Méthodes numériques par quantification optimale en finance

Numerical methods by optimal quantization in finance

Thibaut Montes

Laboratoire de Probabilités, Statistique et Modélisation - UMR 8001
Sorbonne Université

Thèse pour l'obtention du grade de :
Docteur de l'université Sorbonne Université

Sous la direction de : Gilles Pagès
Vincent Lemaire

Rapportée par : Giorgia Callegaro
Benoîte de Saporta

Présentée devant un jury composé de : *(président)*
Gilles Pagès *(directeur)*
Vincent Lemaire *(co-directeur)*
Giorgia Callegaro *(rapporteure)*
Benoîte De Saporta *(rapporteure)*
Benjamin Jourdain *(examinateur)*
Idris Kharroubi *(examinateur)*
Huên Pham *(examinateur)*
Abbas Sagna *(examinateur)*
Jean-Michel Fayolle *(invité)*

École Doctorale de Sciences Mathématiques de
Paris-Centre
Section Mathématiques appliquées

Mercredi 24 Juin 2020

Remerciements

Tout d'abord, je souhaite exprimer ma gratitude envers mes directeurs de thèse Vincent Lemaire et Gilles Pagès qui m'ont accompagné tout au long de mon doctorat, m'ont fait confiance et ont toujours été disponibles pour répondre à mes questions. J'ai énormément appris à leurs côtés et leur en suis extrêmement reconnaissant. Leur supervision complémentaire est tout ce que je souhaite à leurs futurs doctorants. Elle m'a permis d'accomplir plus, à la fois en théorie et en pratique, que je n'aurais pu l'imaginer au début du doctorat. Je remercie également Jean-Michel Fayolle pour avoir rendu possible cette thèse CIFRE et de s'être investi pour rendre possible le dialogue entre ma recherche académique et les développements pratiques d'ICA.

Merci à mes deux rapporteures Giorgia Callegaro et Benoîte de Saporta d'avoir pris le temps de lire mon manuscrit de thèse. J'ai été très honoré de la confiance dont leurs rapports témoignent. Je tiens également à exprimer toute ma reconnaissance envers Benjamin Jourdain, Idris Kharroubi, Huyên Pham et Abass Sagna d'avoir accepté de faire partie de mon jury de thèse.

Je tiens également à remercier l'ensemble des doctorants que j'ai eu la chance de croiser durant ma thèse. On ne peut pas ne pas citer les thésards du bureau 201/3 (Léa, Rancy, Guillermo, Armand, Nicolas et Nicolas, Romain et Babacar) pour les cafés, les déjeuners en soum-soum au Restaurant du Personnel et les bières partagées. Je n'oublie pas non plus tous les autres doctorants que j'ai eu la chance de rencontrer et d'apprendre à connaitre durant les différentes conférences (le CEMRACS avec ses calanques - Métabief et son restaurant de raclette - Padoue et ses spritz). Je souhaite également remercier mes collègues chez ICA avec qui j'ai pu échanger, autour d'un café, d'un déjeuner ou d'un jogging, en particulier, Guillaume, Eric, Vincent, Emmanuel et Lauriane.

Je voudrais ensuite remercier ma famille, Jeanine ma maman, Amélie ma soeur, Joëlle ma marraine, mes oncles et tantes Michel et Elisabeth, Monique et Michel ainsi que mes cousines et cousins, pour avoir toujours cru en moi et avoir été présent, surtout dans les moments difficiles. Je tiens également à remercier Hélène, la maman de Julie, et toute la famille Besnard-Corblet-Hatzopoulos pour m'avoir soutenu et encouragé dès notre rencontre et encore plus durant ma thèse.

Je souhaiterais également remercier tous mes amis pour de merveilleux moments partagés et qui, par effet de bord, m'ont aidé à écrire cette thèse en rendant la vie plus douce. En particulier, je remercie Juju pour toutes ces années d'amitiés qui me sont précieuses ; Pierre,

Patrick et Cécile pour toutes ces soirées / week-ends gastronomalcooliques ; les Scubes et leur +1 (Grégory, Henrik et Anita, Howon, Laure, Léa, Julien et Christina, Anh-Mai, Thomas, Jérôme, Vincent et Gamze, Juliette) pour leur bonne humeur permanente ; Laurène et Robin pour ces longues et étranges heures passées à l'opéra ; Kevin et John pour leurs soirées à thème et pour finir Karine et Myriam pour ces très bonnes années à Aix-en-Provence.

Enfin et surtout, merci Pauline pour ton soutien inconditionnel et pour avoir toujours cru en moi. Tu m'as toujours poussé à aller plus loin, à prendre confiance en moi et ce depuis le début de notre rencontre. Merci pour avoir toujours pris le temps de m'écouter (même lorsque je radote "un peu").

Abstract

This thesis is divided into four parts that can be read independently. In this manuscript, we make some contributions to the theoretical study and financial applications of optimal quantization.

In the first part, we recall the theoretical foundations of optimal quantization as well as the classical numerical methods to build optimal quantizers.

The second part focuses on the problem of numerical integration in dimension 1. This problem arises when one wishes to numerically compute expectations, such as the valuation of derivatives in finance that are expressed as the expectation of a function of a single financial asset. We recall the existing strong and weak error results and extend the results of order 2 convergence rate to other function classes with less regularity. In a second step, we present a weak error development result in one dimension and a second development in a higher dimension when the chosen quantizer is a product quantizer.

In the third part, we look at a first numerical application. We introduce a stationary Heston model in which the initial condition of volatility, instead of being deterministic as in the standard model, is assumed to be randomly distributed with the stationary distribution of the CIR EDS governing volatility. This variant of the original Heston model produces for European options on short maturities a steeper *smile* of implied volatility than the standard model. We then develop a product recursive quantization-based numerical method for the valuation of Bermudan options and barriers.

The fourth and last part deals with a second numerical application, the pricing of Bermudan exchange rate options in a 3 factor model, i.e. where the exchange rate, domestic and foreign interest rates are stochastic. These products are known in the markets as PRDC (Power Reverse Dual Currency). We propose two schemes to evaluate this type of options, both based on optimal product quantization and establish a priori error estimates.

Résumé

Cette thèse est divisée en quatres parties pouvant être lues indépendamment. Dans ce manuscrit, nous apportons quelques contributions à l'étude théorique et aux applications en finance de la quantification optimale.

Dans la première partie, nous rappelons les fondements théoriques de la quantification optimale ainsi que les méthodes numériques classiques pour construire des quantifieurs optimaux.

La seconde partie se concentre sur le problème d'intégration numérique en dimension 1. Ce problème apparait lorsque l'on souhaite calculer numériquement des espérances, tel que l'évaluation de produits dérivés en finance qui s'expriment sous la forme d'un calcul d'espérance d'une fonction d'un unique actif financier. Nous y rappelons les résultats d'erreurs forts et faibles existants et étendons les résultats des convergences d'ordre 2 à d'autres classes de fonctions moins réguliers. Dans un deuxième temps, nous présentons un résultat de développement d'erreur faible en dimension 1 et un second développement en dimension supérieure pour un quantifieur produit.

Dans la troisième partie, nous nous intéressons à une première application numérique. Nous introduisons un modèle de Heston stationnaire dans lequel la condition initiale de la volatilité, au lieu d'être déterministe comme dans le modèle standard, est supposée aléatoire de loi la distribution stationnaire de l'EDS du CIR régissant la volatilité. Cette variante du modèle d'Heston original produit pour les options européennes sur les maturités courtes un *smile* de volatilité implicite plus prononcé que le modèle standard. Nous développons ensuite une méthode numérique à base de quantification récursive produit pour l'évaluation d'options bermudiennes et barrières.

La quatrième et dernière partie traite d'une deuxième application numérique, l'évaluation d'options bermudiennes sur taux de change dans un modèle 3 facteurs, i.e où le taux de change, les taux d'intérêts domestiques et étrangers sont stochastiques. Ces produits sont connus sur les marchés sous le noms de PRDC (Power Reverse Dual Currency). Nous proposons deux schémas pour évaluer ce type d'options toutes deux basées sur de la quantification optimale produit et établissons des estimations d'erreur à priori.

Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

1.1 Optimal Quantization

This thesis is devoted to various theoretical aspects of optimal quantification in relation to numerical integration as well as several applications in finance. Optimal quantization was first introduced by Sheppard in 1897 in [She97]. His work focused on the optimal quantization of the uniform distribution over unit hypercubes. It was then extended to more general laws with or without compact support, motivated by applications to signal transmission in the Bell Laboratory in the 1950s (see [GG82]). Optimal quantization is also linked to an unsupervised learning computational statistical method. Indeed, the "k-means" method, which is a nonparametric automatic classification method consisting, given a set of points and an integer k, in dividing the points into k classes ("clusters"), is based on the same algorithm as the Lloyd method used to build an optimal quantizer. The "k-means" problem was formulated by Steinhaus in [Ste56] and then taken up a few years later by MacQueen in [Mac67]. In the 90s, optimal quantization was first used for numerical integration purposes for the approximation of expectations, see [Pag98], and later used for the approximation of conditional expectations: see [BPP01; BP03; BPP05] for optimal stopping problems applied to the pricing of American options, [PP05; PRS05] for non-linear filtering problems, [BDD13; PCR09; PPP04a; PPP04b] for stochastic control problems, [Gob+05] for discretization and simulation of Zakai and McKean-Vlasov equations and [BSD12; DD12] in the presence of piecewise deterministic Markov processes (PDMP).

1.1.1 Definitions and key findings

Let X be a random vector with values in $\mathbb{R}^d$ provided with a $|\cdot|$ norm, here always Euclidean, with distribution $\mathbb{P}_X$, defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that $X \in L^2_{\mathbb{R}^d}$. The quantization of X consists in approximating X by a random vector $q(X)$ where q is a Borelian function with values in $\Gamma_N = \{x_1^N, \dots, x_N^N\} \subset \mathbb{R}^d$. In addition, we can see that $\mathrm{dist}(X, q(X)) \geqslant \mathrm{dist}(X, \Gamma_N)$ with equality if and only if q is a nearest neighbor projection, denoted $q = \mathrm{Proj}_{\Gamma_N}$. This

nearest neighbor projection Proj_{Γ_N} is associated biunivocally with a Voronoï Borelian partition $\big(C_i(\Gamma_N)\big)_{1\leqslant i\leqslant N}$ of $\mathbb{R}^d$ such that

$$C_i(\Gamma_N) \subset \big\{\xi \in \mathbb{R}, |\xi - x_i^N| \leqslant \min_{j\neq i} |\xi - x_j^N|\big\}.$$

Thus, the associated nearest neighbor projection is defined by

$$\mathrm{Proj}_{\Gamma_N}(\xi) = \sum_{i=1}^{N} x_i^N \, \mathbb{1}_{\xi\in C_i(\Gamma_N)}\,.$$

Such quantization is called "*Voronoï*". We will note $\widehat{X}^{\Gamma_N}$ the closest neighbor projection of X on $\Gamma_N = \{x_1^N, \dots, x_N^N\}$, then

$$\widehat{X}^{\Gamma_N} = \mathrm{Proj}_{\Gamma_N}(X).$$

We lighten the notation from $\widehat{X}^{\Gamma_N}$ to $\widehat{X}^N$ for clarity.

