$\left(\sum_{i \in I} p_i \sigma_i \right)^2 \geq \left\| X - \mathbb{E}[X | \sigma(\{ X \in A_i \}, \ i \in I)] \right\|_1^2 = \left\| X - \mathrm{Proj}_{A,X}(X) \right\|_1^2.$$

4. *In the case of real-valued Lipschitz continuous functions $F : E \to \mathbb{R}$, Equalities (21), (22) and (23) hold as inequalities.*

Proof: We have

$$\sigma_{F,i}^2 = \text{Var}\left(F(X)|X \in A_i\right) = \mathbb{E}\left[|F(X) - \mathbb{E}[F(X)|X \in A_i]|^2 \Big| X \in A_i\right]$$
$$\leq \mathbb{E}\left[|F(X) - F(\mathbb{E}[X|X \in A_i])|^2 \Big| X \in A_i\right].$$

Now using that F is Lipschitz continuous, we get

$$\sigma_{F,i}^2 \leq [F]_{\text{Lip}}^2 \frac{1}{p_i}\mathbb{E}\left[|X - \mathbb{E}[X|X \in A_i]|^2 \, \mathbf{1}_{\{X \in A_i\}}\right] = [F^2]_{\text{Lip}}\sigma_i^2.$$

Items 2 and 3 easily follow from Item 1. Equality follows by considering $F = Id_E$. $\qquad\square$

3.2.1 Universal stratified sampling

Proposition 3.1 suggests, in the case of Lipschitz continuous functionals, to set

$$q_i = \frac{p_i \sigma_i}{\sum\limits_{j \in I} p_j \sigma_j}, \quad j \in I,$$

so that we have uniform efficiency among the class of Lipschitz continuous functionals. This allocation scheme will be further referred to as the "universal stratification" weights. It also shows that, in the Lipschitz continuous case, it is always beneficial to reduce the quadratic distortion associated with the centroidal projection $\text{Proj}_{A,X}$.

Still, this minimization should not be done at the expense of the efficiency of the simulation of the corresponding conditional distributions. We should reach for a balance between the efficiency of the simulation in the strata and the quadratic quantization error controlling the variance reduction. For example, in Section 4, in the functional case, we will use optimal product quantizers, which are rate optimal (and numerically near optimal) and allow for a much more efficient simulation than real optimal functional quantization.

Remark. *We should also mention the adaptive strata design proposed in [29, 14].*

3.3 Simulation in hyper-rectangular strata in the independent Gaussian case

Consider $X \overset{\mathcal{L}}{\sim} \mathcal{N}(0, I_d)$, $d \geq 1$ and $(e_1, \ldots, e_d)$ an orthonormal basis of $\mathbb{R}^d$. Let $N_1, \ldots, N_d \geq 1$ be the number of strata in each direction and for $1 \leq i \leq d$, $-\infty = \gamma_0^i \leq \gamma_1^i \leq \cdots \leq \gamma_{N_i}^i = +\infty$. We define

$$A_{\underline{i}} := \bigcap_{l=1}^d \left\{x \in \mathbb{R}^d \text{ such that } \langle e_l, x \rangle \in [x_{i_l-1}^l, x_{i_l}^l]\right\}, \quad \underline{i} \in \prod_{l=1}^d \{1, \ldots, N_l\}.$$

Then for every $\underline{i} \in \prod_{l=1}^d \{1, \ldots, N_l\}$, $\mathcal{L}\left(X|X \in A_{\underline{i}}\right) = \bigotimes_{l=1}^d \mathcal{L}\left(Z|Z \in \left[\gamma_{i_l-1}^l, \gamma_{i_l}^l\right]\right)$, where $Z \overset{\mathcal{L}}{\sim} \mathcal{N}(0,1)$, $p_{\underline{i}} = \mathbb{P}(A_{\underline{i}}) = \prod_{l=1}^d \left(\mathcal{N}(\gamma_{i_l}^l) - \mathcal{N}(\gamma_{i_l-1}^l)\right)$ and for $-\infty \leq a \leq b \leq \infty$,

$$\mathcal{L}\left(Z|Z \in [a,b]\right) = \mathcal{N}^{-1}\left((\mathcal{N}(b) - \mathcal{N}(a))U + \mathcal{N}(a)\right), \quad U \overset{\mathcal{L}}{\sim} \mathcal{U}([0,1]). \tag{24}$$

4 Functional stratification of Gaussian processes

In the functional case, the state space of the random values are functional spaces. What is usually done is to simulate a scheme to approximate marginals of the underlying process.

In this section, we assume that X is a centered $\mathbb{R}$-valued bi-measurable Gaussian process on $[0, T]$ that satisfies $\int_0^T \mathbb{E}[X_t^2]dt < \infty$. We are interested by the value of $\mathbb{E}[F(X_{t_0}, X_{t_1}, \ldots, X_{t_n})]$ for some real function F, where $0 = t_0 \le t_1 \le \cdots \le t_n = T$ are $n+1$ dates of interest for the underlying process.

(For example, X can be a standard Brownian motion on $[0, T]$, and one computes the risk-neutral expectation of a path-dependent payoff of a diffusion based on X.)

The results of this section can be easily generalized to the multi-dimensional case, like multifactor diffusions. Still we restrict ourselves to the one-dimensional setting for clarity.

Let us assume that $\chi \in \mathcal{O}_{pq}(X, N)$ is a K-L optimal product quantizer of X. The codebook associated with this product quantizer is the set of the paths of the form

$$\chi_{\underline{i}} = \sum_{n \ge 1} \sqrt{\lambda_n^X} \gamma_{i_n}^{(N_n)} e_n^X, \quad \underline{i} = \{i_1, \ldots, i_n, \ldots\},$$

with the same notation as in Section 1.6.2. We now need to be able to simulate the conditional distribution

$$\mathcal{L}(X | X \in A_{\underline{i}})$$

where $A_{\underline{i}}$ is the cell associated with $\chi_{\underline{i}}$ in the codebook. To simulate the conditional distribution $\mathcal{L}(X | X \in A_{\underline{i}})$, one will :

- First, simulate the first K-L coordinates of X, using (24).

- Then simulate the conditional distribution of the marginals of the Gaussian process given its first K-L coordinates.

Remark. *We have chosen to use K-L optimal product quantizers instead of optimal quantizers because in this case, the Voronoi cells in this are hyper-rectangles, which allows us to simulate the first K-L coordinates more easily than in the general case. Moreover, the rate of decay of the quantization errors is rate-optimal under some conditions on the Karhunen-Loève eigenvalues which are verified in the considered examples [18].*

4.1 Simulation of marginals of the Gaussian process, given its d first K-L coordinates

In this setting, the aim is to simulate the conditional distribution

$$\mathcal{L}\Big(X_{t_0}, \ldots, X_{t_n} \,\Big|\, \int_0^T X_s e_1^X(s)ds, \int_0^T X_s e_2^X(s)ds, \ldots, \int_0^T X_s e_d^X(s)ds\Big) \tag{25}$$

where $(X_t)_{t \in [0,T]}$ is an L^2 $\mathbb{R}$-valued Gaussian process, and $(e_k^X, \lambda_k^X)_{k \in \mathbb{N}^*}$ is the Karhunen-Loève system associated with the process X. Hence $\Big(X_{t_0}, \ldots, X_{t_n}, \int_0^T X_s e_1^X(s)ds, \ldots, \int_0^T X_s e_d^X(s)ds\Big)$ is a Gaussian vector.

As a consequence, if we denote $Y := \begin{pmatrix} \int_0^T X_s e_1^X(s)ds \\ \vdots \\ \int_0^T X_s e_d^X(s)ds \end{pmatrix}$ and $V := \begin{pmatrix} X_{t_0} \\ \vdots \\ X_{t_n} \end{pmatrix}$, the conditional distribution

(25) is given by the transition kernel $\nu(y, A) = \mathcal{N}\big(Af_{V|Y}(y), \text{cov}(V - \mathbb{E}[V|Y])\big)$, where $Af_{V|Y} : \mathbb{R}^d \to \mathbb{R}^n$ is an affine function corresponding to the linear regression of V on Y, $Af_{V|Y}(Y) := \mathbb{E}[V|Y]$.

- We have $Af_{V|Y}(Y) = \mathbb{E}[V] + R_{V|Y}Y$ where $R_{V|Y} = \text{cov}(V, Y)\text{cov}(Y)^{-1}$. Using that $\text{cov}(Y) = \big(\lambda_i^X \delta_{ij}\big)_{1 \le i,j \le d}$ and $\text{cov}(V, Y) = (\lambda_k^X e_k^X(t_i))_{0 \le i \le n, 1 \le k \le d}$, we get

$$R_{V|Y} = \big(e_j^X(t_i)\big)_{0 \le i \le n, 1 \le j \le d}. \tag{26}$$

- The covariance matrix is

$$K := \operatorname{cov}\left(V - \mathbb{E}[V|Y]\right) = \mathbb{E}\left[\left(V - R_{V|Y}Y\right)\left(V - R_{V|Y}Y\right)\right]$$
$$= \operatorname{cov}(V) - 2\operatorname{cov}\left(V, R_{V|Y}Y\right) + \operatorname{cov}\left(R_{V|Y}Y\right) = \operatorname{cov}(V) - \operatorname{cov}\left(R_{V|Y}Y\right)$$
$$= \left(\operatorname{cov}(V_l, V_k) - \sum_{i=1}^{d} \lambda_i e_i^X(t_l) e_i^X(t_k)\right)_{0 \le k,l \le n}.$$

The easiest way to simulate according to this probability distribution would be to use the Cholesky factorization of K. However, when using this method, the simulation of a simple path involves the quadratic complexity of an $n \times n$ matrix multiplication, which is not satisfactory for our purpose.

4.2 Faster simulation of conditional paths - Bayesian simulation

As pointed out earlier, the naive simulation method for $\mathcal{L}(V|Y)$ requires for each path a multiplication by a Cholesky transform of K whose cost is $O(n^2)$.

- Yet, the quantization dimension d of the process is close to $\log(N)$ where N is the number of strata, and n, the number of time steps, is usually very large compared to d.

