Functional quantization based stratified sampling methods

Sylvain Corlay, Gilles Pagès

Functional quantization based stratified sampling methods

Sylvain Corlay* † Gilles Pagès †

August 26, 2010

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 the problem of optimal L^2-quantization of a random variable and the variance reduction that can be achieved. We first emphasize on the consistency of quantization for designing strata in stratified sampling methods in both finite dimensional and infinite dimensional frameworks. We show that this strata design has a 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 as the Brownian bridge or an Ornstein-Uhlenbeck process. We derive in detail the quantization of the Ornstein-Uhlenbeck process.

The balance between the algorithmic complexity of the simulation and the variance reduction factor has also been studied.

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

*Natixis, Equity Derivatives and Arbitrage. E-mail: sylvain.corlay@natixis.com. The authors would like to thank the members of Natixis equity derivatives quantitative R&D team for fruitful discussions.

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

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. (See [8].) In this context, quantization was a method of signal discretization. The point of interest was to design the random variable Y in order to minimize the resulting error. This led to the concept of optimal quantization.

More recently, quantization has been introduced in numerical probability to devise numerical integration methods [21] and for solving multi-dimensional stochastic control problems such as American options pricing [1] and swing options pricing [2]. Optimal quantization has many other applications and extensions in various fields like automatic classification (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 (the infinite dimensional case) [17, 22]. Stochastic processes are viewed as random variables taking values in their path spaces such as $L^2_T := L^2([0,T], dt)$.

Still the Monte-Carlo simulation remains the most common numerical methods in the field of numerical probability. One reason is that it is easy to implement in an industrial configuration. In the industry of derivative, banks implement generic Monte-Carlo frameworks for pricing numerous payoffs with a wide variety of models. Another advantage is that the Monte-Carlo method can be parallelized.

Variance reduction methods can be used to reduce dramatically the computation time of a Monte-Carlo simulation, or increase its accuracy. Main variance reduction methods are (adaptive) control variate, pre-conditioning, importance sampling and stratification [9, 16]. The problem is that these methods may strongly depend on the payoff or the model and imply specific changes in the practical implementation of the Monte-Carlo method. Thus, most institutions do not implement the most advanced methods in practice except for marginal cases.

In this paper, we point out theoretical aspects of quantization that lead to a strong link between the problem of optimal L^2-quantization of a random variable and the variance reduction that can be achieved by stratification. We emphasize on the consistency of quantization for designing strata in stratified sampling methods in both finite dimensional and infinite dimensional frameworks. 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 a uniform efficiency among the class of Lipschitz continuous functionals of the Brownian motion. The simulation cost of the conditional path is $O(n)$ where n is the number of discretization dates, like for naive Monte-Carlo simulations. In this context, this stratification based variance reduction method can be considered as a guided Monte-Carlo simulation. (See figure 6.) The method extends to any Gaussian process as soon as its Karhunen-Loève decomposition is explicitly known. So is the case for the Brownian bridge or the Ornstein-Uhlenbeck process. The special case of the Ornstein-Uhlenbeck process is derived in annex B.

One very common situation is the case of Monte-Carlo implementations that are based on multi-factor Brownian diffusions approximated by their Euler scheme. The presented method is particulary adapted to this situation. Even in the multi-dimensional case, no matter how the independent Brownian motions are correlated or used afterwards ; no matter if it is used for diffusing the underlying stock, a stochastic volatility process or an actualization factor. Functional stratification can be used as a generic variance reduction method. The point is that it is used upstream in the Monte-Carlo framework. One does not need to re-implement the whole framework but only the way it is alimented with Brownian motions. Thus quantization-based functional stratification can come along on the top of a computation procedure. In the last section, numerical tests are provided with a benchmark with a Up-In-Call pricing in the Black and Scholes model.

The paper is organized as follows. Section 1 presents the main results about optimal quantization that are required in the following. The emphasis is on the functional quantization of Gaussian processes. Section 2 presents the first historic quantization-based variance reduction method : using quantization as a control variate variable, as proposed in [22, 15]. Then section 3 outlines the links between quantization and stratification. The emphasis is on the Gaussian case. The method is specified in the functional case for Gaussian processes in section 4. We present a simulation method for the Brownian motion and other examples of Gaussian processes (as the Ornstein-Uhlenbeck process and the Brownian bridge) 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. Annex A provides the proof of a closed form expression for the Brownian motion used for functional stratification. Annex B presents the computation of the Karhunen-Loève decomposition of the Orstein-Uhlenbeck process, and the related numerical methods. A pseudo-code for the decomposition computation is provided.

1 Optimal quantization, the abstract framework

1.1 Introduction to quantization of random variables

In the following, $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space, and E is a separable reflexive Banach space. The norm on E is denoted $|\cdot|$. In the following, one will assume that the random variable are defined on $(\Omega, \mathcal{A}, \mathbb{P})$. One denotes $\mathbb{N}^* := \{1, 2, \cdots\}$.

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.

The resulting error of this discretization is the L^p-norm of $X-Y$. One wants to minimize this induced error. This gives the following minimization problem:

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

Definition (Voronoi partition)**.** *Consider $N \in \mathbb{N}^*$, $\Gamma = \{y_1, \cdots, y_N\} \subset E$ and let $C = \{C_1, \cdots, C_N\}$ be a Borel partition of E. C is a Voronoi partition associated with Γ if $\forall i \in [\![1, N]\!]$, $C_i \subset \{\xi \in E, |\xi - y_i| = \min_{j \in [\![1,N]\!]} |\xi - y_j|\}$.*

If $C = \{C_1, \cdots, C_N\}$ is a Voronoi partition associated with $\Gamma = \{y_1, \cdots, y_N\}$, it is clear that $\forall i \in [\![1, N]\!], y_i \in C_i$. C_i is called Voronoi slab associated with y_i in C and y_i is the centre of the slab C_i. One denotes $C_i = \operatorname{slab}_C(y_i)$, and for every $a \in \Gamma$, $W(a|\Gamma)$ is the closed subset of E defined by $W(a|\Gamma) = \left\{ y \in E, |y - a| = \min_{b \in \Gamma} |y - b| \right\}$.

Definition (Nearest neighbour projection)**.** *Let us consider the settled point set $\Gamma = \{y_1, \cdots, y_N\} \subset E$ and $C = \{C_1, \cdots, C_N\}$ the associated Voronoi partition. The nearest neighbour projection on Γ is the application* $\operatorname{Proj}_\Gamma := \sum_{1=1}^{N} y_i \mathbf{1}_{C_i}$.

Proposition 1.1. *Let X be an E-valued L^p random variable, and Y taking its values in the settled point set $\Gamma = \{y_1, \cdots, y_N\} \subset E$ where $N \in \mathbb{N}$. Set $\widehat{X}^\Gamma$ the random variable defined by $\widehat{X}^\Gamma := \operatorname{Proj}_\Gamma(X)$ where $\operatorname{Proj}_\Gamma$ is a nearest neighbour projection on Γ, called a Voronoi Γ-quantizer of X.*

Then we clearly have $|X - \widehat{X}^\Gamma| \leq |X - Y|$ a.s.. Hence $\|X - \widehat{X}^\Gamma\|_p \leq \|X - Y\|_p$.

As a consequence of the previous remark, solving the minimization problem (1) amounts to solving the simpler minimization problem

$$\min \left\{ \|X - \operatorname{Proj}_\Gamma(X)\|_p, \ \Gamma \subset E, \operatorname{card}(\Gamma) \leq N \right\}. \tag{2}$$

The quantity $\|X - \operatorname{Proj}_\Gamma(X)\|_p$ is called the mean L^p-quantization error. When this minimum is reached, one refers to optimal quantization.

The problem of the existence of a minimum have been investigated for decades on its numerical and theoretical aspects in the finite dimensional case [20, 10].

- For every $N \geq 1$, the 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 $\operatorname{card}(\operatorname{supp}(\mathbb{P}_X)) \geq N$. This result stands in the general abstract case of a random variable valued in a reflexive Banach space. (This has been prooved in [17].)
- If $\operatorname{card}(X(\Omega))$ is infinite, this minimum strictly decreases to 0 as N goes to infinity. The rate of convergence is ruled by theorem 1.2 in the finite dimensional case.

Theorem 1.2 (Zador). • SHARP RATE *(See [10]) Let $r > 0$ and $X \in L^{p+\eta}(\mathbb{P})$ for some $\eta > 0$. Let $\mathbb{P}_X(d\xi) = \phi(\xi)d\xi + \mu(d\xi)$ be the canonical decomposition of the distribution of X (μ and the Lebesgue measure are singular). Then, (if $\phi \not\equiv 0$),*

$$e_{N,r}(X, \mathbb{R}^d) \sim \tilde{J}_{r,d} \times \left(\int_{\mathbb{R}^d} \phi^{\frac{d}{d+r}}(u)du \right)^{\frac{1}{d}+\frac{1}{r}} \times N^{-\frac{1}{d}} \quad as \quad N \to \infty, \tag{3}$$

where $\tilde{J}_{r,d} \in (0, \infty)$.

- NON ASYMPTOTIC UPPER BOUND *(See [19]) Let $d \geq 1$. There exists $C_{d,r,\eta} \in]0, \infty[$ such that, for every $\mathbb{R}^d$-valued random vector X,*

$$\forall N \geq 1, \quad e_{N,r}(X, \mathbb{R}^d) \leq C_{d,r,\eta} \|X\|_{r+\eta} N^{-\frac{1}{d}}. \tag{4}$$

This mainly says us that $\min \left\{ \|X - \widehat{X}\|_p, \text{card}(\Gamma) \leq N \right\} \sim C_{\mathbb{P}_X,p,d} N^{-\frac{1}{d}}$. The first statement of the theorem was first prooved for distributions with compact supports by Zador in [27]. Then a first extension to general probability distributions on $\mathbb{R}^d$ is developped in [5]. The first mathematically rigorous proof can be found in [10]. The non asymptotic error bound of the second statement is prooved in [19].

In figure 1, the Voronoi partition of a random N-quantizer and an L^2-optimized N quantizer of the $\mathcal{N}(0, I_2)$ distribution are given.

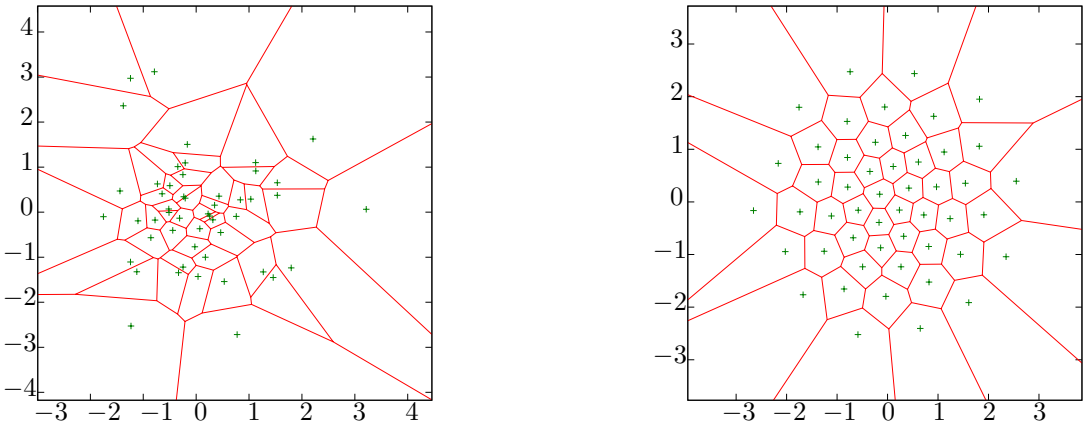

Figure 1: Voronoi partition of a random quantizer and a L^2-optimized N-quantizer of the $\mathcal{N}(0, I_2)$ distribution in $\mathbb{R}^2$. ($N = 20$).

1.2 Stationarity and centroidal Voronoi tessellations

We now assume that E is a separable Hilbert space $(H, \langle ., . \rangle_H)$.

- $\mathcal{C}_N(X)$ is the set of L^2-optimal quantizers of X of level N.
- $e_N(X)$ is the minimal quadratic distortion that can be achieved when approximating X by a quantizer of level N.