Then, the law of a quantizer $\widehat{X}^N$ is entirely characterized by the centroid grid $\Gamma_N = \{x_i^N,\, 1 \leqslant i \leqslant N\}$ in which the quantizer takes its values and the N-tuple of the weights p_i^N which represent the probability that $\widehat{X}^N$ is equal to x_i^N or, equivalently, that X belongs to the Voronoï cell i, i.e.

$$p_i^N = \mathbb{P}\left(\widehat{X}^N = x_i^N\right) = \mathbb{P}\left(X \in C_i(\Gamma_N)\right), \quad i = 1, \dots, N.$$

In this thesis, we'll be working primarily with *optimal quadratic quantization*. The term *optimal* comes from the fact that we look for the best approximation of X in the sense that we will want to minimize the distance between the random vectors X and $\widehat{X}^N$ by optimizing the grid Γ_N for a given size N. This distance is measured in L^2-norm, hence the term *quadratic*. The distance between X and $\widehat{X}^N$, denoted as $\|X - \widehat{X}^N\|_2$, is called the mean quantization error. But we often reason in terms of distortion which is none other than the square of the mean quantization error. For a N-tuple, it is defined by

$$\mathcal{Q}_{2,\mathrm{N}} : x = (x_1^N, \dots, x_N^N) \longmapsto \mathbb{E}\left[\min_{i=1,\dots,N} |X - x_i^N|^2\right] = \|X - \widehat{X}^N\|_2^2.$$

So, we're looking for the grid Γ_N with cardinal at most N such that the quantifier $\widehat{X}^N = \mathrm{Proj}_{\Gamma_N}(X)$ minimizes

$$\min_{\Gamma_N\subset\mathbb{R}^d, |\Gamma_N|\leqslant N} \|X - \widehat{X}^N\|_2^2.$$

Such a grid always exists when $X \in L^2$ (see Theorem 1.1.1 below). In the Figure 1.1, we present two quantizations of size $N = 100$ of a centered Gaussian vector with identity covariance matrix. On the left, we represent an i.i.d. sample of the Gaussian vector and on the right an optimal quantizer. The color of each cell represents the probability p_i^N associated to the cell $C_i(\Gamma_N)$ of centroid x_i^N (red dot).

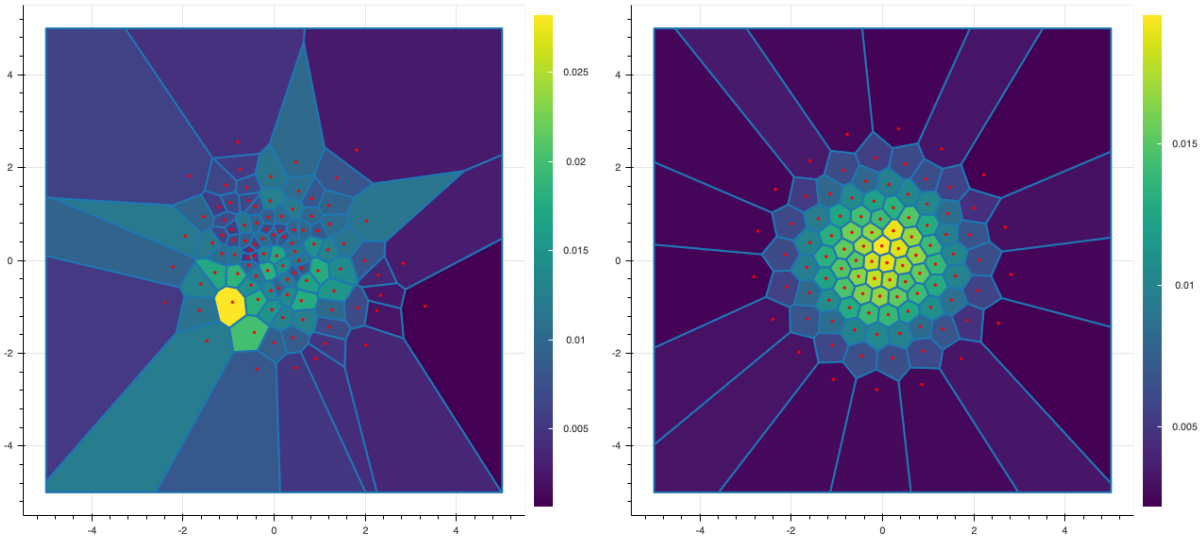


Fig. 1.1 *Two quantizations of size* $N = 100$ *of a centered Gaussian vector with identity covariance matrix.*

The minimization problem being set, several results have been demonstrated in the literature, see for example the two books [GL00; Pag18] for more details on the theory of optimal quantization. Let us note that this theory can be fully developed in a L^p framework and we then speak of L^p-optimal quantization. We first mention a result that ensures the existence of an optimal quantizer.

Theorem 1.1.1. *(Existence of an optimal* N*-quantization) Let* $X \in L^2_{\mathbb{R}^d}(\mathbb{P})$ *and* $N \in \mathbb{N}^\star$*.*

(a) The quadratic distortion function $\mathcal{Q}_{2,\mathrm{N}}$ *at the* N *level reaches a minimum in (at least) one* N*-tuplet* $x^\star = (x_1^N, \ldots, x_N^N)$ *and the associated grid* $\Gamma_N^\star = \{x_i^N, i = 1, \ldots, N\}$ *is called an optimal* N*-quantizer.*

(b) If the $\mathbb{P}_X$ *distribution support of* X *has at least* N *elements, then* $x^\star = (x_1^N, \ldots, x_N^N)$ *has pairwise distinct components and* $\mathbb{P}_X\left(C_i(\Gamma_N^\star)\right) > 0, i = 1, \ldots, N$*. In addition, the sequence* $N \mapsto \inf_{x \in (\mathbb{R}^d)^N} \mathcal{Q}_{2,\mathrm{N}}(x)$ *converges to* 0 *and is strictly decreasing as long as it is strictly positive.*

In addition to knowing that the quadratic distortion decreases towards 0, the exact speed of convergence has been established through the contributions of several authors: [Zad82; BW82; GL00]. The theorem has been demonstrated in the L^p case and thus characterizes the quantization error L^p.

Theorem 1.1.2. *(Zador's Theorem) Let* $d \in \mathbb{N}^\star$ *and* $p \in (0, +\infty)$*.*

(a) Sharp rate. *Let* $X \in L^{p+\delta}_{\mathbb{R}^d}(\mathbb{P})$ *with* $\delta > 0$*. Let* $\mathbb{P}_X(d\xi) = \varphi(\xi) \cdot \lambda_d(d\xi) + \nu(d\xi)$*, where* $\nu \perp \lambda_d$ *i.e.* ν *is singular with respect to the Lebesgue measure* λ_d *on* $\mathbb{R}^d$*. Then, it exists*

a constant $\widetilde{J}_{p,d} \in (0, +\infty)$ *such that*

$$\lim_{N\to+\infty} N^{1/d} \min_{\Gamma_N\subset\mathbb{R}^d, |\Gamma_N|\leqslant N} \|X - \widehat{X}^N\|_p = \widetilde{J}_{p,d}\left[\int_{\mathbb{R}^d} \varphi^{\frac{d}{d+p}} d\lambda_d\right]^{\frac{1}{p}+\frac{1}{d}}$$

where $\widehat{X}^N$ *is an* L^p*-optimal quantization of* X*.*

(b) Non-asymptotic upper-bound [GL00; Pag18]. *Let* $\delta > 0$*. There exists a real constant* $C_{d,p,\delta} \in (0, +\infty)$ *such that, for all random vector* X *with values in* $\mathbb{R}^d$*,*

$$\forall N \geqslant 1, \qquad \min_{\Gamma_N\subset\mathbb{R}^d, |\Gamma_N|\leqslant N} \|X - \widehat{X}^N\|_p \leqslant C_{d,p,\delta}\sigma_{\delta+p}(X)N^{-1/d}$$

where, for $r \in (0, +\infty), \sigma_r(X) = \min_{a\in\mathbb{R}^d} \|X - a\|_r < +\infty$*.*

1.1.2 Construction of an optimal quantizer

There are many methods to build an optimal quantizer. In some very rare cases, centroids are given explicitly, for example when $X \sim \mathcal{U}([a, b])$ where $a, b \in \mathbb{R}$, the Γ_N grid is given by

$$\Gamma_N = \{x_1^N, \ldots, x_N^N\} = \left\{\frac{2i-1}{2N} : i = 1, \ldots, N\right\}.$$

We also refer to [GL00] for Laplace law and [FP02] for semi-closed formulas for exponential law, power law and inverse power law. However, most of the time this is not the case so we have to use iterative methods to construct the grids and weights associated with each of the centroids. These iterative methods are divided into two large families: deterministic methods (Lloyd's algorithm, Newton-Raphson's algorithm and their variants, ...) which are based on explicit knowledge of the density and the distribution function of the X law and methods based on stochastic optimization (Competitive Learning Vector Quantization (CLVQ), randomized version of the Lloyd's algorithm, ...) requiring only the ability to simulate X. These methods are detailed in the Chapter 3.

Case of a real-valued random variable - $d = 1$**.** In the unidimensional case, we have a result of uniqueness of the optimal quantizer when the density of X is log-concave. This theorem has been demonstrated by Kieffer in his [Kie82] (see also [Pag98]).

If X is a random variable ($d = 1$) for which we know the first partial moment $K_X(\cdot)$ and the cumulative distribution function F_X of X

$$K_X(x) := \mathbb{E}[X\,\mathbb{1}_{X\leqslant x}] \qquad \text{and} \qquad F_X(x) := \mathbb{P}(X \leqslant x),$$

then we use in priority deterministic methods which allow to build very quickly an optimal quantizer of X, such as the Lloyd's algorithm introduced in [Llo82] which is a fixed point

search algorithm. It is also possible to apply the Newton-Raphson algorithm by computing the Hessian of the quadratic distortion function (see [PP03] for a detailed example applied to a normal random variable). Other deterministic gradient descents can be used such as Levenberg-Marquardt or quasi-Newton methods. Otherwise, stochastic optimization-based methods such as the stochastic version of the Lloyd's algorithm or a stochastic gradient descent are used (see [Pag98]).