- The idea here is that the conditional distribution $\mathcal{L}(V|Y)$ is determined through the Bayes lemma, by the conditional distribution $\mathcal{L}(Y|V)$ and the two marginal distributions $\mathcal{L}(V)$ and $\mathcal{L}(Y)$.

One knows that $V = \mathbb{E}[V|Y] \overset{\perp\!\!\!\perp}{+} Z$ where $Z \overset{\mathcal{L}}{\sim} \mathcal{N}(0, \operatorname{cov}(V - \mathbb{E}[V|Y]))$ is independent of Y. Hence one is able to simulate according to $\mathcal{L}(V|Y = y)$ if one can simulate the distribution of Z, writing $\mathcal{L}(V|Y = y) = \mathbb{E}[V|Y = y] + \mathcal{L}(Z)$. This decomposition corresponds to the splitting of the Karhunen-Loève expansion:

$$\begin{pmatrix} V_0 \\ \vdots \\ V_n \end{pmatrix} = \underbrace{\sum_{k=1}^{d} \underbrace{\sqrt{\lambda_k^X}\xi_k}_{=Y_k} \begin{pmatrix} e_k^X(t_0) \\ \vdots \\ e_k^X(t_n) \end{pmatrix}}_{=\mathbb{E}[V|Y]} \overset{\perp\!\!\!\perp}{+} \underbrace{\sum_{l \ge d+1} \sqrt{\lambda_k^X}\xi_k \begin{pmatrix} e_k^X(t_0) \\ \vdots \\ e_k^X(t_n) \end{pmatrix}}_{=Z}.$$

To simulate Z, one simulates the distribution of V and the conditional distribution $\mathcal{L}(Z|V)$.

$$\text{We have} \quad \mathcal{L}(Z|V) \overset{\mathcal{L}}{\sim} \delta_V - \mathcal{L}(\mathbb{E}[V|Y]|V) \overset{\mathcal{L}}{\sim} \delta_V - Af_{V|Y}\mathcal{L}(Y|V)$$
$$\overset{\mathcal{L}}{\sim} \delta_V - Af_{V|Y}\mathcal{N}(\mathbb{E}[Y|V], \operatorname{cov}(Y - \mathbb{E}[Y|V])).$$

If $Af_{Y|V}$ is the affine function corresponding to the regression of Y on V and $R_{Y|V}$ its linear part,

$$\operatorname{cov}(Y - \mathbb{E}[Y|V]) = \operatorname{cov}(Y) + \operatorname{cov}(\mathbb{E}[Y|V]) - 2\operatorname{cov}(Y, \mathbb{E}[Y|V]) = \operatorname{cov}(Y) - R_{Y|V}\operatorname{cov}(V)^t R_{Y|V}.$$

This yields $Z = V - Af_{V|Y}(G)$ where $G \overset{\mathcal{L}}{\sim} \mathcal{N}(Af_{Y|V}(V), \operatorname{cov}(Y) - R_{Y|V}\operatorname{cov}(V)^t R_{Y|V})$. Finally, we can use the following method to simulate the conditional distribution of V Y.

- Simulate V. $\hfill O(n).$

- Simulate $G \overset{\mathcal{L}}{\sim} \mathcal{N}\left(Af_{Y|V}(V), \operatorname{cov}(Y) - R_{Y|V}\operatorname{cov}(V)^t R_{Y|V}\right)$ $\hfill O(d \times d).$

- Compute $Z = V - Af_{V|Y}(G)$. $\hfill O(d \times n).$

  The random variable $T := Af_{V|Y}(y) + Z$ satisfies $T \overset{\mathcal{L}}{\sim} \mathcal{L}(V|Y = y)$.

Let us remind that $Af_{V|Y}$ is trivially defined in Equation (26), because coordinates of Y are independent. Other matrices implied in this algorithm are computed prior to any Monte Carlo simulation. In general, $R_{Y|V}$ can simply be computed by performing a numerical least-square regression. Moreover in the special cases of Brownian motion, Bronian bridge and Ornstein-Uhlenbeck processes, there are closed-form expressions for the $R_{Y|V}$, which we present in Appendix B.

In the case of Brownian motion, for a uniform time discretization mesh $t_j = \frac{jT}{n} = jh$, $0 \le j \le n$, this yields $R_{Y|V} = (\alpha_{ij})_{1 \le i \le d, 0 \le j \le n}$, with

- $\alpha_{ij} = \lambda_i^W \frac{2e_i^W(t_j) - e_i^W(t_{j-1}) - e_i^W(t_{j+1})}{h}$ for $j \notin \{0, n\}$,

- $\alpha_{i0} = \lambda_i^W \left(\left(e_i^W\right)'(t_0) - \frac{e_i^W(t_1) - e_i^W(t_0)}{h} \right)$,

- $\alpha_{in} = \lambda_i^W \left(\frac{e_i^W(t_n) - e_i^W(t_{n-1})}{h} - \left(e_i^W\right)'(t_n) \right)$.

We now have a very fast and easy way to simulate the conditional distribution (25) at our disposal.

In Figures 5 and 6, we plot a few paths of the conditional distribution of various Gaussian processes given that they belong to a given L^2 Voronoi cell. The appearance of the drawing suggests to consider the method as a "guided Monte Carlo simulation".
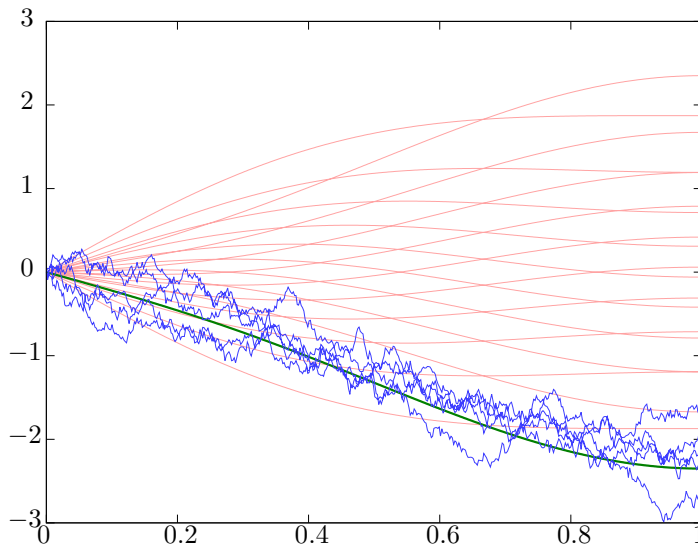

Figure 5: A few paths of the conditional distribution of Brownian motion, given that its path belong to the L^2 Voronoi cell of the highlighted curve in the quantizer.

4.3   Blind optimization procedures for the universal strata design

We have seen in Section 3.2 that the quantity $d(\chi) = \left(\sum\limits_{\chi_{\underline{i}} \in \Gamma} p_{\underline{i}} \sigma_{\underline{i}} \right)^2$ is an upper bound of the variance of the estimator, given in Equation (16) in the case where the functional is 1-Lipschitz continuous. Hence one may want to minimize this criterion instead of the L^2 quantization error. This yields the minimization problem

$$\mathcal{D}_N^{pq} := \min \left\{ d(\chi), \ \chi \in \mathcal{O}_{pq}(X, N) \right\} \tag{27}$$
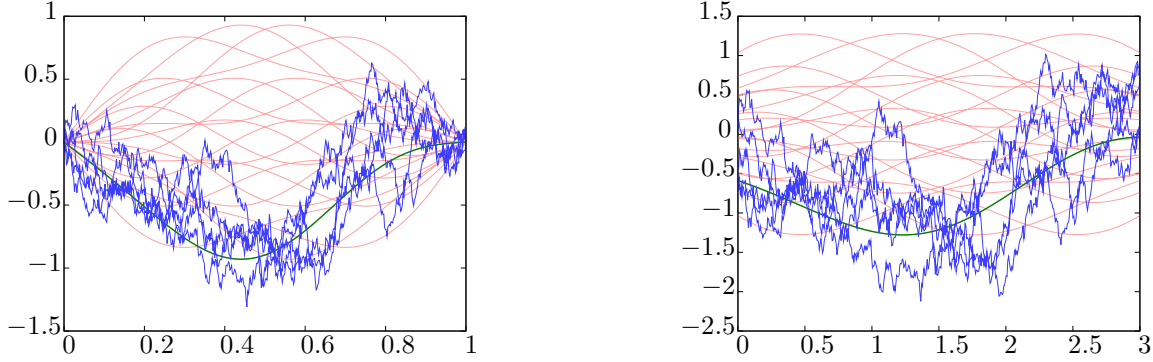
Figure 6: A few paths of the conditional distribution of Brownian bridge (left) and the stationary Ornstein-Uhlenbeck process (right), given that its path belong to the L^2 Voronoi cell of the highlighted curve in the quantizer.

instead of the minimization problem (13).

The same kind of blind optimization procedure as in Section 1.6.3 can be performed. Some values of the optimal decomposition for Brownian motion are given in Table 2.

N	N_{rec}	$d(\chi)$	Product decomposition
1	1	0.5	1
10	10	$9.75689 \cdot 10^{-2}$	5×2
100	96	$5.10548 \cdot 10^{-2}$	$12 \times 4 \times 2$
1000	966	$3.51289 \cdot 10^{-2}$	$23 \times 7 \times 3 \times 2$
10000	9984	$2.63721 \cdot 10^{-2}$	$26 \times 8 \times 4 \times 3 \times 2 \times 2$

Table 2: Record of optimal product decomposition of Brownian motion with respect to the criterion (27).

Optimal product decompositions for both Brownian bridge and Brownian motion and for a wide range of values of N are available on the web site www.quantize.maths-fi.com [26] for download. When comparing decompositions for levels lesser 11000, we notice that in the case of Brownian motion, the optimal decompositions for both criteria are "almost" always the same. The only values where decompositions differ are the ranges $270 - 271$ and $3328 - 3359$, and even then, the two criteria result in similar decompositions. Hence in practice, we can use the same database for both criteria. Nonetheless, in the case of Brownian bridge and Ornstein-Uhlenbeck processes, the optimal decompositions resulting from the two criteria differ more often.