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

$$Y = \mathbb{E}[X|Y]. \tag{5}$$

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

The stationarity is a particularity of the quadratic case ($p = 2$). In other L^p cases, a similar property involving the notion of p-centre occurs. A proof of is available in [11].

A consequence is, if $Y = \text{Proj}_\Gamma(X)$ is an L^2-optimal quantizer, and $C = \{C_1, \cdots, C_n\}$ is the associated Voronoi partition, one has $\forall y \in \Gamma, y = \mathbb{E}[X|X \in \text{slab}_C(y)]$.

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

$$\begin{array}{rccl} D_N^X: & H^N & \to & \mathbb{R}_+ \\ & (x_1,\cdots,x_N) & \to & \mathbb{E}\Big[\min\limits_{1\le i\le N}|X-x_i|_H^2\Big]. \end{array}$$

The distortion function D_N^X is $|.|_H$-differentiable at N-quantizers $x\in H^N$ with pairwise distinct components and

$$\nabla D_N^X(x)=2\Big(\int_{C_i(x)}(x_i-\xi)\mathbb{P}_X(d\xi)\Big)_{1\le i\le N}=2\Big(\mathbb{E}(\hat{X}^{\Gamma(x)}-X)\mathbf{1}_{\{\hat{X}^{\Gamma(x)}=x_i\}}\Big)_{1\le i\le N}. \tag{6}$$

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

Definition (Centroidal projection)**.** *Let $C=\{C_1,\cdots,C_N\}$ be a Borel partition of H. Let us define for $1\le i\le N$, $G_i=\begin{cases}\mathbb{E}[X|X\in C_i] \text{ if } \mathbb{P}[X\in C_i]\neq 0,\\ 0 \text{ in the other case.}\end{cases}$ the centroids associated with X and C.*

The centroidal projection associated C and X is the application $\mathrm{Proj}_{C,X}: x\to\sum\limits_{i=1}^N G_i\mathbf{1}_{C_i}(x)$.

Lemma 1.5 (Huyghens, variance decomposition)**.** *Let $X\in L^2(\mathbb{P})$ be a H-valued L^2 random variable, $N\in\mathbb{N}^*$ and $C=(C_i)_{1\le i\le N}$ a Borel partition of H. Consider* $\mathrm{Proj}_{C,X}=\sum\limits_{i=1}^N G_i\mathbf{1}_{C_i}$ *the associated centroidal projection.*
Then one has,

$$\mathrm{Var}(X)=\underbrace{\mathbb{E}[|X-\mathrm{Proj}_{C,X}(X)|^2]}_{:=(1)}+\underbrace{\mathbb{E}[|\mathrm{Proj}_{C,X}(X)-\mathbb{E}[X]|^2}_{:=(2)}.$$

The variance of the probability distribution X decomposes itself into the ***intraclass inertia*** (1) *plus the* ***interclass inertia*** (2)*.*

Proof of lemma:

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}[|X-\mathrm{Proj}_{C,X}(X)+\mathrm{Proj}_{C,X}(X)-\mathbb{E}[X]|^2]\\ &= \underbrace{\mathbb{E}[|X-\mathrm{Proj}_{C,X}(X)|^2]}_{=(1)}+\underbrace{\mathbb{E}[|\mathrm{Proj}_{C,X}(X)-\mathbb{E}[X]|^2]}_{=(2)}\\ &\quad+\underbrace{2\mathbb{E}[\langle X-\mathrm{Proj}_{C,X}(X),\mathrm{Proj}_{C,X}(X)-\mathbb{E}[X]\rangle]}_{:=(3)}.\end{aligned}$$

Now $(3)=0$ since $\mathrm{Proj}_{C,X}(X)=\mathbb{E}[X|\mathrm{Proj}_{C,X}(X)]$. □

1.3 Optimal quantization and principal component analysis

1.3.1 Reduction of dimension

The aim is now the reduction of the quantization problem to finite dimensional subspaces of H. For any finite dimensional subspace U of H, let Π_U denote the orthogonal projection from H onto U.

Proposition 1.6. *Let U be a finite dimensional linear subspace of H. Then*

$$\begin{aligned}e_N(\Pi_U(X))^2\le e_N(X)^2 &\le \inf\left\{\mathbb{E}\left[\min_{a\in\Gamma}\|X-a\|^2\right],\Gamma\subset U,1\le\mathrm{card}\,\Gamma\le N\right\}\\ &= \mathbb{E}\|X-\Pi_U(X)\|^2+e_N(\Pi_U(X))^2.\end{aligned}$$

In other words, the quadratic quantization error with respect to $\Gamma\subset U$ consists of the projection error and the quantization error of the projected random variable. Let us refer to [17] for a proof.

Notation: Let $d_N(X)=\min\{\dim\mathrm{span}(\Gamma),\Gamma\in\mathcal{C}_N(X)\}$ denotes the quantization dimension of the level N of the quantization problem for X.

It follows from proposition 1.6 that

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

1.3.2 Covariance operator of a random variable

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

1. In the finite dimensional case, 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 $L^2(\mathbb{P})$-process with paths in $L^2([0,T], dt)$ a.s. and covariance function $\Gamma_X(s,t) := \mathbb{E}[X_s X_t]$ satisfying $\int_{[0,T]} \Gamma_X(s,t)ds < +\infty$. Then X can be seen as a $L^2([0,T], dt)$-valued random variable with $\mathbb{E}[\|X\|^2] < \infty$.

$$C_X y = \int_{[0,T]} y(s)\Gamma_X(s,\cdot)ds,\ y \in L^2([0,T], dt). \tag{7}$$

In [17], it is prooved that linear subspaces U of H spanned by n-stationary codebooks of Gaussian measures correspond to principal components of X, in other words, are spanned by eigenvectors of C_X corresponding to the m largest eigenvalues. Thus these subspaces correspond to the first m principal components of X.

Theorem 1.7. *Let Γ be an optimal codebook for X, $U = \mathrm{span}(\Gamma)$ and $m = \dim U$. Then $C_X(U) = U$ and $\mathbb{E}\|X - \Pi_U(X)\|^2 = \sum_{j\geq m+1} \lambda_j^X$, where $\lambda_1^X \geq \lambda_2^X \geq \cdots > 0$ are the ordered non-zero eigenvalues of C_X (written as many times as their multiplicity).*

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

We now deduce the final representation of $e_N(X)$.

$$e_N(X)^2 = \sum_{j\geq m+1} \lambda_j^X + e_N\Big(\bigotimes_{j=1}^m \mathcal{N}(0, \lambda_j^X)\Big)^2 \text{ for } m \geq d_N(X), \tag{8}$$

$$e_N(X)^2 < \sum_{j\geq m+1} \lambda_j^X + e_N\Big(\bigotimes_{j=1}^m \mathcal{N}(0, \lambda_j^X)\Big)^2 \text{ for } 1 \leq m < d_N(X). \tag{9}$$

These two equations (8) and (9) show that for the quantization of a Gaussian process X, as soon as we know its Karhunen-Loève basis $(e_n^X)_{n\in\mathbb{N}^*}$ and its eigenvalues $(\lambda_n^X)_{n\in\mathbb{N}^*}$, the problem of optimal L^2-quantization comes to the problem of the quantization of a Gaussian vector of dimension d_N.

1.4 Product quantization

Let $(e_n)_{n\in\mathbb{N}^*}$ be a Hilbertian basis of H and $I \subset \mathbb{N}^*$ is a non empty finite subset of $\mathbb{N}^*$. For every $k \in I$, consider a N_k-tuple $\Gamma^{N_k} = \{x_1^{N_k}, \cdots, x_{N_k}^{N_k}\} \subset \mathbb{R}$.

An easy way to construct a quantizer is to define the codebook Γ by the set of the points x such that for every $k \in I, \langle x, e_k\rangle \in \Gamma_k$ and for every $k \in \mathbb{N}^*\backslash I, \langle x, e_k\rangle = \mathbb{E}[\langle X, e_k\rangle]$.

The Voronoi cells associated with such a codebook are hyper-parallelepipeds.

Proposition 1.8 (Case of independent marginals)**.** *With the same notations, if one assumes that the marginals of X, $(\langle X, e_1\rangle, \langle X, e_2\rangle, \cdots)$ are independent, then one can choose for each $k \in I$ the values $\Gamma^k = \{x_1^{N_k}, \cdots, x_{N_k}^{N_k}\}$ such that $Y^k = \mathrm{Proj}_{\Gamma^k}(\langle X, e_k\rangle)$ is a stationary quantizer of $\langle X, e_k\rangle$. Then $Y = \mathrm{Proj}_\Gamma(X)$ is a stationary quantizer of X.*

This method yields a stationary quantizer with a simple projection rule.

A drawback of product quantization is that one needs to restrict to the case of independent marginals in order to preserve stationarity.

1.5 Numerical optimal quantization

Various numerical algorithms have been developed to obtain numerically an optimal N-grid with a minimal quadratic quantization error in the finite-dimensional setting. A review of these methods is available in [24]. Let us mention the Lloyd's algorithm for the quadratic case, which is the natural probabilistic counterpart of a classification algorithm due to Forgy [7].

Another algorithm is a stochastic gradient method which is suggested by the fact that the L^2-quantization distortion function is differentiable at any N-tuple having pairwise distinct components and a $\mathbb{P}_X$-negligible Voronoi tessellation boundary and has an integral representation. The algorithm is deeply investigated in [21].

Equation (6) 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, as the Gaussian distribution, the Hessian of the distortion is known and can be represented by a tridiagonal matrix. Hence, it is easy to invert and a Newton-Raphson method can be implemented. It is completely detailed in [21] in the Gaussian case. It remains the fastest way to compute L^2-optimal quantizers of one-dimensional Gaussian variables.

1.6 Quantization of Gaussian processes

1.6.1 Quantization

From now on, we will assume that X is a bi-measurable Gaussian process defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ satisfying $\mathbb{E}[|X|^2_{L^2_T}] = \int_0^T \mathbb{E}[X_s^2]ds < \infty$,.

We have seen in section 1.3 that in this context, as soon as one knows the Karhunen-Loève system $(e_n^X, \lambda_n^X)_{n\in\mathbb{N}^*}$ of the covariance operator of X, the problem of the L^2-optimal quantization of the process X comes to the quantization of the Gaussian vector $\bigotimes_{j=1}^m \mathcal{N}(0, \lambda_j^X)$. The companion parameters of the functional quantizer are easily deduced from the quantizer of $\bigotimes_{j=1}^m \mathcal{N}(0, \lambda_j^X)$ that is used.

All this is valid for any Gaussian process X, except that one needs to know its Karhunen-Loève basis. Several usual Gaussian processes have explicit Karhunen-Loève expansions, like the Brownian motion and the Brownian bridge. The Ornstein-Uhlenbeck process admits a semi-closed form for its Karhunen-Loève expansion. (The formula is derived for normalized parameters in the stationary case in [12], p.195.) In section B, the computation of Karhunen-Loève decomposition of the Ornstein-Uhlenbeck process is detailed in the general Gaussian case ($r_0 \overset{\mathcal{L}}{\sim} \mathcal{N}(m_0, \sigma_0^2)$). As far as we know, the K-L expansion of the fractional Brownian motion is not known.

Further in the paper, numerical illustrations will be given for the following cases.

1. The Brownian motion $(W_t)_{t\in[0,T]}$:

$$e_n^W(t) := \sqrt{\frac{2}{T}} \sin\Big(\pi(n-1/2)\frac{t}{T}\Big), \qquad \lambda_n^W := \Big(\frac{T}{\pi(n-1/2)}\Big)^2, \quad n \geq 1. \tag{10}$$

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

$$e_n^B(t) := \sqrt{\frac{2}{T}} \sin\Big(\pi n\frac{t}{T}\Big), \qquad \lambda_n^B := \Big(\frac{T}{\pi n}\Big)^2, \quad n \geq 1. \tag{11}$$

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

$$dr_t = -\theta r_t dt + \sigma dW_t : \tag{12}$$

$$e_n^{OU}(t) := \left(\frac{1}{\sqrt{\frac{T}{2} - \frac{\sin(2\omega_{\lambda_n}T)}{4\omega_{\lambda_n}}}}\right) \sin(\omega_{\lambda_n} t), \qquad \lambda_n^{OU} := \frac{\sigma^2}{\omega_{\lambda_n}^2 + \theta^2}, \quad n \geq 1, \tag{13}$$

where $(\omega_{\lambda_n})_{n\geq 1}$ are the strictly positive solutions of the equation

$$\theta \sin(\omega_{\lambda_n} T) + \omega_{\lambda_n} \cos(\omega_{\lambda_n} T) = 0,$$

sorted in an increasing order. (Based on results from section B.)