Example 1.1.3. In Figure 1.2, we represent in blue the density of a one-dimensional Gaussian random variable and in red the centroids of the optimal quantizer of size $N = 11$ of this same random variable. We also illustrate what the weights p_i^N associated to the centroids x_i^N represent. Moreover, we can approach the density (if it exists) at each point of the grid by the following relation

$$f(x_i^N) \approx \frac{2p_i^N}{x_{i+1/2}^N - x_{i-1/2}^N}.$$

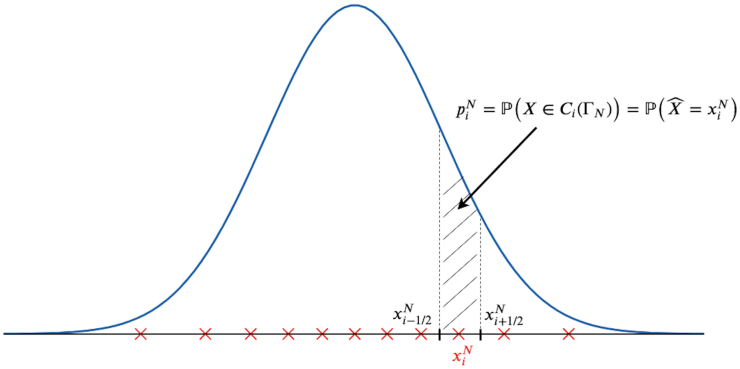


Fig. 1.2 *Density of a reduced centered Gaussian* $\mathcal{N}(0,1)$ *in blue and centroids of an optimal quantizer size* $N = 11$ *in red.*

Case of a random vector - $d \geqslant 2$**.** Now, let us consider a X random vector with values in $\mathbb{R}^d$ ($d \geqslant 2$). Two approaches exist to construct an optimal quantizer of the law of X.

The first approach is to apply the methodology developed in the scalar case directly to the vector case and thus obtain an optimal quantification of X. If we know the density of X then it is still possible in dimension 2 or 3 to apply the deterministic methods (cf. Chapter 3). However, from $d \geqslant 4$, we can only rely on stochastic optimization methods based on the simulation of samples of the X distribution.

The second, product quantization, consists in constructing an optimal quantizer of each of the components of the random vector and then constructing the quantizer by considering the cartesian product between all the optimally quantized components. More precisely, that is $X = (X^\ell)_{\ell=1:d}$, a random vector with values in $\mathbb{R}^d$. We consider the d one-dimensional optimal

quantifiers $\widehat{X}^\ell$ of size N^ℓ of each of the marginal X^ℓ. Each quantizer $\widehat{X}^\ell$ takes its values from the grid $\Gamma_\ell^{N_\ell} = \{z_{i_\ell}^\ell, i_\ell \in \{1, \cdots, N_\ell\}\}$. Thus, the quantizer product of X takes its values in the grid Γ^N which is the Cartesian product of the one-dimensional grids, i.e. $\Gamma^N = \prod_{\ell=1}^d \Gamma_\ell^{N_\ell}$ of size $N = N^1 \times \cdots \times N^d$ or, equivalently,

$$\Gamma^N = \{(x_{i_1}^1, \cdots, x_{i_\ell}^\ell, \cdots, x_{i_d}^d), \quad i_\ell \in \{1, \cdots, N_\ell\}, \quad \ell \in \{1, \cdots, d\}\}.$$

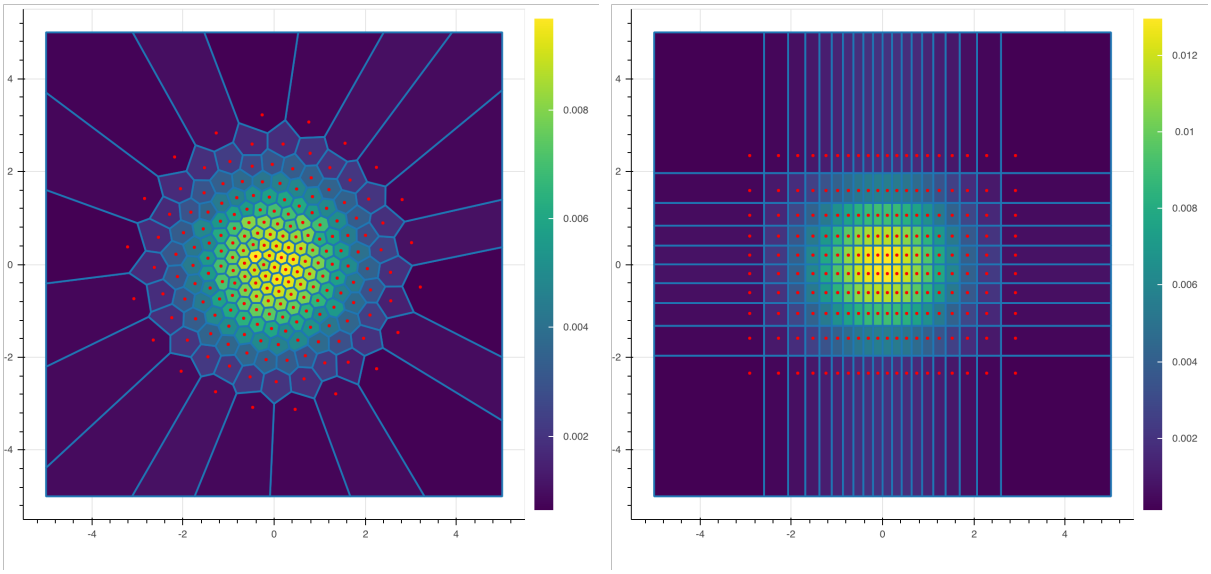


Fig. 1.3 *Two quantizations of size $N = 200$ of a centered Gaussian vector with unit covariance matrix. Optimal quantization on the left and Product quantization on the right.*

In the Figure 1.3, we compare the optimal quantization and the product quantization of a centered Gaussian vector with unit covariance matrix. Both methods have their advantages and disadvantages, the first method produces a better quantization of the random vector X compared to the product quantization but the induced numerical cost for the construction of an optimal quantizer is often much higher.

Case of diffusions. If now, instead of considering a random vector, we are interested in diffusions, i.e.

$$dX_t = b(t, X_t)dt + \sigma(t, W_t)dW_t$$

then there are, again, several solutions to quantize X_t. Specifically, given a time discretization at n-step $(t_k)_{0 \leqslant k \leqslant n}$, we are looking for the quantisers $\widehat{X}_{t_k}^{N_k}$ of size N_k of X_{t_k} that we denote $\widehat{X}_k^{N_k}$ and X_k in order to lighten the notations. The object we're trying to construct is called a *quantization tree*. A tree is characterized by the knowledge of the laws $\left(\Gamma_k, (p_i^k)_{1 \leqslant i \leqslant N_k}\right)$ of the

quantizers $(\widehat{X}_k)_{0\leqslant k\leqslant n}$ and of the transition probabilities $p_{i,j}^k$.

$$\mathbb{P}\left(\widehat{X}_{k+1} = x_j^{k+1} \mid \widehat{X}_k = x_i^k\right).$$

We will not present all the existing approaches that allow us to address the problem of quantization of diffusion but only those that allow us to use deterministic numerical methods for the optimization of the grids. For other approaches, based on stochastic algorithms we refer to the series of papers [BPP01; BP03].

Quantization of marginal laws. The problem of quantization of a diffusion has been initiated and developed in a series of articles [PPP04b; BPP05; BBP09; BBP10; CFG19]. If X_k can be simulated exactly, that is without the help of a time discretization scheme, and that we know the marginal law of X_k, at each instant t_k, then we are brought back to the case of the quantization of a random vector. Indeed, we can optimally quantize each random vector X_k using deterministic numerical methods if $d \leqslant 2$, producing an optimal quantization tree, or we can optimally quantize each of its components and then construct a product quantization of X_k, producing a product quantization tree.

Example 1.1.4. If we consider a Black-Scholes model with constant volatility σ and constant interest rates r

$$dS_t = S_t(rdt + \sigma dW_t), \quad \text{avec} \quad S_0 = s_0,$$

then we have an explicit form for S_t

$$S_t = S_0\, \mathrm{e}^{(r-\sigma^2/2)t+\sigma W_t}$$

so for a given date t, $\log(S_t/S_0) \sim \mathcal{N}\left((r-\sigma^2/2)t, \sigma^2 t\right)$ so we can optimally quantize S_t at each instant that interests us using deterministic methods (cf. Chapter 3). We can also quantize the Brownian W_t which is "more universal".

Recursive quantization. In the case where we do not know by the marginal law of X_k and that we need to use a discretization scheme (Euler-Maruyama, Milstein, ...), we will use a method called *recursive quantization*. Recursive quantization (also called Markovian quantization) was first introduced in [PPP04b] and then studied in depth in [PS15] for the case of a one-dimensional diffusion discretized by an Euler-Maruyama scheme. A fast algorithm based on deterministic methods to build the quantization tree is developed and analyzed. Subsequently, fast recursive quantization was extended to higher order one-dimensional schemes by [McW+18] and to higher dimensions by product quantization (see [PS18b; FSP18; Rud+17; CFG18; CFG17]). This method consists in building recursively in k the quantizers $\widehat{X}_k^{N_k}$ via the recursion

$$\widehat{X}_k^{N_k} = \mathrm{Proj}_{\Gamma_{N_k}}\left(\widetilde{X}_k\right) \quad \text{avec} \quad \widetilde{X}_k = \mathcal{E}_{k-1}\left(\widehat{X}_{k-1}^{N_{k-1}}, Z_k\right)$$

where $\mathcal{E}_{k-1}$ is a discretization scheme.

1.2 Numerical integration

A common problem in practice is to calculate the expectation of a function of X when X is a variable or a random vector, i.e. $\mathbb{E}\big[f(X)\big]$. However, except in very particular cases, it is not possible to calculate explicitly this quantity, it is the case for example if $X = X_T$ the value of a diffusion at the date T. This is why it is necessary to use numerical integration methods. [Pag98] introduces a cubature method based on optimal quantization in order to approximate expectations of the form $\mathbb{E}\big[f(X)\big]$. Let us consider $\widehat{X}^N$ an optimal quantizer of X, the fact that $\widehat{X}^N$ is discrete allows us to easily define the following cubature formula

$$\mathbb{E}\big[f(\widehat{X}^N)\big] = \sum_{i=1}^{N} p_i^N f(x_i^N). \tag{1.1}$$

Furthermore, given that $\widehat{X}^N$ was constructed as the best discrete approximation of X of cardinal at most N then it seems reasonable to think that $\mathbb{E}\big[f(\widehat{X}^N)\big]$ is a good approximation of $\mathbb{E}\big[f(X)\big]$.

In the Chapter 4, taken from the article "New Weak Error bounds and expansions for Optimal Quantization" published in *Journal of Computational and Applied Mathematics*, see [LMP19], we present new results in the real case concerning the error induced by the quantization-based approximation of expectation $\mathbb{E}\big[f(X)\big]$. This is a joint work with Vincent Lemaire and Gilles Pagès and it is accessible in arXiv or HAL. These "weak" results are summarized below.