4.4 Functional stratification of solutions of stochastic differential equations

We consider the SDE
$$dF_t = b(t, F_t)dt + \sigma(t, F_t)dX_t, \quad t \in [0, T], \qquad F_0 = f_0 \tag{28}$$

where X is a centered continuous Gaussian semimartingale starting from 0 and where b and σ are Borel functions, Lipschitz continuous in x uniformly in $t \in [0, T]$ such that $|b(\cdot, 0)| + |\sigma(\cdot, 0)|$ is bounded over $[0, T]$. In this situation, (28) admits a unique strong solution X and $\sup_{t \in [0,T]} |X_t|$ has r-moments for every $r \in (0, \infty)$.

Remark. *In this case, thanks to Fernique's theorem, the continuity assumption on the Gaussian process ensures that $\int_0^T \mathbb{E}[X_s^2]ds < \infty$ and the continuity of the covariance function, (see [13, VIII.3]).*

The most common approach to perform a Monte Carlo simulation with the solution of such a stochastic differential equation, is to use a discretization scheme like the Euler scheme [9]. In this setting, we propose

to simply replace Gaussian process X by a stratified version of X in the Euler scheme. This approach is justified in many ways:

1. In [7], using filtration enlargement techniques, it is proved that under some additional hypothesis on the Gaussian semimartingale X, its conditional distribution in a strata is still a semimartingale with respect to its own filtration. This additional hypothesis is satisfied by Brownian motion, Brownian bridge and Ornstein-Uhlenbeck processes. Therefore, plugging the stratified Euler scheme into the SDE amounts to using the Euler scheme of these conditional stochastic differential equations.

2. In the one-dimensional setting, if we make the additional hypothesis that $\sigma \in C^1([0,T] \times \mathbb{R}, \mathbb{R})$ is positive and bounded, as soon as the drift of the Lamperti transform of the SDE (28) is Lipschitz continuous, the unique strong solution of (28), seen as a functional of the underlying Gaussian process X is $\| \cdot \|_p$-Lipschitz continuous [20]. Hence we can apply the results of Section 3.2 on universal stratification for Lipschitz continuous functionals.

3. The function $(X_{t_0}, \ldots, X_{t_n}) \mapsto (X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}})$ that maps the marginals of Brownian motion to the corresponding increments used in the Euler scheme, is a linear map from $\mathbb{R}^{n+1}$ to $\mathbb{R}^n$ and thus Lipschitz continuous as well.

5   Application to option pricing

The special case of Brownian motion allows us to use functional stratification as a generic variance reduction method for the case of functionals of Brownian diffusions, even in the multidimensional case, regardless of how the Brownian paths are correlated or used afterwards, to drive the diffusion of an underlying stock, a stochastic volatility process or a discount factor. As it only impacts how the Monte Carlo simulation is input with Brownian paths, our approach is easier to implement in a practical setting than adaptive variance reduction methods, which generally require a control loop.

In this section, we study the performance of our method in simple one-dimensional cases. We begin with the case of a continuous-time Up-In Call option in the Black and Scholes model, for which a closed-form expression is known, and used as a Benchmark.

5.1   Benchmark with an Up-In Call in the Black and Scholes model

We evaluate our method in the case of a path dependent option where a reference value can easily be computed: an Up-In Call barrier option in the Black and Scholes model. For the sake of simplicity, we assume that there is no drift (no interest rate and no dividend). There is a closed-form expression for the continuous barrier option price, but we must resort to a numerical approximation [3] (yet very accurate) on that closed-form expression to get the price in the case of discrete dates for the barrier. The total size of the Monte Carlo sample is 100000 in every case.

We price the Up-In Call option with different values of the initial spot S, the strike K, the barrier H, the volatility σ, the maturity T, and the number of fixing dates for the discrete barrier n. In every case, a 95% confidence interval is given. So is the variance of the estimator.

The numerical results are reported in Table 3 when using the method with 20 stratas and Table 4 when using the method with 100 stratas. In this tables, the first column correspond to Broadie and Glasserman's closed-form expression proxy. The second one corresponds to a simple Monte Carlo estimator. The last three columns correspond to stratified sampling estimators with different simulation allocation strategies.

The "natural weights" column stands for the allocation budget of Equation (19). The "Lip.-optimal weights" column stand for the "universal stratification" budget allocation proposed in Section 3.2. In these two cases, we have an explicit allocation rule which does not depend on the payoff function. The last column, "optimal weights" corresponds to an estimation of the optimal budget allocation given in Equation (20).

Parameter Values	Broadie & Glasserman's proxy	Simple estimator	Strat. estimator natural weights	Strat. estimator Lip.-optimal weights	Strat. estimator optimal weights
$S = 100,\ K = 100$ $H = 125,\ \sigma = 0.3,$ $T = 1.5,\ n = 365$	13.9597	14.0379 [13.8705, 14.2053] Var = 729.2518	13.9281 [13.8491, 14.0071] Var = 162.4650	13.9283 [13.8519, 14.0047] Var = 151.9481	13.9364 [13.8827, 13.9901] Var = 75.1319
$S = 100,\ K = 100$ $H = 200,\ \sigma = 0.3,$ $T = 1,\ n = 365$	1.3665	1.4206 [1.3442, 1.4969] Var = 151.6366	1.3659 [1.3106, 1.4211] Var = 79.5118	1.3510 [1.3039, 1.3981] Var = 57.7425	1.3602 [1.3472, 1.3732] Var = 4.4053

Table 3: Numerical results for the Up-In Call option, with 20 stratas.

Parameter Values	Broadie & Glasserman's proxy	Simple estimator	Strat. estimator natural weights	Strat. estimator Lip.-optimal weights	Strat. estimator optimal weights
$S = 100,\ K = 100$ $H = 125,\ \sigma = 0.3,$ $T = 1.5,\ n = 365$	13.9597	14.0379 [13.8705, 14.2053] Var = 729.2518	13.9382 [13.8720, 14.0043] Var = 114.0634	13.9511 [13.8874, 14.0150] Var = 105.8760	13.9483 [13.9047, 13.9919] Var = 49.5071
$S = 100,\ K = 100$ $H = 200,\ \sigma = 0.3,$ $T = 1,\ n = 365$	1.3665	1.4206 [1.3442, 1.4969] Var = 151.6366	1.3296 [1.2825, 1.3768] Var = 57.8899	1.3493 [1.3093, 1.3893] Var = 41.6666	1.3611 [1.3508, 1.3715] Var = 2.8099

Table 4: Numerical results for the Up-In Call option, with 100 stratas.

5.2 Test with an Auto-Call in the CEV model

We assume that the stock follows a CEV model with no drift $dS_t = \sigma S_t^{\frac{\beta}{2}} dW_t,\quad 0 < \beta < 2$. We used the Euler scheme on $\ln(S_t)$, which satisfies the SDE $d\ln(S_t) = -\frac{\sigma^2}{2} S_t^{\beta-2} dt + \sigma S_t^{\frac{\beta}{2}-1} dW_t$.

Description of the Auto-Call payoff:

Let S_t be the stock price and $0 = t_0 < t_1 < \cdots < t_n = T$ be the observation dates. K and H are the "strike" and the "barrier" values. P denotes the "nominal", and C a zero-coupon bond of maturity T.

At the first date t_1 of the schedule, if $S_{t_1} > K$, the holder of the option receives $(1+C)P$ and the contract expires. If $S_{t_1} \le K$, he waits until the second date of the schedule. If $S_{t_2} > K$, the holder gets $(1+C)P$ and the contract expires. And so on... If S_t does not reach K on $[0,T)$, the contract is exerciced as follows: if $S_T > K$, the holder gets $(1+C)P$. If $H < S_T \le K$, the holder gets P and if $S_T \le H$, he gets $P\frac{S_T}{K}$.

The numerical results are reported in Table 5 when implementing the stratification method with 20 and 50 stratas. The parameters of the model are $\beta = 1.5$, $S_0 = 100$, $\sigma = 0.3$. For the payoff, $K = 110$, $H = 80$, $P = 100$, $C = 0.07$. The considered observation dates are $\{1, 2, 3\}$. The number of time steps in the Euler scheme is 300 and the total size of the Monte Carlo sample is 100000 in every case.

Number of strata	Simple estimator	Strat. estimator natural weights	Strat. estimator Lip.-optimal weights	Strat. estimator optimal weights
20	99.0598 [98.9887, 99.1310] Var = 131.8089	99.0839 [99.0438, 99.1239] Var = 41.8067	99.0886 [99.0488, 99.1284] Var = 41.2888	99.0477 [99.0184, 99.0769] Var = 22.2549
50	99.0598 [98.9887, 99.1310] Var = 131.8089	99.0507 [99.0129, 99.0886] Var = 37.3150	99.0790 [99.0414, 99.1166] Var = 36.8408	99.0444 [99.0179, 99.0709] Var = 18.2954

Table 5: Numerical results for the Auto-Call option in the CEV model, with 20 and 50 stratas.

5.3 Test with an Asian straddle in the one-factor Schwartz model

Here, we stand in the case of a stock which follows the following SDE:

$$dS_t = \theta(\alpha - \ln S_t)S_t dt + \sigma S_t dW_t, \tag{29}$$

under the risk-neutral probability. The stochastic process $X = \ln(S)$ is an Ornstein-Uhlenbeck process:

$$dX_t = \theta(\mu - X_t)dt + \sigma dW_t \quad \text{with } \mu = \alpha - \frac{\sigma^2}{2\theta}. \tag{30}$$

This model was proposed by Schwartz in [27]. Such exponentials of Ornstein-Uhlenbeck processes are commonly met in commodity derivatives. One particularity in these markets is that the spot is generally not directly traded. Therefore, the underlyings of derivatives are generaly futures. Still, we use this one-factor "toy" model as a simple case study for our variance reduction method.

The considered payoff is an *Asian straddle* option on a discrete schedule of observation dates $t_0 < \cdots < t_n = T$. K is the "strike" of the option whose payoff is $\left| \frac{1}{n+1} \sum_{k=0}^{n} S_{t_k} - K \right|$.