4. The stationary Ornstein-Uhlenbeck process on $[0,T]$. (See section B.)

On figure 2, one can see an N-optimal L^2-quantizer of the standard Brownian motion.

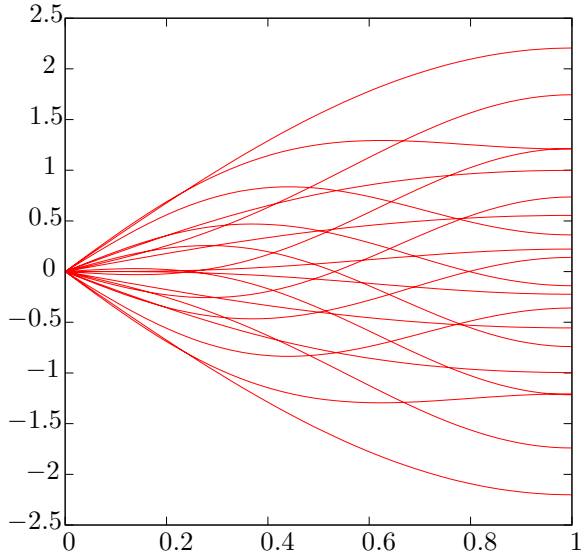

Figure 2: Optimal quantizer of a standard Brownian motion on $[0,1]$.

1.6.2 Product quantization

Thanks to equations (8) and (9), product quantization of the finite dimensional Gaussian vector $\xi \overset{\mathcal{L}}{\sim} \bigotimes_{j=1}^{m} \mathcal{N}(0, \lambda_j^X)$ yields a stationary quantizer of the process X. In this context, let us introduce the following notations:

The quantizer of X is $\widehat{X} = \sum_{n\geq 1} \sqrt{\lambda_n^X}\widehat{\xi}_n e_n^X$, where $\widehat{\xi}_n$ is an optimal N_n-quantizer of ξ_n and $N_1 \times \cdots \times N_n \leq N,\ N_1, \cdots, N_n \geq 1$. (Hence for large enough n, $N_n = 1$ so that $\widehat{\xi}_n = 0$.)

The paths of an $N_1 \times \cdots \times N_n$-quantizer χ and a multi-index $\underline{i} = \{i_1, \cdots, i_n, \cdots\}$ that produces this quantization are of the form

$$\chi_{\underline{i}} = \sum_{n\geq 1} \sqrt{\lambda_n^X} x_{i_n}^{N_n} e_n^X. \tag{14}$$

A quantizer χ defined by equation (14) is called a K-L product quantizer. Furthermore, one denotes by $\mathcal{O}_{pq}(X, N)$ the set of the K-L product quantizers of size at most N of X.

In the case of a product quantization, the counterpart of equation (8) is

$$\begin{aligned} \mathbb{E}[\min_{\underline{i}} |X - \chi_{\underline{i}}|^2] &= \sum_{n=1}^{d_x} \lambda_n^X \mathbb{E}\Big[\min_{1\leq i_n \leq N_n} |\xi_n - x_{i_n}^{(N_n)}|\Big] + \sum_{n\geq d_x+1} \lambda_n^X \\ &= \sum_{n=1}^{d_x} \lambda_n^X \mathbb{E}\Big[\min_{1\leq i_n \leq N_n} |\xi_n - x_{i_n}^{(N_n)}|\Big] + \mathbb{E}[|X|_{L_T^2}^2] - \sum_{n=1}^{d_x} \lambda_n^X. \end{aligned} \tag{15}$$

1.6.3 Product decomposition blind optimization

As a consequence, the lowest quadratic quantization error induced by a K-L-product quantizer having at most N codebooks is obtained as a solution of the minimization problem

$$\min\Big\{e(\chi), \chi \in \mathcal{O}_{pq}(X, N)\Big\}, \tag{16}$$

that is, thanks to equation (15)

$$\min\Big\{\sum_{n=1}^{d} \lambda_n^X \min_{\mathbb{R}^{N_n}} \|\xi - \widehat{\xi}^{(N_n)}\|_2^2 + \sum_{n\geq d+1} \lambda_n^X, N_1 \times \cdots \times N_n \leq N, d \geq 1\Big\}. \tag{17}$$

A solution of (16) is called an optimal K-L product quantizer.

The blind optimization procedure consists of computing the criterium for every possible decomposition $N_1 \times \cdots \times N_n \leq N$. For a given Gaussian process X, results can be kept off-line for a future use. 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` [23] for download. The blind optimization procedure is more thoroughly described in [22]. Let us remind that the optimal decomposition depends on the parameters of the Orstein-Uhlenbeck process (σ and θ in equation (12)) and the maturity.

Some values of optimal decompositions for the stationary Ornstein-Uhlenbeck process are given in table 3.

N	N_{rec}	Squared L^2 Quantization Error	N_{rec} 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

Figure 3: Record of optimal product decomposition values of the stationary centered Ornstein-Uhlenbeck process given by $dr_t = -\theta r_t dt + \sigma dW_t$ on $[0, T]$ with $\theta = 1$, $\sigma = 1$ and $T = 3$.

Proceeding in this article, we will be confronted with other similar optimization problems (with another criterium than the quadratic distortion). The blind optimization procedure will be the way to compute optimal product decomposition databases.

In figure 4, one can see examples of optimal product quantizers of the Brownian motion and the Brownian bridge on $[0, 1]$. In figure 4, one can see optimal product quantizers of the centered Ornstein-Uhlenbeck process starting from $r_0 = 0$ and a stationary Ornstein-Uhlenbeck on $[0, 3]$.

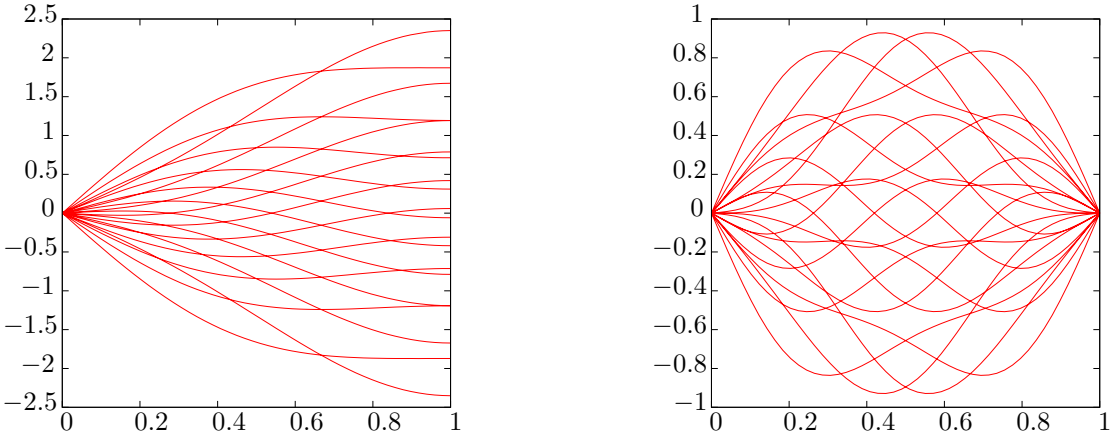

Figure 4: Optimal product quantizer of a standard Brownian motion (left) and a standard Brownian bridge (right) on $[0, 1]$.

1.6.4 Rate of decay for the quantization error

In [17], a precise link between the rate problem and Shannon-Kolmogorov's entropy of X is established. This allowed them to compute the exact rate of convergence of the minimal L^2-quantization error under rather general conditions on the eigenvalues of the covariance operator. Typical rates are $O(\log(n)^{-a})$, $a > 0$. This conditions are fulfilled by a large class of processes as the Ornstein-Uhlenbeck process and the Brownian motion.

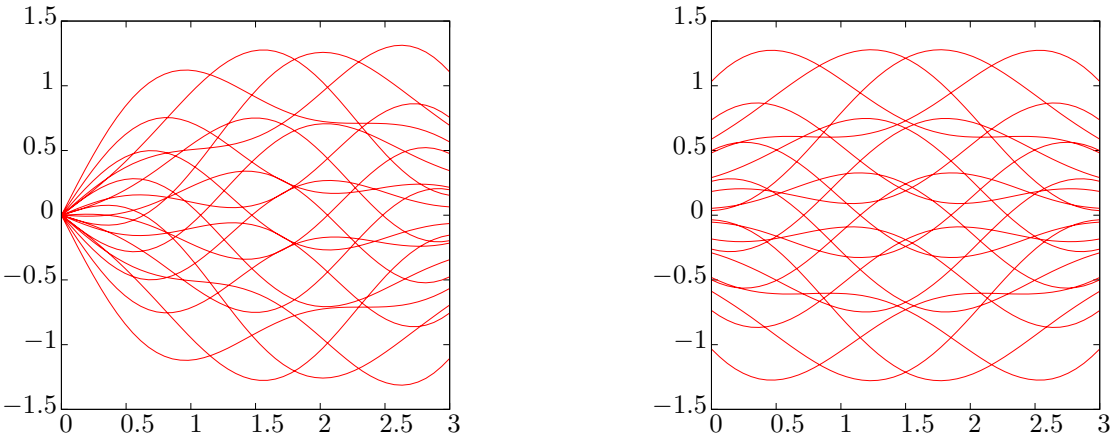

Figure 5: Optimal product quantizer of a centered Ornstein-Uhlenbeck process, starting from $r_0 = 0$ (left) and stationary (right) given by $dr_t = -r_t dt + dW_t$, on $[0, 3]$.

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

This method has been initially proposed in [22].

2.1 Quantization as a control variate variable

Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \to E$ be a square integrable random variable, condider $N \in \mathbb{N}^*$ and let $\Gamma = \{y_1, \cdots, y_N\}$ be an N-codebook. We suppose that we have access to a Γ-valued quantizer $Y = \text{Proj}(X) = \sum_{i=1}^{N} y_i \mathbf{1}_{C_i}(x)$ where $C = \{C_1, \cdots, C_N\}$ is a partition of E. At this step, we do not need Proj to be a nearest neighbour projection on Γ.

Let $F : E \to E$ be a Lipschitz continuous function such that $F(X) \in L^2(\mathbb{P})$. In order to compute $\mathbb{E}[F(X)]$, one writes:

$$\begin{aligned}\mathbb{E}[F(X)] &= \mathbb{E}\big[F(\text{Proj}(X))\big] + \mathbb{E}\big[F(X) - F(\text{Proj}(X))\big] \\ &= \underbrace{\mathbb{E}\big[F(\text{Proj}(X))\big]}_{(a)} + \underbrace{\frac{1}{M}\sum_{m=1}^{M} F(X^{(m)}) - F(\text{Proj}(X^{(m)}))}_{(b)} + R_{N,M}, \end{aligned} \tag{18}$$

where $X^{(m)}, 1 \le m \le M$ are M independent copies of X, and $R_{N,M}$ is a remainder term defined by equation (18).

Here, term (a) can be computed by quantization and term (b) can be computed by a Monte-Carlo simulation. Now

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

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

Consequently, in the d-dimensional case, if F is simply a Lipschitz function and if $(Y_N)_{N\in\mathbb{N}} = (\text{Proj}^N(X))_{N\in\mathbb{N}}$ is a rate optimal sequence of quantizers of X,

$$\|F(X) - F(\text{Proj}^N(X))\|_2 \le [F]_{Lip} \frac{C_X}{N^{1/d}}$$

and

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

Likewise, in the case of the Brownian motion, if $(\widehat{W}^N)_{N\geq 1}$ is a rate optimal sequence of product quantization of the Brownian motion, if F is simply a Lipschitz functional, then

$$\|F(W) - F(\widehat{W}^N)\|_2 \leq [F]_{Lip} \frac{C_W}{\log(N)^{1/2}}$$

and

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

2.2 Practical implementation: the problem of fast nearest neighbour search

• **The complexity of the projection:** Concerning practical implementation, one notices in equation (18) that for every step of the Monte-Carlo method, one has to compute the projection $\mathrm{Proj}(X^{(m)})$. This is the critical part of the algorithm when dealing with optimal quantization. Hence, the efficiency of the quantization as a control variate variable is conditioned by the efficiency of the projection procedure. When dealing with Voronoi quantization, this is the nearest neighbour projection.