1.2.1 Convergence rate of the weak error

In the first part of Chapter 4, we are interested in the rate of convergence from $\mathbb{E}\big[f(\widehat{X}^N)\big]$ to $\mathbb{E}\big[f(X)\big]$ as a function of N for different classes of functions f when X is a random variable with values in $\mathbb{R}$, i.e we look for the largest $\alpha > 0$ such that, for any function f in this class $\mathcal{F}$,

$$\overline{\lim_N} N^\alpha \big|\mathbb{E}\big[f(X)\big] - \mathbb{E}\big[f(\widehat{X}^N)\big]\big| \leqslant C_{f,X} < +\infty.$$

If we naively upper-bound the weak error by the strong error along the Lipschitz continuous functions, we obtain the following upper-bound (with $\alpha = 1$) for a sequence of N-quanifiers L^2-optimal N-quantifiers

$$N\big|\mathbb{E}\big[f(X)\big] - \mathbb{E}\big[f(\widehat{X}^N)\big]\big| \leqslant N[f]_{Lip}\|X - \widehat{X}^N\|_1 \leqslant N[f]_{Lip}\|X - \widehat{X}^N\|_2 \xrightarrow{N\to+\infty} C_f < +\infty$$

where Zador's Theorem (1.1.2) was used. Moreover, if we consider $f(x) = \mathrm{dist}(x, \Gamma_N)$ then f is a Lipschitz continuous function and we have

$$N\big|\,\mathbb{E}\big[f(X)\big] - \mathbb{E}\big[f(\widehat{X}^N)\big]\big| = N\|X - \widehat{X}^N\|_1 \leqslant N\|X - \widehat{X}^N\|_2 \xrightarrow{N\to+\infty} C_f < +\infty.$$

For some classes of functions we can prove that the cubature formula induces a weak error of order 2 ($\alpha = 2$). For example, if we consider functions that are derivable with a Lipschitz continuous derivative then we have an error of order 2, see [Pag98]. Indeed, we use a Taylor expansion with an integral remainder of the form

$$f(x) = f(y) + f'(y)(x-y) + \int_0^1 \big(f'(tx + (1-t)y) - f'(y)\big)(x-y)dt$$

and the stationarity property of an optimal quadratic quantizer as follows

$$\mathbb{E}\big[X \mid \widehat{X}^N\big] = \widehat{X}^N.$$

The first term in the Taylor expansion is zero because

$$\mathbb{E}\big[f'(\widehat{X}^N)(X - \widehat{X}^N)\big] = \mathbb{E}\Big[f'(\widehat{X}^N)\,\mathbb{E}\big[X - \widehat{X}^N \mid \widehat{X}^N\big]\Big] = \mathbb{E}\Big[f'(\widehat{X}^N)\big(\mathbb{E}\big[X \mid \widehat{X}^N\big] - \widehat{X}^N\big)\Big] = 0.$$

Thus, using Lipschitz's property of the derivative and Zador's theorem, we get a weak error of order 2, as expected.

$$\begin{aligned} N^2\big|\,\mathbb{E}\big[f(X)\big] - \mathbb{E}\big[f(\widehat{X}^N)\big]\big| &\leqslant N^2 \int_0^1 \mathbb{E}\big[\big|f'(tX + (1-t)\widehat{X}^N) - f'(\widehat{X}^N)\big||X - \widehat{X}^N|\big]dt \\ &\leqslant \frac{[f']_{Lip}}{2} N^2 \|X - \widehat{X}^N\|_2^2 \xrightarrow{N\to+\infty} C_f < +\infty. \end{aligned}$$

In the first part of Chapter 4, we extend these results concerning the convergence rate of the weak error of order higher than 1 to a wider class of functions with less regularity, more precisely, functions that are either :

- Lipschitz continuous piecewise affine functions with finitely many breaks of affinity,
- Lipschitz continuous convex functions,
- differentiable functions with piecewise-defined locally Lipschitz derivative (K breaks of affinity $\{a_1, \ldots, a_K\}$, such that $-\infty = a_0 < a_1 < \cdots < a_K < a_{K+1} = +\infty$ and the locally Lipschitz property of the derivative is defined by

  $$\forall k = 0, \ldots, K, \quad \forall x, y \in (a_k, a_{k+1}) \quad |f'(x) - f'(y)| \leqslant [f']_{k,Lip,loc}|x - y|\big(g_k(x) + g_k(y)\big)$$

  where $g_k : (a_k, a_{k+1}) \to \mathbb{R}_+$ are non-negative Borel functions,

- differentiable functions with piecewise-defined locally α-Hölder derivative (K breaks of affinity $\{a_1, \dots, a_K\}$, such that $-\infty = a_0 < a_1 < \cdots < a_K < a_{K+1} = +\infty$ and the locally α-Hölder property of the derivative is defined by

$$\forall k = 0, \dots, K, \quad \forall x, y \in (a_k, a_{k+1}), \quad |f'(x) - f'(y)| \leqslant [f']_{k,\alpha,loc} |x - y|^\alpha \big(g_k(x) + g_k(y)\big)$$

  where $g_k : (a_k, a_{k+1}) \to \mathbb{R}_+$ are non-negative Borel functions,

For the first three classes of functions, we show that the weak error is of order 2 and for the last one, of order $1 + \alpha$.

In the numerical part, we illustrate this result by evaluating the price of a European *Call* in a Black-Scholes model given by

$$I_0 := \mathbb{E}\left[\mathrm{e}^{-rT}(S_T - K)_+\right]$$

where $S_t = S_0 \,\mathrm{e}^{(r-\sigma^2/2)t+\sigma W_t}$ with $(W_t)_{t\in[0,T]}$ a Brownian motion. In order to approximate, with the help of quantization, the price of the European *Call* we can rewrite I_0 in two different ways

$$I_0 = \mathbb{E}\left[\varphi(S_T)\right] = \mathbb{E}\left[f(W_T)\right]$$

where φ is a piecewise affine function with one affinity break and f is a differentiable function with a piecewise locally Lipschitz continuous derivative. Thus, when considering quantizers of S_T or W_T and using the cubature formula, we observe, for both approximations, a weak error of order 2.

1.2.2 Weak error expansion of higher order

In the second part of Chapter 4, we are interested by the weak error expansion of the approximation of $\mathbb{E}\left[f(X)\right]$ by $\mathbb{E}\left[f(\widehat{X}^N)\right]$. That is, we're looking expansion of the form

$$\mathbb{E}\left[f(X)\right] = \mathbb{E}\left[f(\widehat{X}^N)\right] + \frac{c_2}{N^2} + O(N^{-(2+\beta)})$$

where $\beta \in (0, 1]$. In the previous section, we have already shown that the optimal quantization-based cubature formula approximation induces an error term of order $O(N^{-2})$ in the best case. Here, we seek to refine the previous results in order to obtain a "controlled" error expansion of order 2 and not a simple convergence rate of order 2.

In Section 4.3, we show that this expansion exists if the function $f : \mathbb{R} \to \mathbb{R}$ is twice differentiable with a Lipschtiz continuous second derivative. This result uses a Taylor expansion of order 2 with an integral remainder of the form

$$f(x) = f(y) + f'(y)(x - y) + \frac{1}{2} f''(y)(x - y)^2 + \int_0^1 (1 - t)\big(f''(tx + (1 - t)y) - f''(y)\big)(x - y)^2 dt$$

where we take the expectation on both sides of the equality and replace x and y with X and $\widehat{X}^N$, respectively. The second term on the righthand side is cancelled using the stationarity property of the optimal quadratic quantizer. For the third term, we rely on [Del+04] (Theorem 6) which states that $\forall g : \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}\left[g(X)\right] < +\infty$

$$\lim_N N^2 \, \mathbb{E}\left[g(\widehat{X}^N)|X - \widehat{X}^N|^2\right] = Q_2(\mathbb{P}_X) \int g(\xi) \, \mathbb{P}_X(d\xi)$$

that we apply to $g = f''$ where $Q_2(\mathbb{P}_X)$ is the Zador's constant. Thus, we already have the first two terms in the expansion of the error. For the last term, we use the Lipschitz property of the second derivative and the rest of the proof is based mainly on a result initially established in [GLP08] and then recently extended in [PS18a], known as "L^r-L^s distortion mismatch", which is formulated as follows : what can be said about the convergence rate of $\mathbb{E}\left[|X - \widehat{X}^N|^s\right]$ knowing that $\widehat{X}^N$ is a L^r-optimal quantizer when $s > r$ and $X \in L^s$? We cite this theorem for $d = 1$, which is the case we're interested in.

Theorem 1.2.1 (L^r-L^s-distorsion mismatch)**.** *Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ a random variable and $r \in (0, +\infty)$. Let $\mathbb{P}_X(d\xi) = \varphi(\xi) \cdot \lambda(d\xi) + \nu(d\xi)$, where $\nu \perp \lambda$ i.e. ν is singular with respect to the Lebesgue measure λ on $\mathbb{R}$ and φ is not identically null. Let $(\Gamma_N)_{N \geqslant 1}$ a sequence of L^r-optimal quantization grids and $s \in (r, r+1)$. If*

$$X \in L^{\frac{s}{1+r-s}+\delta}(\mathbb{P})$$

for a $\delta > 0$, so

$$\limsup_N N \|X - \widehat{X}^N\|_s < +\infty.$$

So, applying this theorem with $r = 2$ and $s = 2 + \beta$, we get a $O(N^{-(2+\beta)})$ for the last term and $\forall \, \beta \in (0, 1)$, we have the following expansion

$$\mathbb{E}\left[f(X)\right] = \mathbb{E}\left[f(\widehat{X}^N)\right] + \frac{c_2}{N^2} + O(N^{-(2+\beta)}).$$

This error expansion allows us to theoretically justify the use of Richardson-Romberg extrapolation which aims to *kill* the first error term of the expansion by linearly combining two quantification cubature formulas, respectively at N and M points, i.e.

$$\mathbb{E}\left[f(X)\right] = \mathbb{E}\left[\frac{M^2 f(\widehat{X}^M) - N^2 f(\widehat{X}^N)}{M^2 - N^2}\right] + O(N^{-(2+\beta)})$$

for $M = kN$ with $k > 1$.

We illustrate this result in the numerical part by valuing a European *spread* option in a 2-dimensional Black-Scholes model whose price is given by

$$I_0 := \mathbb{E}\left[\mathrm{e}^{-rT}(S_T^1 - S_T^2 - K)_+\right].$$

By preconditioning, we express I_0 as follows

$$I_0 = \mathbb{E}\left[\varphi(Z_2)\right]$$

where Z_2 is a standard Gaussian and φ is a twice differentiable function with a Lipschitz second derivative. Thus, considering N-optimal quantizers $\widehat{Z}^N$ of $Z_2 \sim \mathcal{N}(0,1)$, we approximate I_0 using the cubature formula based on optimal quantization (1.1) and observe a weak error of the order of 2. Moreover, using Richardson-Romberg extrapolation, we reach a weak error of the order of 3.