We perform a functional stratified sampling of the Ornstein-Uhlenbeck process. Optimal product decompositions for the criterion (27) are used and available in Table 6 where the numerical results are reported.

The parameters are $S_0 = 100$, $\theta = 0.3$, $\alpha = \ln(110)$, $\sigma = 0.3$ and $K = 100$. The total size of the Monte Carlo sample is 100000 in every case. The observation dates are $\left(i\frac{T}{n}\right)_{i=\{0,\ldots,n\}}$ with $T = 3$ and $n = 36$.

Number of strata and product decomposition	Simple estimator	Strat. estimator natural weights	Strat. estimator Lip.-optimal weights	Strat. estimator optimal weights
20 $20 = 10 \times 2$	17.5393 [17.4504, 17.6282] Var = 205.9375	17.6140 [17.5871, 17.6408] Var = 18.8041	17.6118 [17.5856, 17.6378] Var = 17.5502	17.6240 [17.6006, 17.6477] Var = 14.6363
100 $100 = 10 \times 5 \times 2$	17.5393 [17.4504, 17.6282] Var = 205.9375	17.6101 [17.5850, 17.6351] Var = 16.2945	17.6122 [17.5884, 17.6360] Var = 14.7316	17.6147 [17.5932, 17.6362] Var = 12.0112

Table 6: Numerical results for the Asian straddle option in Schwartz's model, with 20 and 100 stratas.

To perform this computation, one needs to use a non-centered Ornstein-Uhlenbeck quantizer. Building such a quantizer is a straightforward extension of the centered case. As showed in Section A, if X is an Ornstein-Uhlenbeck process on $[0, T]$ following the dynamic $dX_t = \theta(\mu - X_t)dt + \sigma dW_t$, $X_0 \overset{\mathcal{L}}{\sim} \mathcal{N}(m_0, \sigma_0^2)$, with nonzero values of μ and m_0, we have

$$X_t = \underbrace{m_0 e^{-\theta t} + \mu(1 - e^{-\theta t})}_{(1)=\text{non-stochastic path}} + \left(\begin{array}{c} \text{centered Ornstein-Uhlenbeck process} \\ \text{corresponding to } m_0 = \mu = 0 \end{array} \right). \tag{31}$$

In Figure 7, we display the functional product quantizer of a non-centered Ornstein-Uhlenbeck process.

5.4 Comments on the numerical results

"Lipschitz-optimal" strata and weights are not more difficult to compute than the "natural" scheme (19) since all the involved parameters are known. This stratified sampling method does not depend on the payoff function but only on the distribution of the underlying asset which means that it can be plugged upstream in the Monte Carlo simulator. In terms of variance reduction, universal stratification is all the more preforming as the variance of the plain estimator is high, most likely because more strata are "activated". When the payoff function is symmetric, like with the *Asian straddle* it achieves up to 90% of the variance reduction achieved by a payoff-dependent dedicated method like the one devised in [30].

A The Karhunen-Loève expansion of Ornstein-Uhlenbeck processes

In this section, we derive the Karhunen-Loève expansion of the Ornstein-Uhlenbeck process. Proposition A.3 brings the results together. Section A.3 presents the numerical method for computing this expansion.
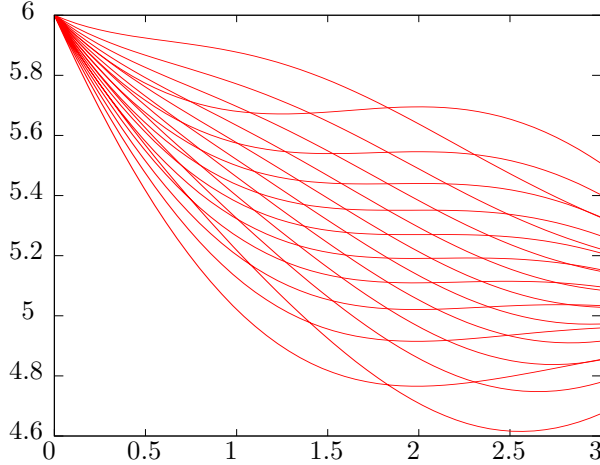
Figure 7: Functional 10×2-product quantizer of an Ornstein-Uhlenbeck process starting from $X_0 = 6$ defined by the diffusion $dX_t = \theta(\mu - X_t)dt + \sigma dW_t$ with $\mu = 5$, $\sigma = 0.3$ and $\theta = 0.8$ on $[0, 3]$.

A.1 The Ornstein-Uhlenbeck process

The Ornstein-Uhlenbeck process is defined by the SDE

$$dX_t = \theta(\mu - X_t)dt + \sigma dW_t, \qquad \text{with } \sigma \geq 0 \text{ and } \theta > 0, \tag{32}$$

which gives

$$X_t = X_0 e^{-\theta t} + \mu(1 - e^{-\theta t}) + \int_0^t \sigma e^{\theta(s-t)} dW_s. \tag{33}$$

We assume that X_0 is Gaussian ($X_0 \stackrel{\mathcal{L}}{\sim} \mathcal{N}(m_0, \sigma_0^2)$) and independent from W. We have $\mathbb{E}[X_t] = m_0 e^{-\theta t} + \mu(1 - e^{-\theta t})$ and $\text{cov}(X_s, X_t) = \frac{\sigma^2}{2\theta} e^{-\theta(s+t)} \left(e^{2\theta \min(s,t)} - 1 \right) + \sigma_0^2 e^{-\theta(s+t)}$. Moreover $\lim_{t \to \infty} \text{Var}(X_t) = \frac{\sigma^2}{2\theta}$ (the long-term variance). If the initial variance σ_0^2 is equal to long-term variance $\frac{\sigma^2}{2\theta}$, X is stationary and the covariance function is given by $\text{cov}(X_s, X_t) = \frac{\sigma^2}{2\theta} e^{-\theta|s-t|}$. The total variance of the process on $[0, T]$ is

$$\|X\|_2^2 = \int_0^T \text{Var}(X_s) ds = \frac{\sigma^2 T}{2\theta} + \left(\sigma_0^2 - \frac{\sigma^2}{2\theta} \right) \left(\frac{1}{2\theta} - \frac{e^{-2\theta T}}{2\theta} \right).$$

A.2 The Ornstein-Uhlenbeck covariance operator

The Ornstein-Uhlenbeck covariance operator is given by

$$T^{OU} f(t) = \int_0^T \frac{\sigma^2}{2\theta} e^{-\theta(s+t)} \left(e^{2\theta \min(s,t)} - 1 \right) f(s) ds + \int_0^T \sigma_0^2 e^{-\theta(s+t)} f(s) ds. \tag{34}$$

Computing the Karhunen-Loève expansion of the Ornstein-Uhlenbeck process
T^{OU} is a compact Hermitian positive operator on the separable Hilbert space $L^2([0, T])$. Hence there exists an orthonormal basis consisting of eigenvectors of T^{OU} and eigenvalues are real and nonnegative. Moreover $\left\| T^{OU} \right\|^2 \leq \frac{\sigma^2 T}{2\theta} + \frac{\sigma^2}{4\theta^2} \left(e^{-2\theta T} - 1 \right)$. We have

$$T^{OU} f(t) = \int_0^t \frac{\sigma^2}{2\theta} e^{\theta(s-t)} f(s) ds + \int_t^T \frac{\sigma^2}{2\theta} e^{\theta(t-s)} f(s) ds + \int_0^T \left(\sigma_0^2 - \frac{\sigma^2}{2\theta} \right) e^{-\theta(s+t)} f(s) ds.$$

Proposition A.1. *If $f \in C([0,1])$, and if $g = T^{OU}f$, then*

$$g'' - \theta^2 g = -\sigma^2 f, \tag{35}$$

with

$$\sigma_0^2 g'(0) = \left(\sigma^2 - \theta\sigma_0^2\right) g(0) \qquad \text{and} \qquad g'(T) = -\theta g(T). \tag{36}$$

Proof:

$$g(t) = \int_0^t \frac{\sigma^2}{2\theta} e^{\theta(s-t)} f(s) ds + \int_t^T \frac{\sigma^2}{2\theta} e^{\theta(t-s)} f(s) ds + \int_0^T \left(\sigma_0^2 - \frac{\sigma^2}{2\theta}\right) e^{-\theta(s+t)} f(s) ds.$$

$$g'(t) = -\frac{\sigma^2}{2} \int_0^t e^{\theta(s-t)} f(s) ds + \frac{\sigma^2}{2} \int_t^T e^{\theta(t-s)} f(s) ds - \left(\theta\sigma_0^2 - \frac{\sigma^2}{2}\right) \int_0^T e^{-\theta(s+t)} f(s) ds$$

$$g''(t) = \frac{\sigma^2\theta}{2} \left(\int_0^t f(s) e^{\theta(s-t)} ds + \int_t^T f(s) e^{\theta(t-s)} ds\right) + \theta \int_0^T \left(\theta\sigma_0^2 - \frac{\sigma^2}{2}\right) e^{-\theta(s+t)} f(s) ds - \sigma^2 f(t).$$

we get $g''(t) = \theta^2 g(t) - \sigma^2 f(t)$. Moreover, Equation (36) comes when identifying expressions with $t = 0$ and $t = T$.