The problem of nearest neighbour projection, also known as the post office problem [14], has been widely investigated in the area of computational geometry. It is encountered for many applications, as pattern recognition and information retrieval.

The problem has been solved near optimally for the case of low dimensions. 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 bigger query time than the obvious brute force algorithm. In fact for dimension $d > \log N$, a brute force algorithm is usually the best choice. This effect is known as the curse of dimensionality. Still, even in low dimension, fast nearest neighbour search is a critical part of the algorithm. Let us refer to [26] for a review about fast nearest neighbour search algorithms.

Concerning vector quantization, the speed of the projection can also be increased by relaxing the hypothesis within the projection on the quantizer is a nearest neighbour projection. It can be done by designing other kind of partitions of the state space.

• **The functional case:** One other drawback of the method, when dealing with the functional case is that one does not simulate the whole trajectory of the stochastic process but only its marginals at discretes dates. Hence it is not possible to compute its projection. This problem finds its solution in the simulation scheme for Gaussian processes derived in section 4.2 for the functional stratification.

A variance reduction technique using a functional quantizer of the Brownian motion as a control variate has been proposed in [15].

3 Application of quantization to stratification

3.1 A short background on stratification

The base idea of stratification is to localize the Monte-Carlo method on the element of a measurable partition of the state space of a L^2 random variable $X : (\Omega, \mathcal{A}) \to (E, \varepsilon)$.

- Let $(A_i)_{i\in I}$ be a finite ε-measurable partition of E. The sets A_i are called *strata*. Assume that the weights $p_i = \mathbb{P}(X \in A_i)$ are known for $i \in I$ and strictly positive.
- Let us define the collection of independent random variables $(X_i)_{i\in I}$ with distribution $\mathcal{L}(X|X \in A_i)$.

Remark: One assumes that one can write $X_i = \phi_i(U)$ where U is uniformly distributed on $[0,1]^{r_i}$ and $\phi : [0,1]^{r_i} \to \mathbb{R}$ is an easily computable function. (One has $r_i \in \mathbb{N} \cup \{+\infty\}$, the case $r_i = +\infty$ occurs for example in the case of the acceptance-rejection method.) This condition simply means that the random variables $X_i \overset{\mathcal{L}}{\sim} \mathcal{L}(X|X \in A_i)$ are easy to simulate on a computer.

It is a major constraint for practical implementation of stratification methods. This simulability condition usually has a strong impact on the possible design of the strata. In the following, one will come back several times on this condition.

Let $F : (E, \varepsilon) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $\mathbb{E}[F^2(X)] < +\infty$.

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

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

$$\overline{F(X)}^I_M := \sum_{i\in I} p_i \frac{1}{M_i} \sum_{k=1}^{M_i} F(X_i^k), \tag{19}$$

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

Proposition 3.1. *With the same notations:*

$$\mathrm{Var}\left(\overline{F(X)}^I_M\right) = \frac{1}{M} \sum_{i\in I} \frac{p_i^2}{q_i} \sigma^2_{F,i}, \tag{20}$$

where $\sigma^2_{F,i} = \mathrm{Var}(F(X)|X \in A_i) = \mathrm{Var}(F(X_i))\ \forall i \in I$.

Proof: Let us denote $Z_i = \frac{1}{M_i} \sum_{k=1}^{M_i} F(X_i^k)$. $(Z_i)_{i\in I}$ are independent.

One has $\overline{F(X)}^I_M = \sum_{i\in I} p_i Z_i$. Hence, by independence,

$$\mathrm{Var}(\overline{F(X)}^I_M) = \sum_{i\in I} p_i^2 \mathrm{Var}(Z_i) = \sum_{i\in I} p_i^2 \frac{1}{M_i} \mathrm{Var}(F(X_i)) = \frac{1}{M} \sum_{i\in I} \frac{p_i^2}{q_i} \sigma^2_{F,i}.$$

□

Optimizing the simulation allocation to each stratus amounts to solving the following minimization problem:

$$\min_{(q_i)\in\mathcal{P}_I} \sum_{i\in I} \frac{p_i^2}{q_i} \sigma^2_{F,i} \quad \text{where } \mathcal{P}_I = \Big\{(q_i)_{i\in I} \in \mathbb{R}_+^I \Big| \sum_{i\in I} q_i = 1\Big\}. \tag{21}$$

3.1.1 Sub-optimal choice

The first natural choice is to set

$$q_i = p_i, \qquad i \in I. \tag{22}$$

The two motivations for this choice are the facts that the weights p_i are known and because it always reduces the variance.

$$\begin{aligned}\sum_{i\in I} \tfrac{p_i^2}{q_i} \sigma^2_{F,i} &= \sum_{i\in I} p_i \sigma^2_{F,i} = \sum_{i\in I} \mathbb{E}\Big[\Big(F(X) - \mathbb{E}[F(X)|X \in A_i]\Big)^2 \mathbf{1}_{A_i}(X)\Big] \\ &= \|F(X) - \mathbb{E}[F(X)|\sigma(\{X \in A_i\}, i \in I)]\|_2^2 \\ &\le \|F(X) - \mathbb{E}[F(X)]\|_2^2 = \mathrm{Var}(F(X)).\end{aligned}$$

3.1.2 Optimal choice

The optimal choice is the solution of the constrained minimization problem (21). The Schwartz inequality yields

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

As a consequence, the solution of the minimization problem corresponds to the equality case into the Schwartz inequality. Hence the solution of the minimization problem is given by

$$q_i^* = \frac{p_i \sigma_{F,i}}{\sum_{j \in I} p_j \sigma_{F,j}}, i \in I \tag{23}$$

and the corresponding minimal variance is given by $\Big(\sum_{i \in I} p_i \sigma_{F,i} \Big)^2$.

At this point, the problem is that one does not know the local inertia $\sigma_{F,i}^2$. Still, using the fact that L^p norms are decreasing with p, one sees that

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

so that

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

In [6], Étoré and Jourdain proposed an algorithm for adaptively modifying the proportion of further drawings in each stratum, that converges to the optimal allocation. This can be used in a general framework.

In section 3.2, we will see that the problem of designing good strata, in term of variance reduction is linked with the problem of optimal quantization. Moreover, the case of quantization based strata have two other advantages:

- The weights p_i are already known, which saves us from evaluating their values during the Monte-Carlo evaluation.
- As concerns the optimal choice for the allocation parameters q_i, one shows in theorem 3.2 that weights can be chosen such that stratification has a uniform efficiency among the class of Lipschitz continuous functionals. This weights have an explicit expression in the case of quantization based stratification.

3.2 Stratification and quantization

The main drawback induced by using quantization as a control variate variable is that it requires repeated computations of projections on the quantizer. (Nearest neighbour search in the case of a Voronoi quantizer.) The point when dealing with stratification is that *one does not have to use a projection procedure.*

The critical point now is the cost of the simulation of conditional distributions $\mathcal{L}(X | X \in A_i)$, $i \in I$.

Theorem 3.2 brings together previous results about stratification and highlights the relationships with the notions of local inertia and intraclass inertia. It stresses the fact that stratification has a uniform efficiency among the class of Lipschitz continuous functionals.

Theorem 3.2 (Universal stratification)**.** *Let $A = (A_i)_{i \in I}$ be a partition (stratification) of E. (Keep in mind the notation* $\mathrm{Proj}_{A,Z}$ *for the centroidal projection of the distribution Z on a partition A, defined in definition 1.2).*

1. *For every $i \in I$, consider the local inertia of the random variable X,*

$$\sigma_i^2 = \mathbb{E}\Big[|X - \mathbb{E}[X | X \in A_i]|^2 \Big| X \in A_i \Big].$$

Then, for every Lipschitz continuous function $F : E \to \mathbb{R}$,

$$\forall i \in I, \quad \sigma_{F,i} \leq [F]_{Lip} \sigma_i \quad \text{so that} \quad \sup_{[F]_{Lip} \leq 1} \sigma_{F,i} \leq \sigma_i. \tag{24}$$

2. In the case of the sub-optimal choice (see section 3.1.1),

$$\begin{aligned}\sup_{[F]_{Lip}\leq 1}\Big(\sum_{i\in I}p_i\sigma_{F,i}^2\Big)\leq\sum_{i\in I}p_i\sigma_i^2 &= \Big\|X-\mathbb{E}[X|\sigma(\{X\in A_i\},i\in I)]\Big\|_2^2\\ &= \Big\|X-\mathrm{Proj}_{A,X}(X)\Big\|_2^2.\end{aligned}\tag{25}$$

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

$$\sup_{[F]_{Lip}\leq 1}\Big(\sum_{i\in I}p_i\sigma_{F,i}^2\Big)\leq\Big(\sum_{i\in I}p_i\sigma_i\Big)^2,\tag{26}$$

and

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

4. If one considers vector-valued Lipschitz continuous functions $F:E\to E$, then inequalities (24), (25) and (26) hold as equalities.

Proof: One has

$$\begin{aligned}\sigma_{F,i}^2 &= \mathrm{Var}(F(X)|X\in A_i)\\ &= \mathbb{E}\big[\big|F(X)-\mathbb{E}[F(X)|X\in A_i]\big|^2\big|X\in A_i\big]\\ &\leq \mathbb{E}\big[\big|F(X)-F(\mathbb{E}[X|X\in A_i])\big|^2\big|X\in A_i\big].\end{aligned}$$

Now using that F is Lipschitz, it follows that

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

Items 2 and 3 easily follow from item 1. Claim 4 is obvious by considering $F=Id_E$. □

The general case: The idea is now to use the partition $\{A_1,\cdots,A_N\}$ and the N-codebook $\Gamma=\{y_1,\cdots,y_N\}$ associated with the projection $\mathrm{Proj}(x)=\sum_{i=1}^N y_i\mathbf{1}_{A_i}(x)$.

In the case of a Voronoi quantization, this amounts to setting $I=\{1,\cdots,N\}$ and $A_i=\mathrm{slab}_A(x_i)$.

Then for every $i\in\{1,\cdots,N\}$, there exists a Borel function $\phi(x_i,.):[0,1]^q\to E$ such that $\phi(x_i,U)\overset{\mathcal{L}}{\sim}\mathcal{L}(X|X\in C_i)=\frac{\mathbf{1}_{C_i}\mathbb{P}_X(d\xi)}{\mathbb{P}[X\in C_i]}$, where $U\overset{\mathcal{L}}{\sim}\mathcal{U}([0,1]^q)$.

Now let (ξ,U) be a couple of independent random variables such that ξ has the distribution of $Y=\mathrm{Proj}(X)$ and $U\overset{\mathcal{L}}{\sim}\mathcal{U}([0,1]^q)$. Then one checks that $\phi(\xi,U)$ has the same distribution as X, so that one may assume without loss of generality that $X=\phi(\mathrm{Proj}(X),U)$ and which in turn implies that $\xi=\mathrm{Proj}(X)$ i.e.

$$X=\phi(\mathrm{Proj}(X),U),\ \text{where } U\overset{\mathcal{L}}{\sim}\mathcal{U}([0,1]^q)\text{ is independent of }\mathrm{Proj}(X).$$

In terms of implementation as mentioned above, one needs a simple form for the function ϕ (in term of computational complexity) which induces some stringent constraints on the choice of the strata.