However, the relevance of the cubature method by optimal quantization when $d = 1$ remains limited because it is in competition with methods based on Gauss points. A multi-dimensional extension is on the other hand very useful as soon as d =3. We consider a function twice differentiable $f : \mathbb{R}^d \mapsto \mathbb{R}$ with a bounded and Lipschitz continuous Hessian. Furthermore, we assume that $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}^d$ has independent components $X_k, k = 1, \ldots, d$ and that the quantizer $\widehat{X}^N$ is a product quantizer of X with d components $(\widehat{X}_k^{N_k})_{k=1,\ldots,d}$ such that $N_1 \times \cdots \times N_d = N$. So, we have

$$\mathbb{E}\left[f(X)\right] = \mathbb{E}\left[f(\widehat{X}^N)\right] + \sum_{k=1}^{d} \frac{c_k}{N_k^2} + O\left(\left(\min_{k=1:d} N_k\right)^{-(2+\beta)}\right).$$

1.2.3 Variance reduction

In the last part of the Chapter 4, we present a new variance reduction method of a Monte Carlo estimator with control variates based on one-dimensional optimal quantization. Other variance reduction methods based on optimal quantization have been developed, see for example [CP15; Pag18] for more details. This approach is motivated by the rate of convergence of order 2 of the weak error induced by the quantization-based cubature formula for various classes of functions, including those mentioned above.

The problematic. Let $(Z_k)_{k=1,\ldots,d} = Z \in L^2_{\mathbb{R}^d}(\mathbb{P})$ a random vector and a function $f : \mathbb{R}^d \to \mathbb{R}$. We're interested in the following quantity

$$I := \mathbb{E}\left[f(Z)\right]. \tag{1.2}$$

Often we cannot compute this quantity explicitly, so a standard approach is to use a Monte Carlo estimator $\bar{I}_M := \frac{1}{M}\sum_{m=1}^{M} f(Z^m)$ by simulation of independent copies Z^m of Z to approximate

I. The convergence of the method and its rate are determined by the strong law of large numbers and the central limit theorem, respectively, which ensure, if Z is of integrable square, that

$$\bar{I}_M \xrightarrow{p.s.} \mathbb{E}\big[f(Z)\big] \quad \text{and} \quad \sqrt{M}\Big(\bar{I}_M - \mathbb{E}\big[f(Z)\big]\Big) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \sigma^2_{f(Z)}\right) \quad \text{when} \quad M \to +\infty$$

where $\sigma^2_{f(Z)} = \mathbb{V}\mathrm{ar}\left(f(Z)\right)$. We notice that, for a given simulation size M, the limiting factor of the method is $\sigma^2_{f(X)}$, so variance reduction methods were developed in order to reduce the value of $\sigma^2_{f(X)}$ and accelerate the convergence of the Monte Carlo estimator to I. The reader can refer to [Pag18; Gla13] for more details on Monte Carlo simulation and variance reduction methods in general such as control variates, antithetic method, stratification, importance sampling, ...

A new method of variance reduction by quantized control variable. Let Ξ^N be a random vector with values in $\mathbb{R}^d$ defined by

$$\Xi^N := (\Xi^N_k)_{k=1,\dots,d},$$

which will be our d-dimensional control variable, each Ξ^N_k component is given by

$$\Xi^N_k := f_k(Z_k) - \mathbb{E}\big[f_k(\widehat{Z}^N_k)\big],$$

where $f_k(z) := f(\mathbb{E}[Z_1], \dots, \mathbb{E}[Z_{k-1}], z, \mathbb{E}[Z_{k+1}], \dots, \mathbb{E}[Z_d])$ and $\widehat{Z}^N_k$ is an optimal quantization of size N of Z_k. We use here a unidimensional optimal quantization in order to take advantage of the weak error results previously shown, indeed the functions $f_k : \mathbb{R} \to \mathbb{R}$ are part of the classes of functions allowing us to reach a weak error of order 2. We introduce $I^{\lambda,N}$ as an approximation for (1.2)

$$I^{\lambda,N} = \mathbb{E}\big[f(Z) - \langle \lambda, \Xi^N \rangle\big] = \mathbb{E}\left[f(Z) - \sum_{k=1}^{d} \lambda_k f_k(Z_k)\right] + \sum_{k=1}^{d} \lambda_k \, \mathbb{E}\big[f_k(\widehat{Z}^N_k)\big] \tag{1.3}$$

where $\lambda \in \mathbb{R}^d$. The terms $\mathbb{E}\big[f_k(\widehat{Z}^N_k)\big]$ in (1.3) can be easily and quickly computed using the discreteness of quantizers.

At this point, we can define $\widehat{I}^{\lambda,N}_M$ the Monte Carlo estimator associated to $I^{\lambda,N}$

$$\widehat{I}^{\lambda,N}_M = \frac{1}{M} \sum_{m=1}^{M} \left(f(Z^m) - \sum_{k=1}^{d} \lambda_k f_k(Z^m_k) \right) + \sum_{k=1}^{d} \lambda_k \, \mathbb{E}\big[f_k(\widehat{Z}^N_k)\big].$$

It is important to notice that we introduce a bias when using such control variates, indeed for every $k \in \{1, \dots, n\}$, $\mathbb{E}[\Xi^N_k] \neq 0$ because $\mathbb{E}\big[f_k(\widehat{Z}^N_k)\big]$ is an approximation of $\mathbb{E}\big[f_k(Z_k)\big]$. However, the quantity that really interests us is not the bias induced by the estimator $\widehat{I}^{\lambda,N}_M$ but

rather the Mean Squared Error (MSE) giving us a bias-variance decomposition

$$\mathrm{MSE}(\widehat{I}_M^{\lambda,N}) = \underbrace{\left(\sum_{k=1}^{d} \lambda_k \Big(\mathbb{E}\left[f_k(\widehat{Z}_k^N)\right] - \mathbb{E}\left[f_k(Z_k)\right]\Big)\right)^2}_{biais^2} + \frac{1}{M} \underbrace{\mathbb{V}\mathrm{ar}\left(f(Z) - \sum_{k=1}^{d} \lambda_k f_k(Z_k)\right)}_{Variance\ du\ Monte\ Carlo}.$$

Thus, we can take higher values of N to make the bias term negligible compared to the variance of the estimator while controlling the total cost induced by the Monte Carlo estimator. In practice, we do not need to take very high values for N. Indeed, the bias term converges to 0 as N^{-4} if f belongs to the right class of functions, so taking optimal quantifiers of size 200 is more than enough to make the bias negligible compared to the variance of the Monte Carlo estimator. We develop this point in the third part of the Chapter 4.

In the numerical part of the Chapter 4, we apply the variance reduction method to the valuation of a basket option in a Black-Scholes model in dimension d. The control variate allows us to divide the variance of the Monte Carlo estimator by 100 for small dimensions ($d = 2$ or $d = 3$) and by 6 for larger dimensions ($d = 10$). We also observe that the bias induced by the quantification becomes negligible for grids with a size greater than 100 ($N > 100$).

1.3 Examples of applications in finance

1.3.1 Stationary Heston Model

In Chapter 5, we are interested in the stationary Heston model and more precisely in the evaluation of European, Bermuda and barrier options in this model as well as the calibration of the model. Chapter 5 corresponds to the preprint "Stationary Heston model: Calibration and Pricing of exotics using Product Recursive Quantization" accessible in arXiv or HAL (see [LMP20]). This article is a joint work with Vincent Lemaire and Gilles Pagès.

The standard Heston model was originally introduced by Heston in [Hes93]. It is a stochastic volatility model where the initial volatility condition is assumed deterministic. This model has become very popular mainly for the following two reasons: it is a stochastic volatility model so it introduces a *smile* in the surface of the implied volatility as observed in the market and the characteristic function of this model is given by a semi closed-form formula which allows us to value European options (*Call* & *Put*) almost instantaneously (see Carr & Madan in [CM99]). However, a remark often made about this model is that the *smile* of implied volatility is not steep enough for short maturities compared to what is observed in the market (see [Gat11]). Noticing that the volatility process is ergodic with a single invariant distribution $\nu = \Gamma(\alpha, \beta)$ where the α and β parameters depend on the volatility diffusion parameters, it has been proposed by Pagès & Panloup in [PP09] to directly consider that the process evolves under its stationary regime instead of starting it at time 0 from a deterministic value. This

choice has the effect of accentuating the volatility *smile* for short maturities while keeping the same behavior as the standard model for longer maturities. Later, the short and long-term behavior of the implied volatility generated by such a model was studied by Jacquier & Shi in [JS17].

Thus, the diffusion of the asset-volatility couple $(S_t^{(\nu)}, v_t^\nu)$ in the stationary Heston model is defined by

$$\begin{cases} \dfrac{dS_t^{(\nu)}}{S_t^{(\nu)}} = (r-q)dt + \sqrt{v_t^\nu}\big(\rho d\widetilde{W}_t + \sqrt{1-\rho^2}dW_t\big) \\ \quad dv_t^\nu = \kappa(\theta - v_t^\nu)dt + \xi\sqrt{v_t^\nu}d\widetilde{W}_t \end{cases}$$

where $v_0^\nu \sim \mathcal{L}(\nu) \sim \Gamma(\alpha,\beta)$ with $\beta = 2\kappa/\xi^2$, $\alpha = \theta\beta$.

Valuation of European Options. First of all, in the first part of the Chapter 5, we recall the method used for the valuation of a *Call* in the standard Heston model. Starting from the knowledge of the characteristic function $\psi\big(\lambda(v), u, T\big)$ of the logarithm of the asset at date T (see [SST04; Gat11; Alb+07] for a robust choice of formula), the price of the *Call* of *strike* K and maturity T on the asset $S_T^{(v)}$ in the standard Heston model where the volatility has as initial condition $v \in \mathbb{R}$ is given by

$$C(\phi(v), K, T) = \mathbb{E}\big[\,\mathrm{e}^{-rT}(S_T^{(v)} - K)_+\big] = s_0\,\mathrm{e}^{-qT}\,P_1\big(\lambda(v), K, T\big) - K\,\mathrm{e}^{-rT}\,P_2\big(\lambda(v), K, T\big)$$

where the quantities $P_1\big(\lambda(v), K, T\big)$ and $P_2\big(\lambda(v), K, T\big)$ are defined by

$$P_1\big(\lambda(v), K, T\big) = \frac{1}{2} + \frac{1}{\pi}\int_0^{+\infty} \mathrm{Re}\left(\frac{\mathrm{e}^{-\mathbf{i}u\log(K)}}{iu}\,\frac{\psi\big(\lambda(v), u-\mathbf{i}, T\big)}{s_0\,\mathrm{e}^{(r-q)T}}\right)du$$
$$P_2\big(\lambda(v), K, T\big) = \frac{1}{2} + \frac{1}{\pi}\int_0^{+\infty} \mathrm{Re}\left(\frac{\mathrm{e}^{-\mathbf{i}u\log(K)}}{\mathbf{i}u}\,\psi\big(\lambda(v), u, T\big)\right)du$$

with $\mathbf{i}$ the base of imaginary numbers (such that $\mathbf{i}^2 = -1$).