$\square$

Proposition A.2. *Conversely, if $g \in C^2([0,T])$ and if functions f and g satisfy Equations (35) and (36) then $g = T^{OU}f$.*

Proof: Computing $T^{OU}g''$ yields:

$$T^{OU}g''(t) = \int_0^t \frac{\sigma^2}{2\theta} e^{\theta(s-t)} g''(s) ds + \int_t^T \frac{\sigma^2}{2\theta} e^{\theta(t-s)} g''(s) ds + \int_0^T \left(\sigma_0^2 - \frac{\sigma^2}{2\theta}\right) e^{-\theta(s+t)} g''(s) ds.$$

An integration by parts yields

$$\begin{aligned}
T^{OU}g'' &= -\sigma_0^2 g'(0) e^{-\theta t} - \sigma^2 g(t) + \tfrac{\sigma^2}{2} g(0) e^{-\theta t} - \left(\theta\sigma_0^2 - \tfrac{\sigma^2}{2}\right) g(0) e^{-\theta t} + \theta^2 T^{OU}g(t) \\
&= -\sigma^2 g(t) + \theta^2 T^{OU}g(t) \quad \text{thanks to (36).}
\end{aligned}$$

$\square$

Now, by necessary conditions, $T^{OU}f = \lambda f \Leftrightarrow \sigma^2 g = \lambda(\theta^2 g - g'')$. We obtain

$$\lambda g'' + (\sigma^2 - \lambda\theta^2)g = 0. \tag{37}$$

Hence the solution of the ordinary differential equation (37) on $[0,T]$ has the form $g(t) = A\cos(\omega t) + B\sin(\omega t)$, with $\omega = \sqrt{\frac{\sigma^2 - \lambda\theta^2}{\lambda}} \Leftrightarrow \lambda = \frac{\sigma^2}{\omega^2 + \theta^2}$. Equation (36) yields $\omega B\sigma_0^2 = (\sigma^2 - \theta\sigma_0^2)A$. Hence, we have $g(t) = K\left(\omega\sigma_0^2 \cos(\omega t) + (\sigma^2 - \theta\sigma_0^2)\sin(\omega t)\right)$, so that Equality (36) yields

$$\omega\sigma^2 \cos(\omega T) + \left(-\omega^2\sigma_0^2 + \theta\sigma^2 - \theta^2\sigma_0^2\right) \sin(\omega T) = 0. \tag{38}$$

Conversely, the same calculation shows that $\lambda_n \in \left]0, \left\|T^{OU}\right\|_2\right]$ is an eigenvalue of T^{OU} if and only if Equality (38) holds.

Proposition A.3. *Finally, if $(\omega_n)_{n\geq 1}$ is the increasingly sorted sequence of the positive solutions of (38), the Karhunen-Loève eigensystem $\left(e_n^{OU}, \lambda_n^{OU}\right)_{n\geq 1}$ of the Ornstein-Uhlenbeck process is*

- $\lambda_n^{OU} = \frac{\sigma^2}{\omega_n^2 + \theta^2}$, *and*

- $e_n^{OU}(t) = K_n\left(\omega_n\sigma_0^2 \cos(\omega_n t) + (\sigma^2 - \theta\sigma_0^2)\sin(\omega_n t)\right)$ *for $n \geq 1$, where K_n is the normalization constant.*

If $(\sigma, \sigma_0) \neq (0,0)$, K_n is given by

$$1/K_n^2 = \frac{1}{2\omega_n}\sigma_0^2(\sigma^2 - \theta\sigma_0^2)\left(1 - \cos(2\omega_n T)\right) + \frac{1}{2}\sigma_0^4\omega_n^2\left(T + \frac{1}{2\omega_n}\sin(2\omega_n T)\right)$$
$$+ \frac{1}{2}(\sigma^2 - \theta\sigma_0^2)^2\left(T - \frac{1}{2\omega_n}\sin(2\omega_n T)\right). \tag{39}$$

Case of a deterministic starting point: In this case ($\sigma_0 = 0$), we have

$$e_n^{OU}(t) = \frac{1}{\sqrt{\frac{T}{2} - \frac{\sin(2\omega_n T)}{4\omega_n}}} \sin(\omega_n t).$$

Stationary case: In the stationary case ($\sigma_0^2 = \frac{\sigma^2}{2\theta}$), we have $e_n^{OU}(t) = C_n\Big(\omega_n \cos(\omega_n t) + \theta \sin(\omega_n t)\Big)$, where C_n is the normalization constant. C_n is given by

$$1/C_n^2 = \frac{\theta}{2}\Big(1 - \cos(2\omega_n T)\Big) + \frac{\omega_n^2}{2}\Big(T + \frac{\sin(2\omega_n T)}{2\omega_n}\Big) + \frac{\theta^2}{2}\Big(T - \frac{\sin(2\omega_n T)}{2\omega_n}\Big).$$

A.3   Numerical computation of the Karhunen-Loève expansion of the Ornstein-Uhlenbeck process

This section focuses on the computation of the positive solutions to (38).

A.3.1   Deterministic starting point

In this case ($\sigma_0 = 0$), we can check that elements of $\big\{\frac{\pi}{2T} + k\frac{\pi}{T},\ k \in \mathbb{N}\big\}$ are not solutions of Equation (38). As a consequence, the equation comes to

$$\theta \tan(\omega T) = -\omega. \tag{40}$$

The case where $\theta = 0$ comes to the case of Brownian motion, hence we assume that $\theta \neq 0$. Solutions of this equation are illustrated in Figure 8. We can show that there is a unique solution ω_n in each interval $\big(\frac{n\pi}{T} - \frac{\pi}{2T}, \frac{n\pi}{T}\big)$, for $n \in \{1, 2, \ldots\}$ and that $\lim_{n\to\infty} \omega_n - \big(\frac{n\pi}{T} - \frac{\pi}{2T}\big) = 0$.
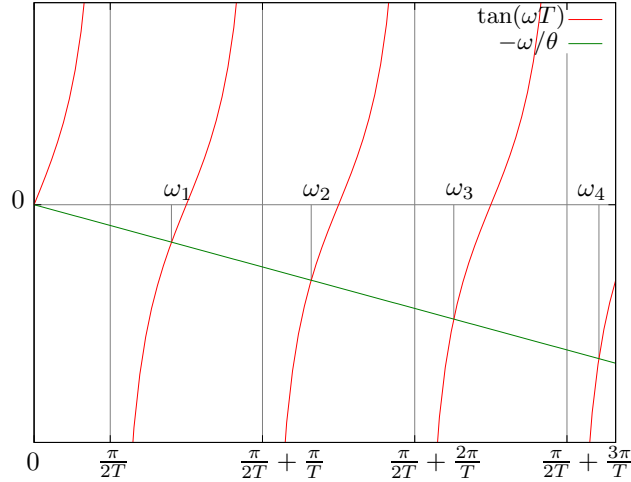

Figure 8: **(Deterministic starting point)**. Solutions of (40). (Ornstein-Uhlenbeck process starting from a fixed point X_0, $\sigma_0 = 0$.) Parameter values are $T = 3$, $\sigma = 1$ and $\theta = 3$.

A.3.2   Non-deterministic starting point

Let us assume now that $\sigma_0 \neq 0$ and consider Equation (38) again. The term $-\omega^2 \sigma_0^2 + \theta\sigma^2 - \theta^2 \sigma_0^2$ never vanishes on $(0, +\infty)$ if $\theta^2 \sigma_0^2 - \theta\sigma^2 \geq 0$.

First case: $\theta^2\sigma_0^2 - \theta\sigma^2 \geq 0$. In this case (38) gives

$$\tan(\omega T) = \frac{\omega\sigma^2}{\omega^2\sigma_0^2 + \theta^2\sigma_0^2 - \theta\sigma^2}. \tag{41}$$

Solutions of this equation are illustrated in Figure 9. We can show that for any $n \in \mathbb{N}^*$, there is a unique solution of (41) in $\left(\frac{n\pi}{T}, \frac{n\pi}{T} + \frac{\pi}{2T}\right)$. Moreover a solution lies in $\left(0, \frac{\pi}{2T}\right)$ if and only if $(\theta^2\sigma_0^2 - \theta\sigma^2)T - \sigma^2 < 0$.
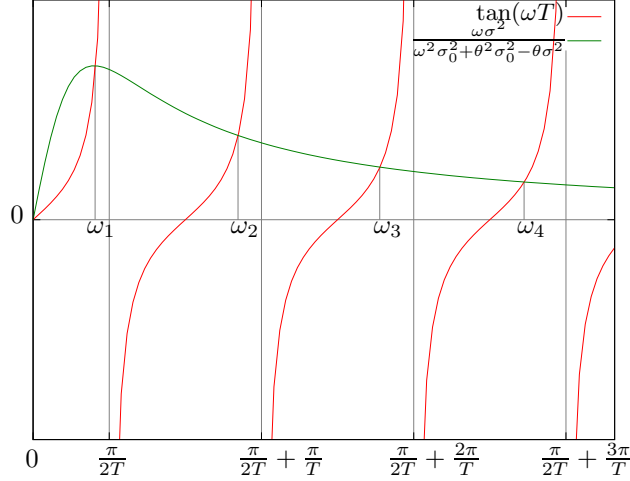

Figure 9: (**Non-deterministic starting point, $\theta^2\sigma_0^2 - \theta\sigma^2 \geq 0$**). Solutions of (40). (Ornstein-Uhlenbeck process starting from $X_0 \overset{\mathcal{L}}{\sim} \mathcal{N}(0, \sigma_0^2)$, $\sigma_0 \neq 0$.) Parameter values are $T = 3$, $\sigma = 1$, $\theta = 3$ and $\sigma_0^2 = 0.4$.

Second case: $\theta^2\sigma_0^2 - \theta\sigma^2 < 0$. Here, the term $-\omega^2\sigma_0^2 + \theta\sigma^2 - \theta^2\sigma_0^2$ vanishes for $\omega = V := \sqrt{\theta\frac{\sigma^2}{\sigma_0^2} - \theta^2}$. If V is not a solution of (38), (*i.e.* if V does not belong to $\left\{\frac{\pi}{2T} + k\frac{\pi}{T} | k \in \mathbb{N}\right\}$), no other element of this set is a solution, and everything comes again to the same Equation (41). Solutions of this equation are illustrated in Figure 10. We can show that there is a unique solution to (38) in each non-empty interval $\left(\frac{n\pi}{T}, \frac{n\pi}{T} + \frac{\pi}{2T}\right) \cap (V, \infty)$ and $\left(\frac{k\pi}{T} - \frac{\pi}{2T}, \frac{k\pi}{T} + \frac{\pi}{2T}\right) \cap (0, V)$, $k \in \mathbb{N}^*$.