3.3 Simulability for hyper-rectangles strata in the independent Gaussian case.

Consider a random variable $X\overset{\mathcal{L}}{\sim}\mathcal{N}(0,I_d)$, $d\geq 1$. Let $(e_1,\cdots,e_d)$ be an orthonormal basis of $E=\mathbb{R}^d$. We set $N_1,\cdots,N_d\geq 1$ the number of strata in each direction. So we consider for $1\leq i\leq d$, $-\infty=x_0^i\leq x_1^i\leq\cdots\leq x_{N_i}^i=+\infty$. The strata are

$$A_{\underline{i}}=\prod_{l=1}^d\Big\{x\in\mathbb{R}^d\text{ such that }\langle e_l,x\rangle\in[y_{i_l-1}^l,y_{i_l}^l]\Big\},\qquad\underline{i}\in\prod_{l=1}^d\{1,\cdots,N_l\}.$$

Then for every multi-index $\underline{i} \in \prod_{l=1}^{d} \{1, \cdots, N_l\}$,

$$\mathcal{L}(X|X \in A_i) = \bigotimes_{l=1}^{d} \mathcal{L}(Z|Z \in [x_{i_l-1}^l, x_{i_l}^l]), \text{ where } Z \overset{\mathcal{L}}{\sim} \mathcal{N}(0,1).$$

Then $p_{\underline{i}} = \mathbb{P}(A_i) = \prod_{k=1}^{d} \big(\mathcal{N}(x_{i_k}) - \mathcal{N}(x_{i_{k-1}})\big)$ and for $-\infty \le a \le b \le \infty$,

$$\mathcal{L}(Z|Z \in [a,b]) = \mathcal{N}^{-1}\Big(\big(\mathcal{N}(b) - \mathcal{N}(a)\big)U + \mathcal{N}(a)\Big), \qquad U \overset{\mathcal{L}}{\sim} \mathcal{U}([0,1]). \tag{27}$$

4 Functional stratification of a Gaussian process

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 an $\mathbb{R}$-valued Gaussian process on $[0,T]$. We are interested in the value of $\mathbb{E}[F(X_{t_0}, X_{t_1}, \cdots, X_{t_n})]$ 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 V the risk-neutral expectation of a path-dependent payoff of a diffusion based on X.)

What is done in this section can be easily generalized to multi-dimensional processes in the case where their coordinates are independent. (For example, when dealing with multi-factor Brownian diffusions, it does not matter how the Brownian motions are being correlated afterward.) 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 product quantizer of X. The codebook associated with this product quantizer is the set of the paths of the form

$$\chi_{\underline{i}} = \sum_{n \ge 1} \sqrt{\lambda_n^X} x_i^{(N_n)} e_n^X, \qquad \underline{i} = \{i_1, \cdots, i_n, \cdots\},$$

with the same notations 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 slab 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 (27).
- Then simulate the conditional distribution of the marginals of the Gaussian process, its first coordinates being settled.

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

$$\mathcal{L}\Big(X_{t_0}, \cdots, X_{t_n} \Big| \int_0^T X_s e_1^X ds, \int_0^T X_s e_2^X(s) ds, \cdots, \int_0^T X_s e_d^X(s) ds\Big) \tag{28}$$

where $(X_t)_{t \in [0,T]}$ is a 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.

As X is a Gaussian process, $\Big(X_{t_0}, \cdots, X_{t_n}, \int_0^T X_s e_1^X(s) ds, \cdots, \int_0^T X_s e_d^X(s) ds\Big)$ is a Gaussian vector.

Hence, 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 (28) is

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

- The conditional expectation writes $Af_{V|Y}(Y) = \mathbb{E}[V] + R_{V|Y}Y$ where $R_{V|Y} = \mathrm{cov}(V, Y)\,\mathrm{cov}(Y)^{-1}$. As $\mathrm{cov}(Y) = \left(\left(\lambda_i^X \delta_{ij}\right)\right)_{1\le i,j\le d}$, and $\mathrm{cov}(V, Y) = ((\lambda_k^X e_k^X(t_i)))_{1\le k\le d, 0\le i\le n}$, one has

$$R_{V|Y} = \left(\left(e_j^X(t_i)\right)\right)_{0\le i\le n, 1\le j\le d}. \tag{29}$$

- The covariance matrix is

$$\begin{aligned} K := \mathrm{cov}(V - \mathbb{E}[V|Y]) &= \mathbb{E}\Big[(V - R_{V|Y}Y)(V - R_{V|Y}Y)\Big] \\ &= \mathrm{cov}(V) - 2\,\mathrm{cov}(V, R_{V|Y}Y) + \mathrm{cov}(R_{V|Y}Y) \\ &= \mathrm{cov}(V) - \mathrm{cov}(R_{V|Y}Y) \end{aligned}$$

$$= \left(\left(\mathrm{cov}(V_l, V_k) - \sum_{i=1}^{d} \lambda_i e_i^X(t_l) e_i^X(t_k)\right)\right)_{0\le k,l\le n}.$$

Now, we are able to simulate according to this probability distribution.

The easiest way of doing this in the definite positive case is to compute the Cholesky factorization of the matrix K, but in this case, the simulation of a simple path requires an $n \times n$ matrix multiplication, which complexity is quadratic. This solution is not satisfactory for our purpose.

4.2 Faster simulation of conditional paths - Bayesian simulation

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

- Yet, in the context of this paper, d is the quantization dimension of the process. It is close to $\log(N)$ if N is the number of strata, and n, the number of time steps, is usually very large compared to d.
- Moreover, we make the assumption that the cost of the simulation of $(X_{t_0}, \cdots, X_{t_n})$ is $O(n)$. (So is the case for the Brownian motion, the Ornstein-Uhlenbeck process or the Brownian bridge for example.)
- 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 $Z \overset{\mathcal{L}}{\sim} \mathcal{N}(0, \mathrm{cov}(V - \mathbb{E}[V|Y]))$ 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 $\mathcal{L}(V|Y = y) = E[V|Y = y] + \mathcal{L}(Z)$.

This decomposition corresponds to the splitting of the Karhunen-Loève expansion:

$$\begin{pmatrix} V_0 \\ \vdots \\ V_n \end{pmatrix} = \underbrace{\sum_{k=1}^{d} \underbrace{\sqrt{\lambda_k^X}\xi_k}_{=Y_k} \begin{pmatrix} e_k^X(t_1) \\ \vdots \\ e_k^X(t_n) \end{pmatrix}}_{=\mathbb{E}[V|Y]} \overset{\perp\!\!\!\perp}{+} \underbrace{\sum_{l\ge d+1} \sqrt{\lambda_k}\xi_k \begin{pmatrix} e_k^X(t_1) \\ \vdots \\ e_k^X(t_n) \end{pmatrix}}_{=Z}.$$

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

$$\begin{aligned} \text{One has} \qquad \mathcal{L}(Z|V) &\overset{\mathcal{L}}{\sim} V - \mathcal{L}(\mathbb{E}[V|Y]|V) \overset{\mathcal{L}}{\sim} V - Af_{V|Y}\mathcal{L}(Y|V) \\ &\overset{\mathcal{L}}{\sim} V - Af_{V|Y}\mathcal{N}(\mathbb{E}[Y|V], \mathrm{cov}(Y - \mathbb{E}[Y|V])). \end{aligned}$$

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

$$\begin{aligned} \mathrm{cov}(Y - \mathbb{E}[Y|V]) &= \mathrm{cov}(Y) + \mathrm{cov}(\mathbb{E}[Y|V]) - 2\,\mathrm{cov}(Y, \mathbb{E}[Y|V]) \\ &= \mathrm{cov}(Y) - R_{Y|V}\,\mathrm{cov}(V)^t R_{Y|V}. \end{aligned}$$

This yields $Z = V - Af_{V|Y}(G)$ where $G \overset{\mathcal{L}}{\sim} \mathcal{N}(Af_{Y|V}(V), \mathrm{cov}(Y) - R_{Y|V}\,\mathrm{cov}(V)^t R_{Y|V})$.

Finally, the algorithm writes:

- Simulate V. *(cost of $O(n)$.)*
- Simulate $G \overset{\mathcal{L}}{\sim} \mathcal{N}(Af_{Y|V}(V), \mathrm{cov}(Y) - R_{Y|V}\,\mathrm{cov}(V)^t R_{Y|V})$ *(cost of $O(d \times d)$).*
- Compute $Z = V - Af_{V|Y}(G)$. *(cost of $O(d \times n)$).*
- The random variable $T = Af_{V|Y}(y) + Z$ satisfies $T \overset{\mathcal{L}}{\sim} \mathcal{L}(V|Y = y)$.

Let us remind the fact that the affine function $Af_{V|Y}$ is trivially defined in equation (29), because coordinates of Y are independent. Other matrices implied in this algorithm are computed prior to any Monte-Carlo simulation.

In the general case, the matrix $R_{Y|V}$ needed by the method can be computed by performing a numerical least square regression.

Still, in the case of the Brownian motion, there is a closed form for the matrix $R_{Y|V}$. If $t_j = \frac{jT}{n} = jh$, $0 \leq j \leq n$, this yields $R_{Y|V} = ((\alpha_{ij}))_{1 \leq i \leq d, 0 \leq j \leq n}$, with

- for $j \notin \{0, n\}$, $\alpha_{ij} = \lambda_i^W \frac{2e_i^W(t_j) - e_i^W(t_{j-1}) - e_i^W(t_{j+1})}{h}$,
- $\alpha_{i0} = \lambda_i^W \left(\left(e_i^{W\prime}\right)(t_0) - \frac{e_i^W(t_1) - e_i^W(t_0)}{h} \right)$,
- $\alpha_{in} = \lambda_i^W \left(\frac{e_i^W(t_n) - e_i^W(t_{n-1})}{h} - \left(e_i^W\right)'(t_n) \right)$.

The proof is available in section A. The case of a non uniform subdivision is handled as well.

Now, we have a very fast and easy way to simulate the conditional distribution (28) at our disposal.

In figures 6 and 7, we plot a few paths of the conditional distribution of various Gaussian processes knowing 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".

4.3 Blind optimization procedures

We have seen in section 3.2 that the quantity $d(\chi) = \left(\sum_{\underline{i} \in \Gamma} p_{\underline{i}} \sigma_{\underline{i}} \right)^2$ is an upper bound of the variance of the estimator, given in equation (19) in the case where the functional is 1-Lipschitz continuous. Hence one may want to minimize this criterium instead of the L^2-quantization error. This yields the minimization problem

$$\min \left\{ d(\chi), \chi \in \mathcal{O}_{pq}(X, N) \right\} \tag{30}$$

instead of the minimization problem (16).

The same kind of blind optimization procedure as in section 1.6.3 can be performed. Some values of the optimal decomposition for the standard Brownian motion are given in table 8.

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` [23] for download. When comparing all the decompositions obtained for a quantizer size smaller than 11000, one notices that in the case of the 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 $3328 - 3359$. The two criteria do not have very different values for the two decompositions. Therefore, in practice, one can use the same decomposition database for the two applications.

Nonetheless, in the case of the Brownian bridge and the Ornstein-Uhlenbeck process, one notices that the optimal decompositions for the variance and the optimal decomposition for the L^2-distortion differ more often.

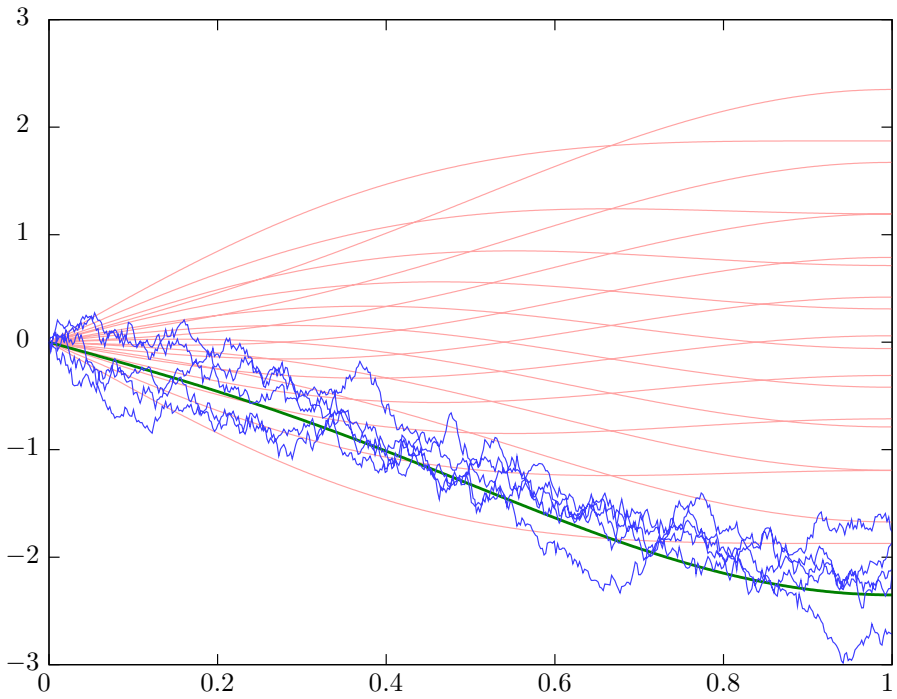

Figure 6: Plot of a few paths of the conditional distribution of the Brownian motion, knowing that its path belong to the L^2 Voronoi cell of the highlighted curve in the quantizer.