From this formula, we derive a method to compute the price I_0 of a *Call* in the stationary Heston model. Indeed, by preconditioning by v_0^ν, we have

$$I_0 = \mathbb{E}\big[\,\mathrm{e}^{-rT}\,\varphi(S_T^{(\nu)})\big] = \mathbb{E}\big[C(\phi(v_0^\nu), K, T)\big].$$

Thus, in order to obtain an approximation of I_0, we propose two methods. The first, based on optimal quantization, consists in building an optimal quantizer of the gamma law $\Gamma(\alpha,\beta)$ and then to use the cubature formula studied in the Chapter 4. The second method is to use a quadrature formula based on Laguerre polynomials.

Calibration. Once we are able to price European options in the stationary Heston model, we calibrate the model on market data to study the behavior of short-term implied volatility.

We also calibrate the standard Heston model to compare its implied volatility surface to the one of the stationary model. Both models are calibrated on the implied volatility surface of the Euro Stoxx 50 (see Figure 1.4). Since we are interested in the short-term behaviour of the implied volatility surface, the calibration of the models is performed on options with a maturity of 50 days ($T = 50/365$). We then observe the implied volatilities generated by the models for short-term maturities.

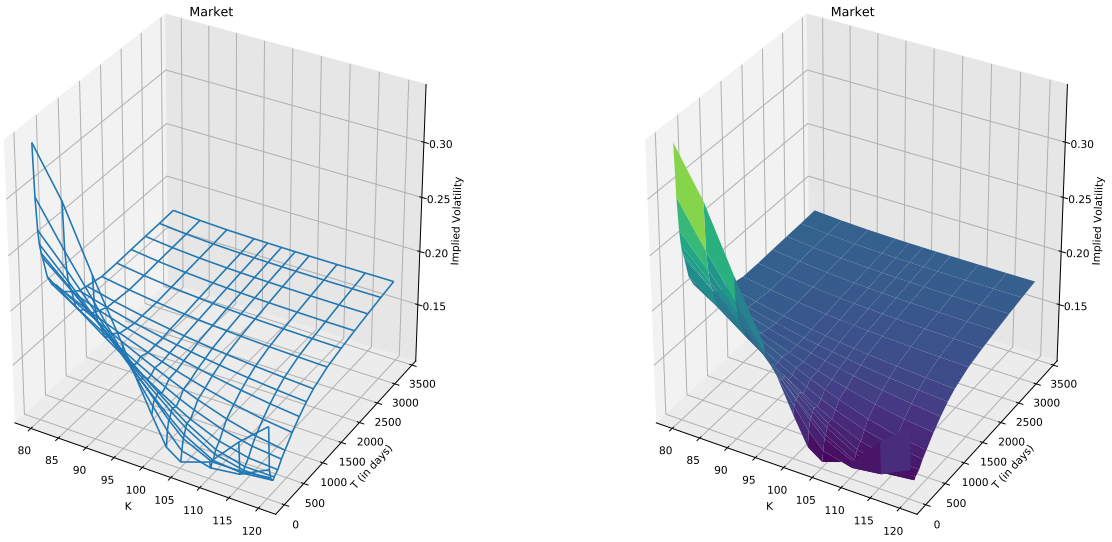


Fig. 1.4 *Implicit volatility area of the* Euro Stoxx *50 on September 26, 2019. ($S_0 = 3541$, $r = -0.0032$ and $q = 0.00225$)*

The set of 4 parameters of the stationary Heston model to be calibrated is defined by

$$\mathcal{P}_{SH} = \big\{(\theta, \kappa, \xi, \rho) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+ \times [-1, 1]\big\}$$

and the 5 standard model parameters $\mathcal{P}_H$ by

$$\mathcal{P}_H = \big\{(x, \theta, \kappa, \xi, \rho) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+ \times [-1, 1]\big\}.$$

The other parameters are directly observed in the market.

We can notice that the stationary model has one less parameter to be calibrated compared to the standard model, which makes its calibration more robust than the standard model which is known to be over-parameterized (see [**guarantee2009fitting**]). In practice, we observe that the calibration of the standard model is very dependent on the set of parameters used to initialize the optimization algorithm whereas it is not the case for the stationary model.

For the calibration of the models, the standard method consists in solving the following optimization problem

$$\min_{\phi\in\mathcal{P}} \sum_{K} \left(\frac{\sigma_{\text{IV}}^{Market}(K,T) - \sigma_{\text{IV}}^{Model}(\phi,K,T)}{\sigma_{\text{IV}}^{Market}(K,T)} \right)^2$$

where the quantities $\sigma_{\text{IV}}^{Market}(K,T)$ and $\sigma_{\text{IV}}^{Model}(\phi,K,T)$ are, respectively, market implied volatilities and those calculated with a Heston model of parameter $\phi = (\theta,\kappa,\xi,\rho)$ or $\phi = (x,\theta,\kappa,\xi,\rho)$ in appropriate cases.

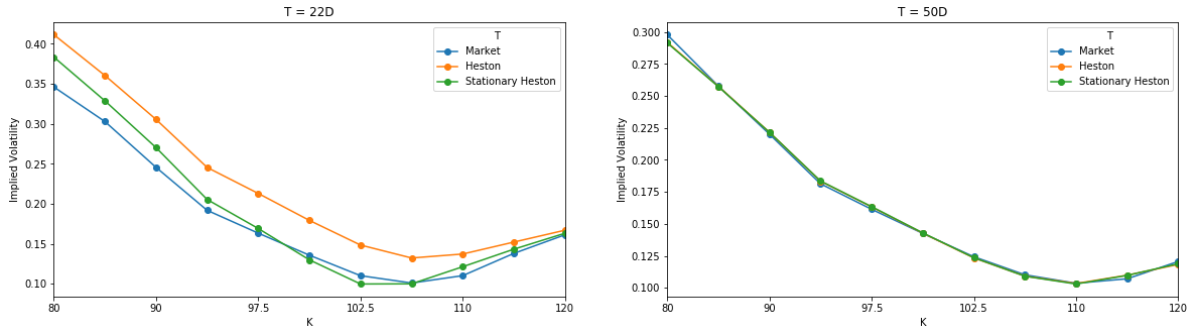


Fig. 1.5 *Implied volatility for maturity options* 22 *(left) and* 50 *(right) days after calibration without penalty.*

In Figure 1.5, we compare the implied volatility curves generated by the two models after calibration to European options with 50-day maturity. We observe that the stationary model produces a *smile* of volatility that is steeper than the standard model for options with a 22-day maturity. However, when we perform the calibration, we notice that the parameters obtained do not satisfy the Feller's condition

$$\xi^2 \leqslant 2\kappa\theta$$

which ensures the strict positivity of volatility. This property is important for the numerical valuation of exotic options discussed in the last part of the chapter.

Thus, to obtain parameters that satisfy the Feller condition, we constrain the parameters by adding a penalty to the minimization problem that becomes

$$\min_{\phi\in\mathcal{P}} \sum_{K} \left(\frac{\sigma_{\text{IV}}^{Market}(K,T) - \sigma_{\text{IV}}^{Model}(\phi,K,T)}{\sigma_{\text{IV}}^{Market}(K,T)} \right)^2 + \lambda \max(\xi^2 - 2\kappa\theta, 0)$$

where λ is the penalty factor adjusted during the procedure.

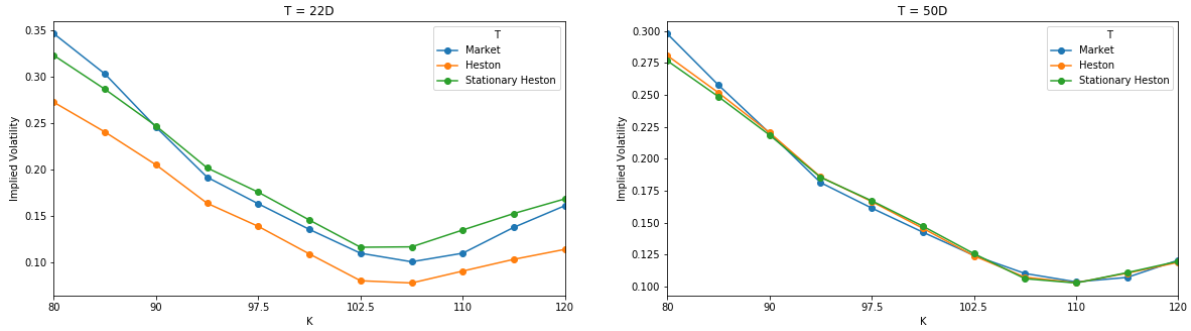


Fig. 1.6 *Implied volatility for maturity options* 22 *(left) and* 50 *(right) days after calibration with penalty.*

In the Figure 1.6, we make the same comparison as before. We notice that the addition of the penalty deteriorated the quality of the calibration at maturity 50 days. As for maturity 22 days, we observe that the stationary model again succeeds in producing a *smile* of volatility closer to that of the market than the standard model.

Exotic Options Valuation by Recursive Product Quantification. In the last part of the Chapter 5, we address the pricing of exotic options such as Bermudan options and barrier options using a *Backward Dynamic Principle Programming*. The numerical method we propose is based on recursive product quantization. We extend the methodology previously developed by [FSP18; CFG18; CFG17] where a Euler-Maruyama scheme was considered for the discretization in time of both assets and volatility.