In Algorithm 1, we detail the procedure for the computation of the nth eigenvalue of the Ornstein-Uhlenbeck covariance operator. The function **search**$(a, \text{left}, \text{right})$ stands for a root finding method. It returns the root of Equation (38) that is bracketed by $[\text{left}, \text{right}]$.

A.3.3 A numerical guess for ω_n.

We use $\psi(x) := \frac{\frac{4(8-\pi^2)x^3}{\pi^4} + x}{1 - \frac{4x^2}{\pi^2}}$ as an approximation of $\tan(x)$ on $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. We have $\|\tan -\psi\|_\infty^{\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)} = \frac{10-\pi^2}{2\pi} \approx 0.02075$. Plugging this into (40), we obtain

$$\theta\psi(\omega_n T + n\pi) = -\omega_n \quad n \geq 1. \tag{42}$$

This results into a polynomial equation of degree 3 having a unique (closed-form) solution $\omega_n^{guess} \in \left(\frac{n\pi}{T} - \frac{\pi}{2T}, \frac{n\pi}{T}\right)$ which can be used as a starting point for the root finding procedure.

B Closed-form expression for $R_{Y|V}$ in the cases of Brownian motion, Brownian bridge and Ornstein-Uhlenbeck processes

We use the same notation as in Section 4.2. In this Section, we derive closed-form expressions of the matrix $R_{Y|V} := (\alpha_{ij})_{1 \leq i \leq d, 0 \leq j \leq n} \in M_{d,n}(\mathbb{R})$ corresponding to the affine function $Af_{Y|V}$ defined by $\mathbb{E}[Y|V] =$
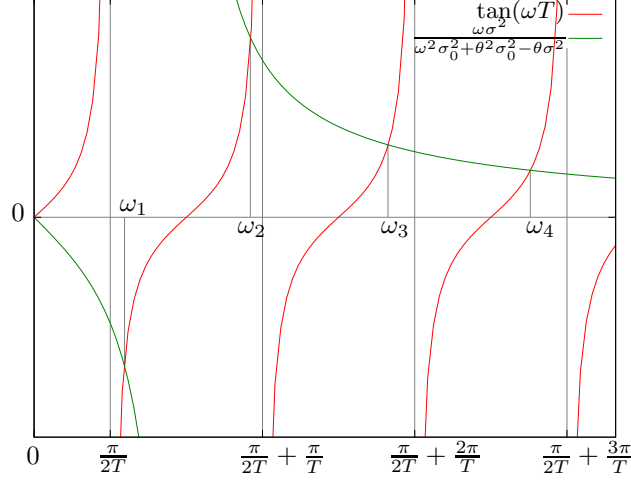
Figure 10: **(Non-deterministic starting point, $\theta^2\sigma_0^2 - \theta\sigma^2 < 0$).** Solutions of (40). (Ornstein-Uhlenbeck process starting from $X_0 \overset{\mathcal{L}}{\sim} \mathcal{N}(0, \sigma_0^2)$, $\sigma_0 \neq 0$.) Parameter values are $T = 3$, $\sigma = 1$, $\theta = 3$ and $\sigma_0^2 = 0.3$.

$Af_{Y|V}(V)$, in the cases of Brownian motion, Brownian bridge and Ornstein-Uhlenbeck processes.

In the general case, this linear least-square minimization can be performed numerically, but this preliminary stage can become time-consuming when the number of simulation dates grows.

If $t_0 = 0 \leq t_1 \leq \cdots \leq t_n = T$ is a subdivision of $[0, T]$, and X is a Gaussian Markov process, we define the affine functions f_j^i by

$$\mathbb{E}\left[\int_0^T X_s e_i^X(s)ds \,\bigg|\, X_{t_0}, \ldots, X_{t_n}\right] = \sum_{j=0}^{n-1} \mathbb{E}\left[\int_{t_j}^{t_{j+1}} X_s e_i^X(s)ds \,\bigg|\, X_{t_j}, X_{t_{j+1}}\right] =: f_j^i(X_{t_j}, X_{t_{j+1}}). \qquad (43)$$

B.1 The case of Brownian motion

Now, assuming that $X = W$ is a standard Brownian motion on $[0, T]$, using Equation (43) we obtain, for $t_j \neq t_{j+1}$, $f_j^i(x, y) = \mathbb{E}\left[\int_{t_j}^{t_{j+1}} \left(x + \frac{s - t_j}{t_{j+1} - t_j}(y - x) + Y_{s - t_j}^{B, t_{j+1} - t_j}\right) e_i^W(s)ds\right]$, where $Y_{s - t_j}^{B, t_{j+1} - t_j}$ is a standard Brownian bridge on $[t_j, t_{j+1}]$. Hence,

$$f_j^i(x, y) = x \underbrace{\left(\int_{t_j}^{t_{j+1}} \frac{t_{j+1} - s}{t_{j+1} - t_j} e_i^W(s)ds\right)}_{:=A_j^i} + y \underbrace{\left(\int_{t_j}^{t_{j+1}} \frac{s - t_j}{t_{j+1} - t_j} e_i^W(s)ds\right)}_{:=B_j^i} = xA_j^i + yB_j^i.$$

Simple algebra leads to

$$\int_{t_j}^{t_{j+1}} e_i^W(s)ds = \sqrt{\frac{2}{T}} \frac{T}{\pi\left(i - \frac{1}{2}\right)} \left(\cos\left(\pi\left(i - \frac{1}{2}\right)\frac{t_j}{T}\right) - \cos\left(\pi\left(i - \frac{1}{2}\right)\frac{t_{j+1}}{T}\right)\right),$$

and

$$\int_{t_j}^{t_{j+1}} s e_i^W(s)ds = \sqrt{\frac{2}{T}} \frac{T}{\pi\left(i - \frac{1}{2}\right)} \left(t_j \cos\left(\pi\left(i - \frac{1}{2}\right)\frac{t_j}{T}\right) - t_{j+1} \cos\left(\pi\left(i - \frac{1}{2}\right)\frac{t_{j+1}}{T}\right)\right)$$

$$+ \sqrt{\frac{2}{T}} \left(\frac{T}{\pi\left(i - \frac{1}{2}\right)}\right)^2 \left(\sin\left(\pi\left(i - \frac{1}{2}\right)\frac{t_{j+1}}{T}\right) - \sin\left(\pi\left(i - \frac{1}{2}\right)\frac{t_j}{T}\right)\right).$$

Algorithm 1 Ornstein-Uhlenbeck eigenvalue $(\theta, \sigma, \sigma_0, T, n)$

---

if $\sigma_0 = 0$ **then**   ▷ There is a unique solution ω_n of (38) in $\left(\frac{n\pi}{T} - \frac{\pi}{2T}, \frac{n\pi}{T}\right)$.

     search$\left(\omega_n, \frac{n\pi}{T} - \frac{\pi}{2T}, \frac{n\pi}{T}\right)$.

else

     if $(\theta^2 \sigma_0^2 - \theta\sigma^2) \geq 0$ **then**   ▷ The vertical asymptote in (41) lies on the left of the origin.

         if $(\theta^2 \sigma_0^2 - \theta\sigma^2)T - \sigma^2 < 0$ **then**   ▷ (38) has a unique solution in $\left(0, \frac{\pi}{2T}\right)$.

             search$\left(\omega_n, \frac{(n-1)\pi}{T}, \frac{(n-1)\pi}{T} + \frac{\pi}{2T}\right)$.

         else   ▷ The smallest positive solution ω_1 of (38) lies in $\left(\frac{\pi}{2T}, \frac{\pi}{T}\right)$.

             search$\left(\omega_n, \frac{n\pi}{T}, \frac{n\pi}{T} + \frac{\pi}{2T}\right)$.

     else   ▷ The vertical asymptote of the right-hand side of (41) lies on the right the origin.

         if $\frac{(n-1)\pi}{T} - \frac{\pi}{2T} > \sqrt{\theta\frac{\sigma^2}{\sigma_0^2} - \theta^2}$ **then**

             search$\left(\omega_n, \frac{(n-1)\pi}{T}, \frac{(n-1)\pi}{T} + \frac{\pi}{2T}\right)$.

         else if $\frac{(n+1)\pi}{T} - \frac{\pi}{2T} < \sqrt{\theta\frac{\sigma^2}{\sigma_0^2} - \theta^2}$ **then**

             search$\left(\omega_n, \frac{n\pi}{T} - \frac{\pi}{2T}, \frac{n\pi}{T}\right)$.

         else if $\frac{n\pi}{T} - \frac{\pi}{2T} < \sqrt{\theta\frac{\sigma^2}{\sigma_0^2} - \theta^2}$ and $\frac{(n+1)\pi}{T} - \frac{\pi}{2T} > \sqrt{\theta\frac{\sigma^2}{\sigma_0^2} - \theta^2}$ **then**

             search$\left(\omega_n, \frac{n\pi}{T} - \frac{\pi}{2T}, \sqrt{\theta\frac{\sigma^2}{\sigma_0^2} - \theta^2}\right)$.

         else

             search$\left(\omega_n, \sqrt{\theta\frac{\sigma^2}{\sigma_0^2} - \theta^2}, \frac{n\pi}{T} - \frac{\pi}{2T}\right)$.

   return $\lambda_n \leftarrow \frac{\sigma^2}{\omega_n^2 + \theta^2}$.

Hence $\mathbb{E}\left[\int_0^T W_s e_i^W(s)ds \Big| W_{t_1}, \ldots, W_{t_n}\right] = \sum_{j=0}^{n-1} A_j^i W_{t_j} + B_j^i W_{t_{j+1}} = \sum_{i=0}^{n} \alpha_{ij} W_{t_i}$ with, for every $1 \leq j < n$, $\alpha_{ij} = A_j^i + B_{j-1}^i$, $\alpha_{i0} = A_0^i$ and $\alpha_{in} = B_{n-1}^i$. Finally, we get the following closed-form expression for $R_{Y|V} := (\alpha_{ij})_{1 \leq i \leq d, 0 \leq j \leq n}$.