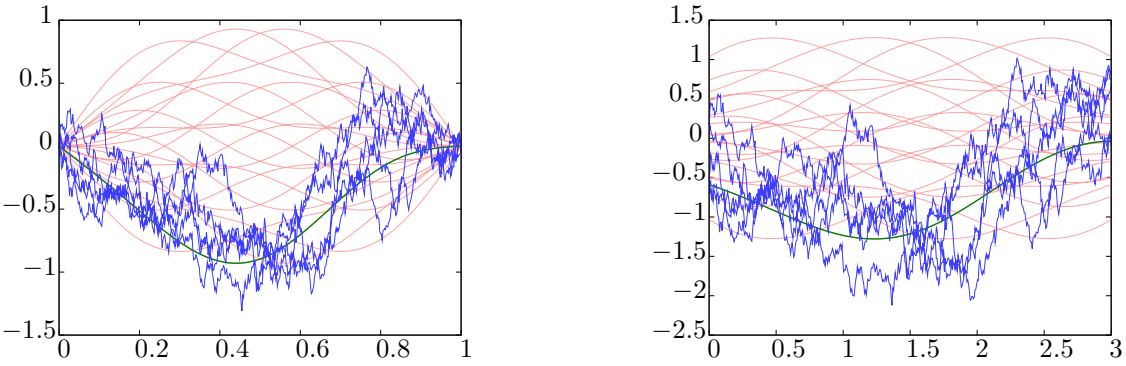

Figure 7: Plot of a few paths of the conditional distribution of the Brownian bridge (left) and the stationary Ornstein-Uhlenbeck process (right), knowing that its path belong to the L^2 Voronoi cell of the highlighted curve in the quantizer.

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

Figure 8: Record of optimal product decomposition record values of the standard Brownian motion with respect to the criterium (30).

4.4 Functional stratification of diffusions

When dealing with diffusions with respect with a (multi-dimensional) Brownian motion, most Monte-Carlo methods imply a discretization scheme as the Euler scheme [9]. In this case, one replaces the

Brownian motion by a stratified Brownian motion in the Euler scheme. This method is justified in many aspects:

1. Let us consider the diffusion

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \quad X_0 = x_0, \quad t \in [0, T], \quad \text{where} \tag{31}$$

   - $\sigma \in \mathcal{C}^1([0, T] \times \mathbb{R}, \mathbb{R})$ positive and bounded,
   - $\forall (t, x) \in [0, T] \times \mathbb{R}$, $|b(t, x)| \leq C(1 + |x|)$.

   In the one dimensional setting, as soon as the drift of the Lamperti transform of the SDE (31) is Lipschitz continuous, it is prooved in [18] that the unique strong solution of (31), seen as a functional of the underlying Brownian motion is $\|\cdot\|_p$-Lipschitz continuous.

   Hence one stands in the case of a Lipschitz continuous functional where one can use the results of section 3.2 about universal stratification.

2. The function $(W_{t_0}, \cdots, W_{t_n}) \to \left(W_{t_1} - W_{t_0}, \cdots, W_{t_n} - W_{t_{n-1}}\right)$ that maps the marginals of the Brownian motion to the corresponding Euler scheme is linear from $\mathbb{R}^{n+1}$ to $\mathbb{R}^n$ and thus Lipschitz continuous as well.

In next section (5), numerical examples are given when replacing the Brownian motion by a stratified Brownian motion in the Euler scheme.

5 Application to option pricing

Now, we are able to simulate the conditional distribution of a Gaussian process, given one of its Voronoi cell in a product quantizer. One condition is to know an orthonormal Hilbert basis that diagonalizes its covariance operator. The cases of the Brownian motion, the Brownian bridge and the Ornstein-Uhlenbeck process have been handled.

The particular case of the Brownian motion allows to use functional stratification as a generic variance reduction method for the case of functionals of Brownian diffusions. Even in the multi-dimensional case, no matter how the independent Brownian motions are correlated or used afterwards ; no matter if it is used for diffusing the underlying stock, a stochastic volatility process or an actualization factor. It can be used as a variance reduction method.

Hence, this is a very interesting variance reduction method to be used in an industrial way, independently of the path-dependent payoff or the model (as soon as it uses Brownian diffusions or one of the other proposed Gaussian processes). Users do not have to set up complicated adjustments when using it.

In the following of this section, the method is used to illustrate its performance on simple one dimensional cases. One begins with the case of a continuous time Up-In Call in the Black and Scholes model, for which a closed formula is known, and used as a Benchmark.

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

Here, one benchmarks the numerical method for a path dependent option in a case where a theoretical value is known : a barrier option in the Black and Scholes Model.

For the sake of simplicity, consider a log-normal Black and Scholes diffusion with no drift (no interest rate and no dividend).

One has a closed form for the continuous barrier option. A numerical correction proposed by Broadie and Glasserman [4] is done to get the closed-form price to be compared to. The number of Monte-Carlo simulations is 100000 in every case.

One prices an Up-In-Call 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 9 when using the method with 20 stratas and table 10 when using the method with 100 stratas. In this tables, the first column correspond to the Broadie and

Glasserman's closed form proxy. The second one corresponds to a simple Monte-Carlo estimator. The last three columns correspond to a stratified sampling estimator with different simulation allocation for each strata.

The "sub-optimal weights" column stands for the allocation budget of equation (22). The "Lip.-optimal weights" column stand for the "universal stratification" budget allocation proposed in theorem 3.2. Both these two case have explicit allocation rules. Last column, "Optimal weights" corresponds to an estimation of the optimal budget allocation given in expression (23).

Parameters	Broadie & Glasserman's proxy	Simple Estimator	Strat. Estimator sub-optimal 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

Figure 9: Numerical results for the Up In Call option, with 20 stratas.

Parameters	Broadie & Glasserman's proxy	Simple Estimator	Strat. Estimator sub-optimal 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

Figure 10: Numerical results for the Up In Call option, with 100 stratas.

5.2 Test with an auto-call pricing in the CEV model.

Here, we stand in the case were the stock follows a CEV model with no drift

$$dS_t = \sigma S_t^{\frac{\beta}{2}} dW_t, \quad 0 \leq \beta < 2.$$

The simulation scheme that is used here is a Euler scheme on $\ln(S_t)$. One has

$$d\ln(S_t) = -\frac{\sigma^2}{2} S_t^{\beta-2} dt + \sigma S_t^{\frac{\beta}{2}-1} dW_t.$$

Let us remind the fact that there are closed-form expressions for vanilla option pricing in this model that can be expressed as a function of the noncentral chi-square distribution [13]. A first test of consistency for the method was to check that we could find the same price when performing such a Monte-Carlo simulation. The tested path-dependent payoff that we consider here is the so-called "auto-call" payoff.

Description of the auto-call payoff:

S_t is the stock price at time t and $t_1 < \cdots < t_n = T$ is a schedule of observation dates. K and H, the "strike" and the "barrier" are two settled values with which S will be compared to. P denotes the "nominal", and C a bound.

At the first date t_1 of the schedule, if $S_{t_1} > K$, the holder of the option gets $(1+C)P$ and the product stops. If $S_{t_1} \leq K$, one waits until the second date of the schedule. If $S_{t_2} > K$, the holder gets $(1+C)P$ and the product stops. And so on... If S_t does not reach K until the last date $t_n = T$.

At t_n, if $S_T > K$, the holder gets $(1+C)P$. If $B < S_T \leq K$, the holder gets P and if $S_T \leq B$, he gets $P\frac{S_T}{K}$.

The numerical results are reported in table 11 when using the 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 one performs 100000 Monte-Carlo simulations in every case.

Number of strata	Simple Estimator	Strat. Estimator sub-optimal 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

Figure 11: Numerical results for the auto-call option in the CEV model, with 20 and 50 stratas.

5.3 Test with an Asian option pricing in the one-factor Schwartz's model.

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

$$dS = \theta(\alpha - \ln S)Sdt + \sigma S dW_t, \tag{32}$$

under the risk neutral probability.

The stochastic process $X = \ln(S)$ is an Ornstein-Uhlenbeck process:

$$dX = \theta(\mu - X)dt + \sigma dW_t \quad \text{with } \mu = \alpha - \frac{\sigma^2}{2\theta}. \tag{33}$$

This model, proposed by Schwartz in [25] is an example of stochastic behaviour of commodity prices that takes into account mean reversion. Such exponentials of Ornstein-Uhlenbeck processes are very common in commodity derivatives models. One particularity in these markets is that the spot is not directly observed. Derivatives mostly rely on futures of the considered commodity. Still, one takes this one factor "toy" model as a simple case study for our variance reduction method.

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

One uses the stratified estimator with the Ornstein-Uhlenbeck process. Optimal product decompositions for the criteria (30) are used and available in table 12 where the numerical results are reported.

The numerical parameters are $S_0 = 100$, $\theta = 0.3$, $\alpha = \ln(110)$, $\sigma = 0.3$ and $K = 100$. One performs 100000 Monte-Carlo simulations in every case. The observation dates are $\left(i\frac{T}{n}\right)_{i=\{0,\cdots,n\}}$ with $T = 3$ and $n = 36$.

Number of strata and product decomposition	Simple Estimator	Strat. Estimator sub-optimal weights	Strat. Estimator Lip.-optimal weights	Strat. Estimator Optimal weights
20 $20 = 10 \times 2$	9.8485 [9.7508, 9.9462] Var = 248.3156	9.8867 [9.8632, 9.9102] Var = 14.3132	9.8848 [9.8624, 9.9073] Var = 13.1090	9.8846 [9.8695, 9.8997] Var = 5.9547
50 $48 = 10 \times 5$	9.8485 [9.7508, 9.9462] Var = 248.3156	9.8835 [9.8608, 9.9061] Var = 13.4003	9.87862 [9.8555, 9.8983] Var = 11.8787	9.8845 [9.8702, 9.8987] Var = 5.2949
100 $100 = 10 \times 5 \times 2$	9.8485 [9.7508, 9.9462] Var = 248.3156	9.8883 [9.8661, 9.9105] Var = 12.8434	9.8924 [9.8716, 9.9133] Var = 11.3508	9.8844 [9.8706, 9.8782] Var = 4.9664

Figure 12: Numerical results for the Asian option in the Schwartz's model, with 20 and 100 stratas.

To perform this computation, one had to use a non-centered Ornstein-Uhlenbeck quantizer. Building such a quantizer is a straightforward extension of the centered case. As showed in section B, if r is an Ornstein-Uhlenbeck process on $[0, T]$ following the dynamic $dr_t = \theta(\mu - r_t)dt + \sigma dW_t$, $r_0 \sim \mathcal{N}(m_0, \sigma_0^2)$, with non zero values of μ and m_0, one has

$$X_t = \underbrace{m_0 e^{-\theta t} + \mu(1 - e^{-\theta t})}_{\text{(1)=non stochastic path}} + \begin{pmatrix} \text{centered Ornstein-Uhlenbeck process} \\ \text{corresponding to } m_0 = \mu = 0 \end{pmatrix}. \tag{34}$$

Hence, one only needs to add the expectation (1) to the centered optimal (product) quantizer to get an optimal (product) quantizer for the non centered case. An example of such a non centered Ornstein-Uhlenbeck product quantizer is available in figure 13.