Time discretization of diffusions. We made the choice to consider a hybrid scheme composed of an Euler-Maruyama scheme for the dynamics of the log-active $X_t = \log(S_t^{(\nu)})$ and a Milstein scheme for the *boosted* volatility process $Y_t = \mathrm{e}^{\kappa t}\, v_t^{\nu}$. Thus, we have

$$\begin{cases} \bar{X}_{t_{k+1}} = \mathcal{E}_{b,\sigma}\big(t_k, \bar{X}_{t_k}, \bar{Y}_{t_k}, Z_{k+1}^1\big) \\ \bar{Y}_{t_{k+1}} = \mathcal{M}_{\tilde{b},\tilde{\sigma}}\big(t_k, \bar{Y}_{t_k}, Z_{k+1}^2\big) \end{cases}$$

with $t_k = \frac{k}{n}$, n the number of discretization time steps, $Z_{k+1}^1 \sim \mathcal{N}(0,1)$ and $Z_{k+1}^2 \sim \mathcal{N}(0,1)$ such that $\mathbb{C}\mathrm{orr}(Z_{k+1}^1, Z_{k+1}^2) = \rho$. The Euler-Maruyama scheme is defined by

$$\mathcal{E}_{b,\sigma}(t,x,y,z) = x + b(t,x,y)h + \sigma(t,x,y)\sqrt{h}\, z$$

with

$$b(t,x,y) = r - q - \frac{\mathrm{e}^{-\kappa t}\, y}{2} \qquad \text{and} \qquad \sigma(t,x,y) = \mathrm{e}^{-\kappa t/2}\sqrt{y},$$

and the Milstein schema put into its canonical form

$$\mathcal{M}_{\widetilde{b},\widetilde{\sigma}}(t,x,z) = x - \frac{\widetilde{\sigma}(t,x)}{2\widetilde{\sigma}'_x(t,x)} + h\left(\widetilde{b}(t,x) - \frac{(\widetilde{\sigma}\widetilde{\sigma}'_x)(t,x)}{2}\right) + \frac{(\widetilde{\sigma}\widetilde{\sigma}'_x)(t,x)h}{2}\left(z + \frac{1}{\sqrt{h}\widetilde{\sigma}'_x(t,x)}\right)^2$$

with

$$\widetilde{b}(t,x) = \mathrm{e}^{\kappa t}\,\kappa\theta, \qquad \widetilde{\sigma}(t,x) = \xi\sqrt{x}\,\mathrm{e}^{\kappa t/2} \quad \text{and} \quad \widetilde{\sigma}'_x(t,x) = \frac{\xi\,\mathrm{e}^{\kappa t/2}}{2\sqrt{x}}.$$

Product Markovian Recursive Quantization. Once the choice of the discretization scheme in time has been made, we are interested in the discretization in space of the asset-volatility couple.

To do this, we first construct a Markovian quantization tree $(\widehat{Y}_{t_k})_{k=0,\dots,n}$. It is advantageous to notice that the volatility is autonomous and therefore we face a one-dimensional problem. Thus, the quantizers $\widehat{Y}_{t_k}$ are recursively constructed, i.e. $\widehat{Y}_{t_{k+1}}$ is an optimal quantizer of $\widetilde{Y}_{t_{k+1}}$ defined by

$$\widetilde{Y}_{t_{k+1}} = \mathcal{M}_{\widetilde{b},\widetilde{\sigma}}\big(t_k, \widehat{Y}_{t_k}, Z^2_{k+1}\big), \qquad \widehat{Y}_{t_{k+1}} = \mathrm{Proj}_{\Gamma^Y_{N_{2,k+1}}}\big(\widetilde{Y}_{t_{k+1}}\big).$$

Numerically, we use the methods based on deterministic algorithms for the 1 dimension developed in Chapter 3.

Now, using the fact that Y_t has already been quantized, we construct a Markov quantization tree $(\widehat{X}_{t_k})_{k=0,\dots,n}$ of X_t. Again we are brought back to a one-dimensional problem and we construct the quantizers $\widehat{X}_{t_k}$ recursively, i.e. $\widehat{X}_{t_{k+1}}$ is an optimal quantizer of $\widetilde{X}_{t_{k+1}}$ defined by

$$\widetilde{X}_{t_{k+1}} = \mathcal{E}_{b,\sigma}\big(t_k, \widehat{X}_{t_k}, \widehat{Y}_{t_k}, Z^1_{k+1}\big), \qquad \widehat{X}_{t_{k+1}} = \mathrm{Proj}_{\Gamma^X_{N_{1,k+1}}}\big(\widetilde{X}_{t_{k+1}}\big).$$

In order to alleviate the notations, we shall denote $\widehat{X}_k$ and $\widehat{Y}_k$ instead of $\widehat{X}_{t_k}$ and $\widehat{Y}_{t_k}$.

Now that we have calibrated the stationary Heston model and are able to construct a quantization tree for the asset-volatility couple, we are interested in the evaluation of exotic options and more specifically Bermudan or barrier options.

Bermudan options. The price on date t_k of a Bermudan option exercisable on dates $\{t_k, \cdots, t_n\}$ with payoff $\psi_{t_k}(X_{t_k}, Y_{t_k})$ on the date t_k is given by the Snell envelope V_k

$$V_k = \sup_{\tau\in\mathcal{T}^n_k} \mathbb{E}\Big[\,\mathrm{e}^{-r\tau}\,\psi_\tau(X_\tau, Y_\tau) \mid \mathcal{F}_{t_k}\Big],$$

where $\mathcal{T}^n_k$ represents the set of stopping times τ taking values in $\{t_k, t_1, \dots, t_n\}$. The *Backward Dynamic Principle Programming* allows to rewrite V_k as follows

$$\begin{cases} V_n = \mathrm{e}^{-rt_n}\,\psi_n(X_n, Y_n), \\ V_k = \max\big(\mathrm{e}^{-rt_k}\,\psi_k(X_k, Y_k), \mathbb{E}[V_{k+1} \mid \mathcal{F}_k]\big), \qquad 0 \leqslant k \leqslant n-1. \end{cases}$$

We then apply the methodology employed by [BP03; BPP05; Pag18] which consists in replacing X_k and Y_k by the quantizers $\widehat{X}_k$ and $\widehat{Y}_k$. By construction of the recursive quantization, the couple $(\widehat{X}_k, \widehat{Y}_k)$ is Markovian so we obtain the following *Quantized Backward Dynamic Principle Programming*

$$\begin{cases} \widehat{V}_n = \psi_n(\widehat{X}_n, \widehat{Y}_n), \\ \widehat{V}_k = \max\left(\psi_k(\widehat{X}_k, \widehat{Y}_k), \mathbb{E}[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{Y}_k)]\right), \qquad k = 0, \dots, n-1. \end{cases}$$

Finally, the price of the Bermudan option is given by $\mathbb{E}\left[\widehat{V}_0\right]$.

Barrier Options. The price on date t_k of a barrier option with maturity T, payoff f and barrier L is given by

$$P_{UO} = \mathrm{e}^{-rT}\, \mathbb{E}\left[f(X_T)\, \mathbb{1}_{\sup_{t\in[0,T]} X_t \leqslant L}\right].$$

For the valuation of the barrier option, we apply the algorithm based on the conditional law of Euler's scheme, see [Gla13; Sag10; Pag18]. Thus, once the asset-volatility couple is discretized in time, the price P_{UO} is rewritten as follows

$$\bar{P}_{UO} = \mathrm{e}^{-rT}\, \mathbb{E}\left[f(\bar{X}_T)\, \mathbb{1}_{\sup_{t\in[0,T]} \bar{X}_t \leqslant L}\right] = \mathrm{e}^{-rT}\, \mathbb{E}\left[f(\bar{X}_T) \prod_{k=0}^{n-1} G^k_{(\bar{X}_k, \bar{Y}_k), \bar{X}_{k+1}}(L)\right]$$

where

$$G^k_{(x,y),z}(u) = \left(1 - \mathrm{e}^{-2n\frac{(x-u)(z-u)}{T\sigma^2(t_k,x,y)}}\right) \mathbb{1}_{\{u \geqslant \max(x,z)\}}.$$

Finally, replacing $\bar{X}_k$ and $\bar{Y}_k$ by $\widehat{X}_k$ and $\widehat{Y}_k$ and using a recursive algorithm to approach $\bar{P}_{UO}$ by $\mathbb{E}[\widehat{V}_0]$, we have

$$\begin{cases} \widehat{V}_n = \mathrm{e}^{-rT}\, f(\widehat{X}_n), \\ \widehat{V}_k = \mathbb{E}\left[G^k_{(\widehat{X}_k, \widehat{Y}_k), \widehat{X}_{k+1}}(L)\, \widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{Y}_k)\right], \qquad 0 \leqslant k \leqslant n-1. \end{cases}$$

1.3.2 Pricing of Bermudan options in a 3-factor model (PRDC)

In the Chapter 6, we address the problem of Bermudan exchange rate option pricing where stochastic domestic and foreign interest rates are considered. In this case, we refer to a three-factor model. Chapter 6 corresponds to the article "Quantization-based Bermudan option pricing in the FX world" submitted to *Journal of Computational Finance* and accessible in arXiv or HAL (see [Fay+19]). This article is a joint work with Jean-Michel Fayolle, Vincent Lemaire and Gilles Pagès.

The need to evaluate such products originated in Japan at the end of the 20th century. Indeed, the persistence of low interest rates during the last decades of the century was one of the main reasons that led to the creation of exchange rate structured financial products. These

products met the need of investors seeking higher coupons than those based on the yen. As financial products became more and more complex, they became known as power reverse dual currency (PRDC) products, see [Wys17].

Even though these products were issued towards the end of the 20th century, they are still present in banks' portfolios and must be taken into account when evaluating counterparty risk such as Credit Valuation Adjustment (CVA), Debt Valuation Adjustment (DVA), Funding Valuation Adjustment (FVA), Capital Valuation Adjustment (KVA), ..., in short xVA (see [BMP13; CBB14; Gre15] for more details on the subject).

The model. $P(t,T)$ is defined as the value at time t of one unit of the selected currency delivered (i.e. paid) at time T, also known as the zero coupon price or discount factor. We will note the zero coupon with superscript d when we speak of a zero coupon in the domestic currency ($P^d(t,T)$) and with superscript f for the zero coupon in the foreign currency. The model used for the diffusion of the domestic and foreign zero coupons belongs to the Heath-Jarrow-Morton (HJM) family of yield curve models. For more details and theory on its models, we may refer to the following articles [EFG96; EMV92; HJM92; BS73].

Thus the diffusion of the domestic zero-coupon curve under the domestic risk-neutral probability $\mathbb{P}$ is given by

$$\frac{dP^d(t,T)}{P^d(t,T)} = r_t^d dt + \sigma_d(T-t)dW_t^d,$$

where W^d is a $\mathbb{P}$-Brownian Motion, r_t^d is the instantaneous domestic rate at time t and σ_d is the volatility. For the foreign zero-coupon curve, the dynamic is given, under the foreign risk-neutral probability $\widetilde{\mathbb{P}}$, by the diffusion

$$\frac{dP^f(t,T)}{P^f(t,T)}P^f(t,T) = r_t^f dt + \sigma_f(T-t)d\widetilde{W}_t^f,$$

where $\widetilde{W}^f$ is a $\widetilde{\mathbb{P}}$-Brownian Motion, r_t^f is the instantaneous foreign rate at time t and σ_f is the volatility. The two probabilities $\widetilde{\mathbb{P}}$ and $\mathbb{P}$ are supposed to be equivalent, i.e. $\widetilde{\mathbb{P}} \sim \mathbb{P}$ and there is ρ_{df} defined as the limit of the quadratic variation $\langle W^d, \widetilde{W}^f \rangle_t = \rho_{df} t$.