- If $t_{j-1} < t_j < t_{j+1}$,

$$\alpha_{ij} = \lambda_i^W \frac{(t_{j+1} - t_{j-1})e_i^W(t_j) - (t_{j+1} - t_j)e_i^W(t_{j-1}) - (t_j - t_{j-1})e_i^W(t_{j+1})}{(t_{j+1} - t_j)(t_j - t_{j-1})}.$$

   If $t_{j-1} = t_j < t_{j+1}$, $\alpha_{ij} = \lambda_i^W \left(\left(e_i^W\right)'(t_j) - \frac{e_i^W(t_{j+1}) - e_i^W(t_j)}{t_{j+1} - t_j} \right)$.

   If $t_{j-1} < t_j = t_{j+1}$, $\alpha_{ij} = \lambda_i^W \left(\frac{e_i^W(t_j) - e_i^W(t_{j-1})}{t_j - t_{j-1}} - \left(e_i^W\right)'(t_j) \right)$.

   If $t_{j-1} = t_j = t_{j+1}$, $\alpha_{ij} = 0$.

- $\alpha_{i0} = \begin{cases} \lambda_i^W \left(\left(e_i^W\right)'(t_0) - \frac{e_i^W(t_1) - e_i^W(t_0)}{t_1 - t_0} \right) & \text{if } t_1 \neq t_0, \\ 0 \text{ otherwise.} \end{cases}$

- $\alpha_{in} = \begin{cases} \lambda_i^W \left(\frac{e_i^W(t_n) - e_i^W(t_{n-1})}{t_n - t_{n-1}} - \left(e_i^W\right)'(t_n) \right) & \text{if } t_n \neq t_{n-1}, \\ 0 \text{ otherwise.} \end{cases}$

The equality case is useful when dealing with small time steps that make the numerical evaluation of the divided differences $(e_i^W(t_{j+1}) - e_i^W(t_j))/(t_{j+1} - t_j)$ inaccurate.

B.2 The case of Brownian bridge

If $X = B$ is a standard Brownian bridge on $[0, T]$, using Equation (43), we get for $t_j \neq t_{j+1}$, $f_j^i(x, y) = \mathbb{E}\left[\int_{t_j}^{t_{j+1}} \left(x + \frac{s-t_j}{t_{j+1}-t_j}(y-x) + \left(Y_{s-t_j}^{B, t_{j+1}-t_j}\right)\right) e_i^B(s)ds\right]$, where $Y_{s-t_j}^{B, t_{j+1}-t_j}$ is a standard Brownian bridge on $[t_j, t_{j+1}]$. Hence, very similarly to the case of Brownian motion,

$$f_j^i(x, y) = x \underbrace{\left(\int_{t_j}^{t_{j+1}} \frac{t_{j+1}-s}{t_{j+1}-t_j} e_i^B(s)ds\right)}_{:=A_j^i} + y \underbrace{\left(\int_{t_j}^{t_{j+1}} \frac{s-t_j}{t_{j+1}-t_j} e_i^B(s)ds\right)}_{:=B_j^i} = xA_j^i + yB_j^i.$$

Using that

$$\int_{t_j}^{t_{j+1}} e_i^B(s)ds = \sqrt{\frac{2}{T}} \frac{T}{\pi i} \left(\cos\left(\pi i \frac{t_j}{T}\right) - \cos\left(\pi i \frac{t_{j+1}}{T}\right)\right),$$

and

$$\int_{t_j}^{t_{j+1}} se_i^B(s)ds = \sqrt{\frac{2}{T}} \frac{T}{\pi i} \left(t_j \cos\left(\pi i \frac{t_j}{T}\right) - t_{j+1}\cos\left(\pi i \frac{t_{j+1}}{T}\right)\right) + \sqrt{\frac{2}{T}} \left(\frac{T}{\pi i}\right)^2 \left(\sin\left(\pi i \frac{t_{j+1}}{T}\right) - \sin\left(\pi i \frac{t_j}{T}\right)\right),$$

we get $\mathbb{E}\left[\int_0^T B_s e_i^B(s)ds \Big| B_{t_1}, \ldots, B_{t_n}\right] = \sum_{j=0}^{n-1} \left(A_j^i B_{t_j} + B_j^i B_{t_{j+1}}\right) = \sum_{i=0}^{n} \alpha_{ij} B_{t_i}$ where, for every $1 \leq j < n$, $\alpha_{ij} = A_j^i + B_{j-1}^i$, $\alpha_{i0} = A_0^i$ and $\alpha_{in} = B_{n-1}^i$. Moreover, we have

$$A_j^i = \lambda_i^B \left(\left(e_i^B\right)'(t_j) - \frac{e_i^B(t_{j+1}) - e_i^B(t_j)}{t_{j+1} - t_j}\right), \quad \text{and} \quad B_j^i = \lambda_i^B \left(\frac{e_i^B(t_{j+1}) - e_i^B(t_j)}{t_{j+1} - t_j} - \left(e_i^B\right)'(t_{j+1})\right).$$

Finally we obtain the following closed-form expression for $R_{Y|V} := (\alpha_{ij})_{1 \leq i \leq d, 0 \leq j \leq n}$.

- If $t_{j-1} < t_j < t_{j+1}$,

$$\alpha_{ij} = \lambda_i^B \frac{(t_{j+1} - t_{j-1})e_i^B(t_j) - (t_{j+1} - t_j)e_i^B(t_{j-1}) - (t_j - t_{j-1})e_i^B(t_{j+1})}{(t_{j+1} - t_j)(t_j - t_{j-1})}.$$

  If $t_{j-1} = t_j < t_{j+1}$, $\alpha_{ij} = \lambda_i^B \left(\left(e_i^B\right)'(t_j) - \frac{e_i^B(t_{j+1}) - e_i^B(t_j)}{t_{j+1} - t_j}\right)$.

  If $t_{j-1} < t_j = t_{j+1}$, $\alpha_{ij} = \lambda_i^B \left(\frac{e_i^B(t_j) - e_i^B(t_{j-1})}{t_j - t_{j-1}} - \left(e_i^B\right)'(t_j)\right)$.

  If $t_{j-1} = t_j = t_{j+1}$, $\alpha_{ij} = 0$.

- $\alpha_{i0} = \begin{cases} \lambda_i^B \left(\left(e_i^B\right)'(t_0) - \frac{e_i^B(t_1) - e_i^B(t_0)}{t_1 - t_0}\right) & \text{if } t_1 \neq t_0, \\ 0 \text{ otherwise.} \end{cases}$

- $\alpha_{in} = \begin{cases} \lambda_i^B \left(\frac{e_i^B(t_n) - e_i^B(t_{n-1})}{t_n - t_{n-1}} - \left(e_i^B\right)'(t_n)\right) & \text{if } t_n \neq t_{n-1}, \\ 0 \text{ otherwise.} \end{cases}$

Remark. *We obtain the same expression as for Brownian motion, where (e_n^W, λ_n^W) is replaced with (e_n^B, λ_n^B).*

B.3 The case of centered Ornstein-Uhlenbeck processes

If X is an Ornstein-Uhlenbeck process, solution of the SDE $dX_t = -\theta X_t dt + \sigma dW_t$, with $X_0 \overset{\mathcal{L}}{\sim} \mathcal{N}(0, \sigma_0^2)$ independent of W. Consider $t_0 = 0 \le t_1 \le \cdots \le t_n = T$ a subdivision of $[0, T]$. Using Equation (43) and the conditional Fubini theorem, we obtain

$$f_i^j(X_{t_j}, X_{t_{j+1}}) = \mathbb{E}\left[\int_{t_j}^{t_{j+1}} X_s e_i^{OU}(s) ds \,\middle|\, X_{t_j}, X_{t_{j+1}}\right] = \int_{t_j}^{t_{j+1}} \mathbb{E}\left[X_s \middle| X_{t_j}, X_{t_{j+1}}\right] e_i^{OU}(s) ds,$$

Assuming that $t_0 < t_1 < \cdots t_n$, we easily prove that

$$\mathbb{E}\left[X_s \middle| X_{t_j}, X_{t_{j+1}}\right] = X_{t_j} \frac{e^{\theta(t_{j+1}-s)} - e^{-\theta(t_{j+1}-s)}}{e^{\theta(t_{j+1}-t_j)} - e^{-\theta(t_{j+1}-t_j)}} + X_{t_{j+1}} \frac{e^{\theta(s-t_j)} - e^{-\theta(s-t_j)}}{e^{\theta(t_{j+1}-t_j)} - e^{-\theta(t_{j+1}-t_j)}}.$$

Hence

$$f_i^j(x, y) = x \underbrace{\left(\int_{t_j}^{t_{j+1}} \frac{e^{\theta(t_{j+1}-s)} - e^{-\theta(t_{j+1}-s)}}{e^{\theta(t_{j+1}-t_j)} - e^{-\theta(t_{j+1}-t_j)}} e_i^{OU}(s) ds\right)}_{:=A_j^i} + y \underbrace{\left(\int_{t_j}^{t_{j+1}} \frac{e^{\theta(s-t_j)} - e^{-\theta(s-t_j)}}{e^{\theta(t_{j+1}-t_j)} - e^{-\theta(t_{j+1}-t_j)}} e_i^{OU}(s) ds\right)}_{:=B_j^i},$$

where $\left(e_n^{OU}\right)_{n \ge 1}$ are the Karhunen-Loève eigenfunctions of X. Finally, we have $\mathbb{E}\left[\int_0^T X_s e_i^{OU}(s) ds \middle| X_{t_1}, \ldots, X_{t_n}\right] = \sum_{j=0}^{n-1} \left(A_j^i X_{t_j} + B_j^i X_{t_{j+1}}\right) = \sum_{i=0}^{n} \alpha_{ij} X_{t_i}$, with for every $1 \le j < n$, $\alpha_{ij} = A_j^i + B_{j-1}^i$, $\alpha_{i0} = A_0^i$ and $\alpha_{in} = B_{n-1}^i$.