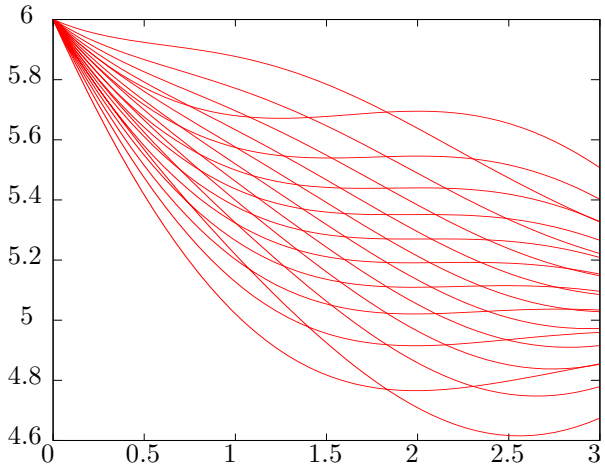

Figure 13: Functional 10×2-product quantizer of an Ornstein-Uhlenbeck process starting from $r_0 = 6$ defined by the diffusion $dr_t = \theta(\mu - r_t)dt + \sigma dW_t$ with $\mu = 5$, $\sigma = 0.3$ and $\theta = 0.8$ on $[0, 3]$.

5.4 Commentaries on the numerical results

In every tested case, one notices that the quantization-based stratified sampling method reduces noticeably the variance of the Monte-Carlo estimator. The "universal stratification" allocation proposed in theorem 3.2 overcomes the sub-optimal weight allocation. Still in the case of the auto-call, its advantage is not very perceptible.

Moreover, the "optimal allocation" estimation yields a very good variance reduction factor. This suggests to implement either a simple prior rough estimation of the optimal allocation or a more sophisticated algorithm as the one proposed in [6] by Étoré and Jourdain.

A Special case of the Brownian motion for $R_{Y|V}$ computation.

In this section, one uses the same notations as in section 4.2. We give the closed form of the matrix $R_{Y|V} := ((\alpha_{ij}))_{1\leq i\leq d, 0\leq j\leq n} \in M_{d,n}(\mathbb{R})$ which corresponds to the affine function $Af_{Y|V}$ defined by $\mathbb{E}[Y|V] = Af_{Y|V}(V)$.

Consider $t_0 = 0 \leq t_1 \leq \cdots \leq t_n = T$ a subdivision of $[0, T]$.

$$\mathbb{E}\Big[\int_0^T W_s e_i^W(s)ds | W_{t_0}, \cdots, W_{t_n}\Big] = \sum_{j=0}^{n-1} \underbrace{\mathbb{E}\Big[\int_{t_j}^{t_{j+1}} W_s e_i^W(s)ds \Big| W_{t_j}, W_{t_{j+1}}\Big]}_{=f_j^i(W_{t_j}, W_{t_{j+1}})},$$

where f_j^i is an affine function.

If $t_j \neq t_{j+1}$, $f_j^i(x, y) = \mathbb{E}\Big[\int_{t_j}^{t_{j+1}} \Big(x + \frac{s-t_j}{t_{j+1}-t_j}(y-x) + (Y_{s-t_j}^{B,t_{j+1}-t_j})\Big) e_i^W(s)ds\Big]$ (where $Y_{s-t_j}^{B,t_{j+1}-t_j}$ is a standard Brownian bridge)

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

Simple computations lead to:

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

and

$$\begin{aligned}\int_{t_j}^{t_{j+1}} s e_i^W(s)ds &= \sqrt{\tfrac{2}{T}}\tfrac{T}{\pi(i-\frac{1}{2})}\Big(t_j\cos(\pi(i-\tfrac{1}{2})\tfrac{t_j}{T}) - t_{j+1}\cos(\pi(i-\tfrac{1}{2})\tfrac{t_{j+1}}{T})\Big)\\ &\quad+\sqrt{\tfrac{2}{T}}\Big(\tfrac{T}{\pi(i-\frac{1}{2})}\Big)^2\Big(\sin(\pi(i-\tfrac{1}{2})\tfrac{t_{j+1}}{T}) - \sin(\pi(i-\tfrac{1}{2})\tfrac{t_j}{T}))\Big).\end{aligned}$$

Hence $\mathbb{E}\Big[\int_0^T W_s e_i^W(s)ds|W_{t_1},\cdots,W_{t_n}\Big] = \sum_{j=0}^{n-1} A_j^i W_{t_j} + B_j^i W_{t_{j+1}} = \sum_{i=0}^{n}\alpha_{ij}W_{t_i}$ with, for every $1\le j < n$, $\alpha_{ij} = A_j^i + B_{j-1}^i$, $\alpha_{i0} = A_0^i$ and $\alpha_{in} = B_{n-1}^i$.

Finally one gets the following closed forms for $R_{Y|V} := ((\alpha_{ij}))_{1\le i\le d, 0\le j\le n}$.

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

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

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

- $\alpha_{i0} = A_0^i = \begin{cases}\lambda_i^W\left(e_i^{W\prime}(t_0) - \frac{e_i^W(t_1)-e_i^W(t_0)}{t_1-t_0}\right) & \text{if } t_1\neq t_0,\\ 0 \text{ in the other case.}\end{cases}$
- $\alpha_{in} = B_{n-1}^i = \begin{cases}\lambda_i^W\left(\frac{e_i^W(t_n)-e_i^W(t_{n-1})}{t_n-t_{n-1}} - e_i^{W\prime}(t_n)\right) & \text{if } t_n\neq t_{n-1},\\ 0 \text{ in the other case.}\end{cases}$

(The equality cases are useful when dealing with time steps that make the numerical evaluation of $e_i^W(t_{j+1}) - e_i^W(t_j)$ to close to zero.)

B Computation of the Karhunen-Loève decomposition the Ornstein-Uhlenbeck process

In this section, one details the Karhunen-Loève decomposition of the Ornstein-Uhlenbeck process. Proposition B.3 brings the results together. Section B.3 presents the numerical method for computing this decomposition.

B.1 The Ornstein-Uhlenbeck process

The Ornstein-Uhlenbeck process is defined by the SDE

$$dr_t = \theta(\mu - r_t)dt + \sigma dW_t, \qquad \text{with } \sigma\ge 0 \text{ and } \theta > 0. \tag{35}$$

The equation is solved by applying Itô's formula to the process $U_t := r_t e^{\theta t}$. One gets

$$r_t = r_0 e^{-\theta t} + \mu(1-e^{-\theta t}) + \int_0^t \sigma e^{\theta(s-t)}dW_s. \tag{36}$$

If one assumes that r_0 is Gaussian ($r_0 \overset{\mathcal{L}}{\sim} \mathcal{N}(m_0,\sigma_0^2)$) and is independent from W, the process $(r_t)_{t>0}$ is Gaussian. One has $\mathbb{E}[r_t] = m_0 e^{-\theta t} + \mu(1-e^{-\theta t})$ and $\mathrm{cov}(r_s,r_t) = \frac{\sigma^2}{2\theta}e^{-\theta(s+t)}(e^{2\theta\min(s,t)}-1) + \sigma_0^2 e^{-\theta(s+t)}$.

Moreover $\lim_{t\to\infty}\mathrm{Var}(r_t) = \frac{\sigma^2}{2\theta}$ (the long term variance). If the initial variance σ_0^2 is equal to long term variance $\frac{\sigma^2}{2\theta}$, the process is stationary and the covariance writes $\mathrm{cov}(r_s,r_t) = \frac{\sigma^2}{2\theta}e^{-\theta|s-t|}$.

The total variance of the process on $[0,T]$ is

$$\|r_2\|_2^2 = \int_0^T \mathrm{Var}(r_s)ds = \frac{\sigma^2 T}{2\theta} + \Big(\sigma_0^2 - \frac{\sigma^2}{2\theta}\Big)\Big(\frac{1}{2\theta} - \frac{e^{-2\theta T}}{2\theta}\Big).$$

B.2 The Ornstein-Uhlenbeck covariance operator:

The Ornstein-Uhlenbeck covariance operator is given by

$$T^{OU}f(s) = \int_0^T \frac{\sigma^2}{2\theta} e^{-\theta(s+t)} \Big(e^{2\theta \min(s,t)} - 1\Big) f(s)ds + \int_0^T \sigma_0^2 e^{-\theta(s+t)} f(s)ds. \tag{37}$$

Computing the Karhunen-Loève decomposition of the Ornstein-Uhlenbeck process

T^{OU} is a compact Hermitian positive operator on the separable Hilbert space $L^2([0,T])$. Hence there is an orthonormal basis of V consisting of eigenvectors of T^{OU} and each eigenvalue is real and strictly positive. Moreover $\|T^{OU}\|^2 \leq \frac{\sigma^2 T}{2\theta} + \frac{\sigma^2}{4\theta^2}\Big(e^{-2\theta T} - 1\Big)$. One has

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

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

$$g'' - \theta^2 g = -\sigma^2 f, \tag{38}$$

with

$$\sigma_0^2 g'(0) = \Big(\sigma^2 - \theta\sigma_0^2\Big) g(0) \qquad \text{and} \qquad g'(T) = -\theta g(T). \tag{39}$$

Proof:

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

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

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

One gets $g''(t) = \theta^2 g(t) - \sigma^2 f(t)$. Moreover, equation (39) comes when identifying expressions with $t = 0$ and $t = T$. □

Proposition B.2. *Conversely, if $g \in C^2([0,T])$ and if f satisfies equations (38) and (39) then $g = T^{OU}f$.*

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

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

An integration by parts yields

$$\begin{aligned} T^{OU}g'' &= -\sigma_0^2 g'(0)e^{-\theta t} - \sigma^2 g(t) + \tfrac{\sigma^2}{2} g(0)e^{-\theta t} - \Big(\theta\sigma_0^2 - \tfrac{\sigma^2}{2}\Big) g(0)e^{-\theta t} + \theta^2 T^{OU}g(t) \\ &= -\sigma^2 g(t) + \theta^2 T^{OU}g(t) \quad \text{thanks to (39).} \end{aligned}$$

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

$$\lambda g'' + (\sigma^2 - \lambda\theta^2) g = 0. \tag{40}$$

□

Hence the solution of the ordinary differential equation (40) 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 (39) yields $\omega B\sigma_0^2 = (\sigma^2 - \theta\sigma_0^2)A$. Hence, function $g(x)$ writes

$$g(t) = K\Big(\omega\sigma_0^2\cos(\omega t) + (\sigma^2 - \theta\sigma_0^2)\sin(\omega t)\Big).$$

Hence $g'(T) = -\theta g(T)$ yields

$$\omega\sigma^2\cos(\omega T) + \big[-\omega^2\sigma_0^2 + \theta\sigma^2 - \theta^2\sigma_0^2\big]\sin(\omega T) = 0. \tag{41}$$

Conversely, by the same computation, one sees that $\lambda_n \in]0, \|T^{OU}\|_2]$ is an eigenvalue of T^{OU} if and only if equation (41) is fulfilled.

Proposition B.3. *Finally, if one knows the sorted increasing sequence (ω_n) of the strictly positive solutions of equation (41), the Karhunen-Loève basis $(\lambda_n^{OU}, e_n^{OU})$ of the Ornstein-Uhlenbeck covariance operator T^{OU} are given by:*

- $\lambda_n^{OU} = \frac{\sigma^2}{\omega_n^2+\theta^2}$, *and*
- $e_n^{OU}(t) = K_n\Big(\omega_n\sigma_0^2\cos(\omega_n t) + (\sigma^2 - \theta\sigma_0^2)\sin(\omega_n t)\Big)$, *where K_n is the normalization constant.*

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

$$1/K_n^2 = \frac{1}{2w_n}\sigma_0^2(\sigma^2 - \theta\sigma_0^2)[1-\cos(2\omega_n T)] + \frac{1}{2}\sigma_0^4\omega_n^2\Big(T + \frac{1}{2\omega_n}\sin(2\omega_n T)\Big) + \frac{1}{2}(\sigma^2 - \theta\sigma_0^2)^2\Big(T - \frac{1}{2\omega_n}\sin(2\omega_n T)\Big). \quad (42)$$

Case of a deterministic start point: In this case ($\sigma_0 = 0$), one has

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

Stationary case: In the stationary case, $\sigma_0^2 = \frac{\sigma^2}{2\theta}$, one has

$$e_n^{OU}(t) = C_n\Big(\omega_n\cos(\omega_n t) + \theta\sin(\omega_n t)\Big),$$

where C_n is the normalization constant. C_n is given by

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

B.3 Numerical computation of the Karhunen-Loève decomposition of the Ornstein-Uhlenbeck process

As we have seen in the previous section, everything comes to evaluate numerically the strictly positive solutions of equation (41).