B.3.1 The Ornstein-Uhlenbeck process starting from 0

In this case ($\sigma_0^2 = 0$), as proved in Appendix A, the Karhunen-Loève eigensystem is

$$e_n^{OU}(t) := \frac{1}{\sqrt{\frac{T}{2} - \frac{\sin(2\omega_n T)}{4\omega_n}}} \sin(\omega_n t), \qquad \lambda_n^{OU} := \frac{\sigma^2}{\omega_n^2 + \theta^2}, \quad n \ge 1, \tag{44}$$

where ω_n are the increasingly sorted positive solutions of $\theta \sin(\omega_n T) + \omega_n \cos(\omega_n T) = 0$. This gives

$$A_j^i = \frac{-\theta}{\theta^2 + \omega_i^2} \left(\frac{2 e_i^{OU}(t_{j+1})}{e^{\theta(t_{j+1}-t_j)} - e^{-\theta(t_{j+1}-t_j)}} - e_i^{OU}(t_j) \frac{e^{\theta(t_{j+1}-t_j)} + e^{-\theta(t_{j+1}-t_j)}}{e^{\theta(t_{j+1}-t_j)} - e^{-\theta(t_{j+1}-t_j)}}\right) + \frac{1}{\theta^2 + \omega_i^2} \left(e_i^{OU}\right)'(t_j),$$

and

$$B_j^i = \frac{\theta}{\theta^2 + \omega_i^2} \left(e_i^{OU}(t_{j+1}) \frac{e^{\theta(t_{j+1}-t_j)} + e^{-\theta(t_{j+1}-t_j)}}{e^{\theta(t_{j+1}-t_j)} - e^{-\theta(t_{j+1}-t_j)}} - \frac{2 e_i^{OU}(t_j)}{e^{\theta(t_{j+1}-t_j)} - e^{-\theta(t_{j+1}-t_j)}}\right) - \frac{1}{\theta^2 + \omega_i^2} \left(e_i^{OU}\right)'(t_{j+1}).$$

The coefficients $(\alpha_{ij})_{1 \le i \le d, 0 \le j \le n}$ are given by $\alpha_{ij} = A_j^i + B_{j-1}^i$ for $1 \le j < n$, $\alpha_{i0} = A_0^i$ and $\alpha_{in} = B_{n-1}^i$. The terms involving $\left(e_i^{OU}\right)'$ vanish. Furthermore, we can show that $\lim_{t_{j+1} \to t_j} A_j^i = 0$ and $\lim_{t_{j-1} \to t_j} B_{j-1}^i = 0$ and deduce the corresponding formula when some dates in the schedule are equal.

B.3.2 The general Ornstein-Uhlenbeck process

In this case ($\sigma_0^2 > 0$), as proved in Appendix A, the Karhunen-Loève eigensystem is given by

$$e_n^{OU}(t) := K_n \left(\omega_n \sigma_0^2 \cos(\omega_n t) + \left(\sigma^2 - \theta \sigma_0^2\right) \sin(\omega_n t)\right), \qquad \lambda_n^{OU} := \frac{\sigma^2}{\omega_n^2 + \theta^2}, \quad n \ge 1, \tag{45}$$

where ω_n are the increasingly sorted positive solutions of $\omega_n \sigma^2 \cos(\omega_n T) + (\theta \sigma^2 - \theta^2 \sigma_0^2 - \omega_n^2 \sigma_0^2) \sin(\omega_n T) = 0$, and

$$\frac{1}{K_n^2} = \frac{1}{2\omega_n} \sigma_0^2 \left(\sigma^2 - \theta \sigma_0^2\right) \left(1 - \cos(2\omega_n T)\right) + \frac{1}{2} \sigma_0^4 \omega_n^2 \left(T + \frac{\sin(2\omega_n T)}{2\omega_n}\right) + \frac{1}{2} \left(\sigma^2 - \theta \sigma_0^2\right)^2 \left(T - \frac{\sin(2\omega_n T)}{2\omega_n}\right).$$

This gives $e_n^{OU}(t) := K_n \sqrt{\omega_n^2 \sigma_0^4 + (\sigma^2 - \theta \sigma_0^2)^2} \sin(\omega_n t + \phi_n)$, with $\phi_n = \arccos\left(\frac{\sigma^2 - \theta \sigma_0^2}{\sqrt{\omega_n^2 \sigma_0^4 + (\sigma^2 - \theta \sigma_0^2)^2}}\right)$ and $\lambda_n^{OU} := \frac{\sigma^2}{\omega_n^2 + \theta^2}$, $n \geq 1$. Using that for $K \in \mathbb{R}$, $\omega \in \mathbb{R}^*$ and $(t_a, t_b) \in \mathbb{R}^2$, we get

$$\int_{t_a}^{t_b} \exp(Ks) \sin(\omega s + \phi) ds = \frac{K}{K^2 + \omega^2} \left(e^{K t_b} \sin(\omega t_b + \phi) - e^{K t_a} \sin(\omega t_a + \phi)\right)$$
$$- \frac{\omega}{K^2 + \omega^2} \left(e^{K t_b} \cos(\omega t_b + \phi) - e^{K t_a} \cos(\omega t_a + \phi)\right), \quad (46)$$

we see that the expressions for $(\alpha_{ij})_{1 \leq i \leq d, 0 \leq j \leq n}$ established in Section B.3.1 remain valid in this case.

References

[1] Vlad Bally, Gilles Pagès, and Jacques Printems. A quantization tree method for pricing and hedging multidimensional American options. *Mathematical Finance*, 15(1):119–168, 2005.

[2] Olivier Bardou, Sandrine Bouthemy, and Gilles Pagès. Optimal quantization for the pricing of swing options. *Applied Mathematical Finance*, 16(2):183–217, 2009.

[3] Mark Broadie, Paul Glasserman, and Steven Kou. A continuity correction for discrete barrier options. *Mathematical Finance*, 7:325–349, 1997.

[4] James A. Bucklew and Gary L. Wise. Multidimensional asymptotic quantization theory with rth power distortion measures. *IEEE Transactions On Information Theory*, IT-28(2 pt 1):239–247, 1982.

[5] Sylvain Corlay. A fast nearest neighbor search algorithm based on vector quantization. *Preprint*, 2011.

[6] Sylvain Corlay. Properties of the Ornstein-Uhlenbeck bridge. *Preprint*, 2013.

[7] Sylvain Corlay. Partial functional quantization and generalized bridges. *Bernoulli Journal*, 20:716–746, 2014.

[8] Allen Gersho and Robert M. Gray. *Vector quantization and signal compression*. Kluwer Academic Publishers, 1991.

[9] Paul Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer-Verlag New York, Inc., 2004.

[10] Siegfried Graf and Harald Luschgy. *Foundations of Quantization for Probability Distributions*. Springer-Verlag Berlin and Heidelberg GmbH & Co. K, 2000.

[11] Siegfried Graf, Harald Luschgy, and Gilles Pagès. Optimal quantizers for Radon random vectors in a Banach space. *J. Approx. Theory*, 144(1):27–53, 2007.

[12] Francis Hirsch and Gilles Lacombe. *Elements d'analyse fonctionnelle ; Cours et exercices avec réponses*. Dunod, 2009.

[13] Svante Janson. *Gaussian Hilbert spaces*. Cambridge university press, 1997.

[14] Benjamin Jourdain, Bernard Lapeyre, and Piergiacomo Sabino. Convenient multiple directions of stratification. *International Journal of Theoretical and Applied Finance*, 2011.

[15] Donald E. Knuth. *Art of Computer Programming, Volume 3: Sorting and Searching (2nd Edition)*. Addison-Wesley Professional, April 1998.

[16] Antoine Lejay and Victor Reutenauer. A variance reduction technique using a quantized Brownian motion as a control variate. *J. Comput. Finance*, 2008.

[17] Vincent Lemaire and Gilles Pagès. Unconstrained recursive importance sampling. *Ann. Appl. Probab.*, 20(3):1029–1067, 2010.

[18] Harald Luschgy and Gilles Pagès. Functional quantization of Gaussian processes. *Journal of Functional Analysis*, 196(2):486–531, 2002.

[19] Harald Luschgy and Gilles Pagès. Sharp asymptotics of the functional quantization problem for Gaussian processes. *Annals of Probability*, 32(2):1574–1599, 2004.

[20] Harald Luschgy and Gilles Pagès. Functional quantization of a class of Brownian diffusions: A constructive approach. *Stochastic Processes and their Applications*, 116(2):310–336, 2006.

[21] Harald Luschgy and Gilles Pagès. Functional quantization rate and mean regularity of processes with an application to Lévy processes. *Ann. Appl. Probab.*, 18(2):427–469, 2008.

[22] Gilles Pagès. A space quantization method for numerical integration. *J. Comput. Appl. Math.*, 89:1–38, 1998.

[23] Gilles Pagès, Huyên Pham, and Jacques Printems. Optimal quantization methods and applications to numerical problems in finance. In Svetlozar T. Rachev and George A. Anastassiou, editors, *Handbook on numerical methods in finance*, pages 253–297. Birkhäuser, Boston, MA, 2004.

[24] Gilles Pagès and Jacques Printems. Optimal quadratic quantization for numerics: the Gaussian case. *Monte Carlo Methods and Applications*, 9:135–166, 2003.

[25] Gilles Pagès and Jacques Printems. Functional quantization for numerics with an application to option pricing. *Monte Carlo Methods and Appl.*, 11(11):407–446, 2005.

[26] Gilles Pagès and Jacques Printems. `http://www.quantize.maths-fi.com`, 2005. "Web site devoted to optimal quantization".

[27] Eduardo S. Schwartz. The stochastic behavior of commodity prices: Implications for valuation and hedging. *Journal of Finance*, 52(3):923–73, July 1997.

[28] Paul L. Zador. Asymptotic quantization error of continuous signals and the quantization dimension. *IEEE Trans. Inform. Theory*, IT-28(2):139 –149, March 1982.

[29] Pierre Étoré, Gersende Fort, Benjamin Jourdain, and Éric Moulines. On adaptive stratification. *Ann. Oper. Res.*, 2011.

[30] Pierre Étoré and Benjamin Jourdain. Adaptive optimal allocation in stratified sampling methods. *Methodology and Computing in Applied Probability*, 2008.