B.3.1 Deterministic start point

In this case, ($\sigma_0 = 0$), one can check that elements of $\{\frac{\pi}{2} + k\frac{\pi}{T} | k \in \mathbb{N}\}$ are not solutions of equation (41), thus the equation comes to

$$\theta\tan(\omega T) = -\omega. \quad (43)$$

The case where $\theta = 0$ comes to the case of the Brownian motion, hence one assumes that $\theta \neq 0$. Solutions of this equation are illustrated in figure 14. One can easily show that a unique solution w_n lies in each interval $\Big]\frac{n\pi}{T} - \frac{\pi}{2T}, \frac{n\pi}{T}\Big[$, for $n \in \mathbb{N}^*$.

$$\lim_{n\to\infty}\omega_n - \Big(\frac{n\pi}{T} - \frac{\pi}{2T}\Big) = 0.$$

B.3.2 Non deterministic start point

Let us assume now that $\sigma_0 \neq 0$ and consider equation (41) 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: $\theta^2\sigma_0^2 - \theta\sigma^2 \geq 0$**.**

Here, everything comes to the equation

$$\tan(\omega T) = \frac{\omega\sigma^2}{\omega^2\sigma_0^2 + \theta^2\sigma_0^2 - \theta\sigma^2}. \quad (44)$$

Solutions of this equation are illustrated in figure 15.

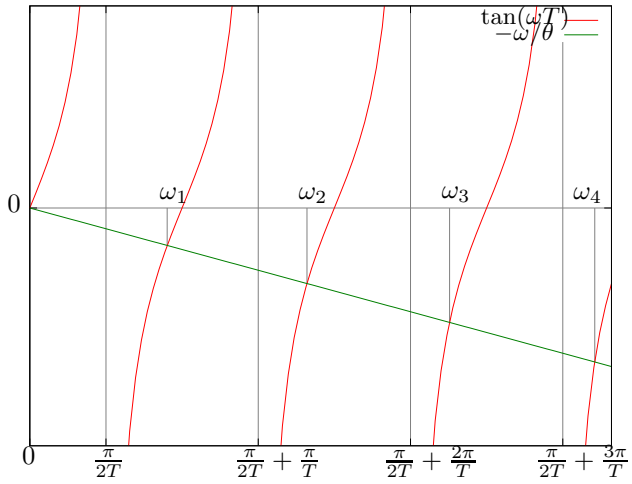

Figure 14: **(Deterministic start point)**. Solutions of equation (43). (Ornstein-Uhlenbeck process starting from a determined point r_0, $\sigma_0 = 0$.) Values for this figure are $T = 3$, $\sigma = 1$ and $\theta = 3$.

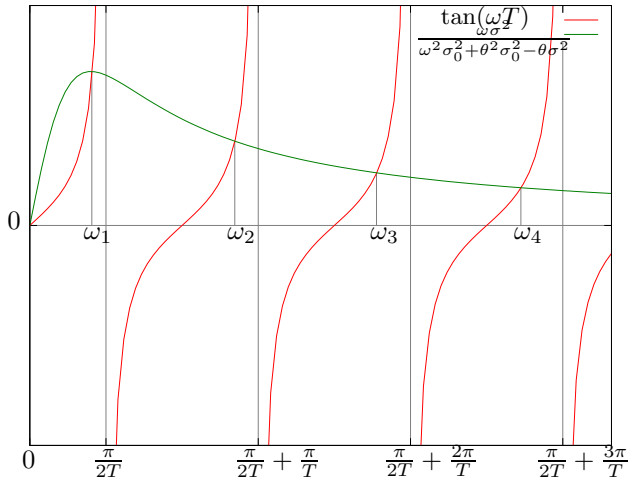

Figure 15: **(Non deterministic start point, $\theta^2\sigma_0^2 - \theta\sigma^2 \geq 0$)**. Solutions of equation (43). (Ornstein-Uhlenbeck process starting from $r_0 \overset{\mathcal{L}}{\sim} \mathcal{N}(0, \sigma_0^2)$, $\sigma_0 \neq 0$.) Values for this figure are $T = 3$, $\sigma = 1$, $\theta = 3$ and $\sigma_0^2 = 0.4$.

One can easily show that $\forall n \in \mathbb{N}^*, \exists! \omega \in \left]\frac{n\pi}{T}, \frac{n\pi}{T} + \frac{\pi}{2T}\right[$, which is solution of equation (41). Moreover a solution lies in $]0, \frac{\pi}{2T}[$ if and only if $(\theta^2\sigma_0^2 - \theta\sigma^2)T - \sigma^2 < 0$.

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 equation (41), (i.e. if V does not belong to $\{\frac{\pi}{2T} + k\frac{\pi}{T} | k \in \mathbb{N}\}$), no other value of this set is a solution, and everything comes again to the same equation (44).

Solutions of this equation are illustrated in figure 16. One can then easily show that $\forall n \in \mathbb{N}^* \cap]V, +\infty[, \exists! \omega \in \left]\frac{n\pi}{T}, \frac{n\pi}{T} + \frac{\pi}{2T}\right[$, which is solution of equation (41). Moreover, in every non empty interval $\left]\frac{k\pi}{T} - \frac{\pi}{2T}, \frac{k\pi}{T} + \frac{\pi}{2T}\right[\cap \left]0, V\right[$, $k \in \mathbb{N}^*$ there is another solution of the equation.

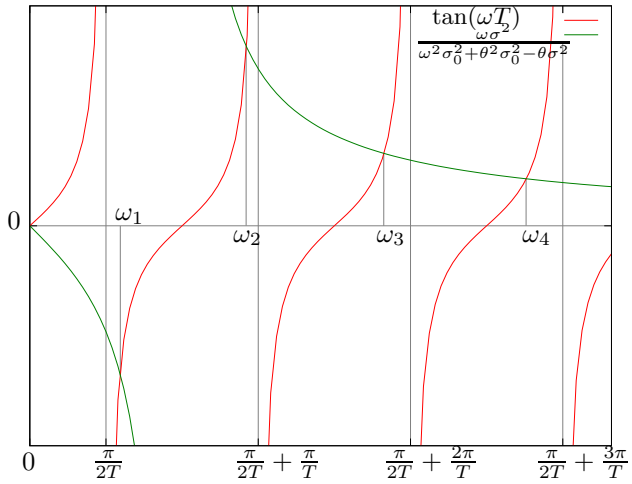

Figure 16: **(Non deterministic start point, $\theta^2\sigma_0^2 - \theta\sigma^2 < 0$).** Solutions of equation (43). (Ornstein-Uhlenbeck process starting from $r_0 \overset{\mathcal{L}}{\sim} \mathcal{N}(0, \sigma_0^2)$, $\sigma_0 \neq 0$.) Values for this figure are $T = 3$, $\sigma = 1$, $\theta = 3$ and $\sigma_0^2 = 0.3$.

B.3.3 Pseudo-Code for computing Ornstein-Uhlenbeck eigenvalues

A pseudo-code for the computation of the n-th eigenvalue of the Ornstein-Uhlenbeck covariance operator is given in algorithm 1. In this algorithm, the function **search**$(a, \text{left}, \text{right})$ stands for a root finding method. It fills argument a with the root of equation (41) that is bracketed by [left, right].

In the author's implementation, one uses the Brent's method [3] as a reliable root finding method. As Newton-like methods, Brent method can take advantage of a guess of the value of the root. (One needs a bracketing method: the idea is to start from a small interval around the guess that is geometrically expanded, until the limiting range [left, right] is reached.)

B.3.4 A numerical guess for the value of ω_n.

As we have seen, we use a root finding method for evaluating numerically the value of ω_n. In this section, we propose a numerical guess for this quantity, that can be used as a starting point in the root finding method.

Function tan is approximated on $]-\frac{\pi}{2}, \frac{\pi}{2}[$ by the rational fraction $\psi(x) := \frac{\frac{4(8-\pi^2)x^3}{\pi^4}+x}{1-\frac{4x^2}{\pi^2}}$, which is a good uniform approximation of tan on this interval. $\| \tan -\psi \|_\infty^{]-\frac{\pi}{2},\frac{\pi}{2}[} = \frac{10-\pi^2}{2\pi} \approx 0.02075$.

Now, in the case of the Ornstein-Uhlenbeck eigenvalue computation, for a deterministic start point, equation (43) can be approximated by

$$\theta\psi(\omega_n T + n\pi) = -\omega_n \quad n \geq 1. \tag{45}$$

This comes to a polynomial equation of degree 3 for every $n > 0$ which has a unique solution $w_n^{guess} \in]\frac{n\pi}{T} - \frac{\pi}{2T}, \frac{n\pi}{T}[$. This numerical guess yields a good accuracy for approximating the value of ω_n.

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, May 15 2007. Comment: 27p.

[3] R.P. Brent. *Algorithms for minimization without derivatives*. Prentice-Hall,. Englewood Cliffs, N.J., 1972.

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

Require: $\theta > 0,\ \sigma \geq 0,\ \sigma_0 \geq 0,\ T \geq 0,\ n \geq 1.$

if $\sigma_0 = 0$ **then**

{There is a unique solution w_n of (41) in the interval $]\frac{n\pi}{T} - \frac{\pi}{2T}, \frac{n\pi}{T}[$. }

search$(w_n, \frac{n\pi}{T} - \frac{\pi}{2T}, \frac{n\pi}{T})$.

else

{Here $\sigma_0 > 0$.}

if $(\theta^2\sigma_0^2 - \theta\sigma^2) \geq 0$ **then**

{The vertical asymptote of the right hand of equation 44 lies on the left of 0. }

if $(\theta^2\sigma_0^2 - \theta\sigma^2)T - \sigma^2 < 0$ **then**

{There is a unique solution w_n of (41) in the interval $]0, \frac{\pi}{2T}[$. }

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

else

{The smallest stricly positive solution w_1 of equation (41) lies in the interval $]\frac{\pi}{2T}, \frac{\pi}{T}[$.}

search$(w_n, \frac{n\pi}{T}, \frac{n\pi}{T} + \frac{\pi}{2T})$.

end if

else

{The vertical asymptote of the right hand of equation 44 lies on the right of 0. }

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

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

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

search$(w_n, \frac{n\pi}{T} - \frac{\pi}{2T}, \frac{n\pi}{T})$.

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$(w_n, \frac{n\pi}{T} - \frac{\pi}{2T}, \sqrt{\theta\frac{\sigma^2}{\sigma_0^2} - \theta^2})$.

else

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

end if

end if

end if

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

Return λ_n.

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

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

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

[7] E. W. Forgy. Cluster analysis of multivariate data: efficiency vs interpretability of classifications. *Biometrics*, 21:768 769, 1965.

[8] Allen Gersho and Robert M. Gray. *Vector quantization and signal compression*. Kluwer Academic Publishers, Norwell, MA, USA, 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 New York, Inc., Secaucus, NJ, USA, 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 Lacombes. *Elements d'analyse fonctionnelle ; Cours et exercices avec réponses.* Dunod, 2009.

[13] Y.L. Hsu, T.I. Lin, and C.F. Lee. Constant elasticity of variance (cev) option pricing model: Integration and detailed derivation. *Mathematics and Computers in Simulation*, 79(1):60 – 71, 2008.

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

[15] Antoine Lejay and Victor Reutenauer. A variance reduction technique using a quantized Brownian motion as a control variate. *Preprint*, 2008.

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

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

[18] 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, Feb 2006.

[19] 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.

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

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

[22] 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.

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

[24] Gilles Pagès, Jacques Printems, and Huyen Pham. Optimal quantization methods and applications to numerical problems in finance. 2003.

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

[26] Panayotis Tsaparas. Nearest-neighbor search in multidimensional spaces. *Qualifying Depth Oral Report 319-02, Dept. of Computer Science, University of Toronto*, 1999.

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