For the exchange rate (FX), we refer to S_t as the value at time $t > 0$ of one unit of foreign currency in the domestic currency. The dynamics of $(S_t)_{t\geqslant 0}$ is of Black-Scholes type and given by

$$\frac{dS_t}{S_t} = (r_t^d - r_t^f)dt + \sigma_S dW_t^S,$$

where r_t^d is the instantaneous rate of the domestic currency at time t, r_t^f is the instantaneous rate of the foreign currency at time t, σ_S is the volatility and W^S is a standard Brownian motion under the domestic risk-neutral probability.

The problem. Our objective is to price Bermudan options on the exchange rate S_t exercisable at $n+1$ dates: $\{t_0, \dots, t_n\}$. Thus, the date price t_k of the Bermuda option is given by the *Snell envelope* V_k of the obstacle $\left(\mathrm{e}^{-\int_0^{t_k} r_s^d ds}\, \psi_{t_k}(S_{t_k})\right)_{k=0:n}$.

$$V_k = \sup_{\tau \in \mathcal{T}_k^n} \mathbb{E}\left[\, \mathrm{e}^{-\int_0^{\tau} r_s^d ds}\, \psi_\tau(S_\tau) \mid \mathcal{F}_{t_k}\,\right]$$

where τ is a stopping-time with values in $\{t_k, \dots, t_n\}$ and $\mathcal{T}_k^n$ represents all such stopping-time.

Example 1.3.1. The payoff we consider in Chapter 6 is one of a PRDC coupon (see the example in Figure 1.7) defined by

$$\psi_{t_k}(x) = \min\left(\max\left(\frac{C_f(t_k)}{S_0}x - C_d(t_k), \mathrm{Floor}(t_k)\right), \mathrm{Cap}(t_k)\right)$$

where $\mathrm{Floor}(t_k)$ and $\mathrm{Cap}(t_k)$ are the floor/cap values chosen when creating the product, as well as $C_f(t_k)$ and $C_d(t_k)$ which are the coupon values of the foreign and domestic currencies to which we wish to compare ourselves.

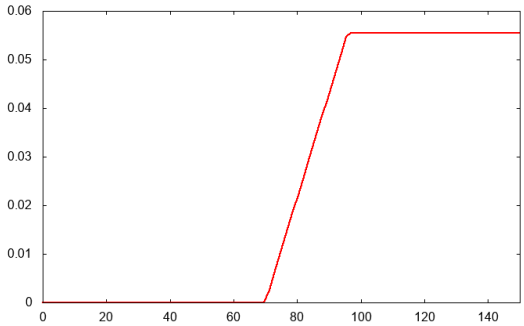


Fig. 1.7 *Example of a PRDC payoff* $\psi_{t_k}(S_{t_k}) = \min\left(\left(0.189\frac{S_{t_k}}{88.17} - 0.15\right)_+, 0.0555\right)$ *at date* t_k.

Backward Dynamic Principle Programming. The *Backward Dynamic Principle Programming* allows us to rewrite V_k as follows:

$$\begin{cases} V_n = \mathrm{e}^{-\int_0^{t_n} r_s^d ds}\, \psi_n(S_{t_n}), \\ V_k = \max\left(\mathrm{e}^{-\int_0^{t_k} r_s^d ds}\, \psi_k(S_{t_k}), \mathbb{E}[V_{k+1} \mid \mathcal{F}_{t_k}]\right), \qquad 0 \leqslant k \leqslant n-1. \end{cases}$$

Furthermore, we notice that the obstacle $\mathrm{e}^{-\int_0^t r_s^d ds}\, \psi_t(S_t)$ can be rewritten as a function h_t of two processes X_t and Y_t

$$\mathrm{e}^{-\int_0^t r_s^d ds}\, \psi_t(S_t) = h_t(X_t, Y_t)$$

where the couple (X, Y) is defined by

$$(X_t, Y_t) = \left(\sigma_S W_t^S + \sigma_f \int_0^t (t-s) dW_s^f, -\sigma_d \int_0^t (t-s) dW_s^d\right).$$

Thus, this new expression for the obstacle allows us to rewrite the Snell envelope problem in the form of

$$V_k = \sup_{\tau \in \mathcal{T}_k^n} \mathbb{E}\left[h_\tau(X_\tau, Y_\tau) \mid \mathcal{F}_{t_k}\right].$$

However, the couple (X_k, Y_k) is not Markovian and this poses a problem in the Principle of Dynamic Program because the conditioning that appears in the conditional expectation cannot be replaced by (X_k, Y_k). This is why we are led to consider the random vector (X, W^f, Y, W^d) which is Markovian. Thus the *Backward Dynamic Principle Programming* can be rewritten as follows

$$\begin{cases} V_n = h_n(X_n, Y_n), \\ V_k = \max\left(h_k(X_k, Y_k), \mathbb{E}\left[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\right]\right), \qquad 0 \leqslant k \leqslant n-1. \end{cases} \tag{1.4}$$

Quantization based numerical solution. We are now interested in the practical part of numerically computing the values V_k. In the Chapter 6, we have opted for a numerical method based on optimal quantization as introduced in [BPP01] and developed in [BP03; PPP04b; BPP05] for the evaluation of Bermudan options but with the variant that consists in using a product optimal quantization tree. This approach has the advantage of being fast, stable and accurate in small dimensions. However, as the dimension grows, the computation time can be very expensive and the convergence speed of the method is degraded because of the "curse of dimension" that affects the optimal quantization.

The first idea we present, when we want to discretize (1.4) by optimal quantization, is the most natural one. We replace the random variables X_k, W_k^f, Y_k and W_k^d by their optimal quantization $\widehat{X}_k$, $\widehat{W}_k^f$, $\widehat{Y}_k$ and $\widehat{W}_k^d$, of size N_k^X, $N_k^{W^f}$, $N_k^{W^f}$, N_k^Y and $N_k^{W^d}$ respectively, and we "force", in a certain sense, the Markov property by introducing the "forced" *Quantized Backward Dynamic Principle Programming* defined by

$$\begin{cases} \widehat{V}_n = h_n(\widehat{X}_n, \widehat{Y}_n), \\ \widehat{V}_k = \max\left(h_k(\widehat{X}_k, \widehat{Y}_k), \mathbb{E}\left[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)\right]\right), \qquad 0 \leqslant k \leqslant n-1. \end{cases}$$

The term "forced" is justified because $(\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)_k$ is not a Markov chain so this Backward Dynamic Principle Programming is not naturally associated to the Snell envelope. We denote by $N_k = N_k^X \times N_k^{W^f} \times N_k^Y \times N_k^{W^d}$ the global size of the quantization grid produced.

For this approximation, we provide an a priori quadratic error for $\|V_k - \widehat{V}_k\|_2$, $k = 0, \ldots, n$.

Theorem 1.3.2. *If the payoff functions $(\psi_{t_k})_{k=0:n}$ are Lipschitz continuous with compactly supported (right) derivative. Then the quadratic error induced by the quantization approximation $(\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)$ is upper-bounded by*

$$\left\|V_k - \widehat{V}_k\right\|_2 \leqslant \left(\sum_{l=k}^{n} C_{X_l} \left\|X_l - \widehat{X}_l\right\|_{2p}^2 + C_{Y_l} \left\|Y_l - \widehat{Y}_l\right\|_{2p}^2 + C_{W_l^d} \left\|W_l^d - \widehat{W}_l^d\right\|_{2p}^2 + C_{W_l^f} \left\|W_l^f - \widehat{W}_l^f\right\|_{2p}^2 \right)^{1/2},$$

where $1 < p < 3/2$ and $q \geqslant 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$ and the constants $C_{X_l}, C_{W_l^d}, C_{Y_l}, C_{W_l^f}$ are finite. So, by taking $\bar{N} = \min_k N_k$, we have

$$\lim_{\bar{N} \to +\infty} \left\|V_k - \widehat{V}_k\right\|_2^2 = 0.$$

The major problem with the approach we have just presented is the algorithmic complexity associated with this method due to the size of the product quantization grids. This complexity makes the computation of the conditional expectations appearing in the backward dynamic programming principle very expensive. Our objective is thus to reduce the size of the problem. To do so, we remove the processes W^d and W^f from the product-quantization tree to keep only X and Y. By doing so, we lose the Markov property of the random vector we are considering, but we significantly reduce the numerical complexity of the problem. In this context, (1.4) is approached by

$$\begin{cases} \widehat{V}_n = h_n(\widehat{X}_n, \widehat{Y}_n), \\ \widehat{V}_k = \max\left(h_k(\widehat{X}_k, \widehat{Y}_k), \mathbb{E}\left[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{Y}_k)\right]\right), \qquad 0 \leqslant k \leqslant n-1. \end{cases}$$

We denote $N_k = N_k^X \times N_k^Y$ the size of the quantization grid.

Again, we provide an a priori quadratic error for $\left\|V_k - \widehat{V}_k\right\|_2$, $k = 0, \ldots, n$.

Theorem 1.3.3. *If the payoff functions $(\psi_{t_k})_{k=0:n}$ are Lipschitz continuous with compactly supported (right) derivative. Then the quadratic error induced by the quantization approximation $(\widehat{X}_k, \widehat{Y}_k)$ is upper-bounded by*

$$\left\|V_k - \widehat{V}_k\right\|_2 \leqslant \left(\sum_{l=k}^{n-1} C_{W_{l+1}^f} \left\|W_{l+1}^f - \mathbb{E}[W_{l+1}^f \mid (X_l, Y_l)]\right\|_{2p}^2 + C_{W_{l+1}^d} \left\|W_{l+1}^d - \mathbb{E}[W_{l+1}^d \mid (X_l, Y_l)]\right\|_{2p}^2 \right.$$
$$\left. + C_{X_l} \left\|X_l - \widehat{X}_l\right\|_{2p}^2 + C_{Y_l} \left\|Y_l - \widehat{Y}_l\right\|_{2p}^2 \right)^{1/2}$$

where $1 < p < 3/2$ *and* $q \geqslant 1$ *such that* $\frac{1}{p} + \frac{1}{q} = 1$ *and the constants* $C_{X_l}, C_{W^d_{l+1}}, C_{Y_l}, C_{W^f_{l+1}}$ *are finite. So, by taking* $\bar{N} = \min N_k$*, we have*

$$\lim_{\bar{N}\to+\infty} \left\|V_k - \widehat{V}_k\right\|_2^2 = \sum_{l=k}^{n-1} C_{W^f_{l+1}} \left\|W^f_{l+1} - \mathbb{E}[W^f_{l+1} \mid (X_l, Y_l)]\right\|_{2p}^2 + C_{W^d_{l+1}} \left\|W^d_{l+1} - \mathbb{E}[W^d_{l+1} \mid (X_l, Y_l)]\right\|_{2p}^2.$$

We can thus notice that the approximation we made by replacing the preconditioning in (X_k, W^f_k, Y_k, W^d_k) by (X_k, Y_k), even if it considerably reduces the complexity of the problem, induces a systematic error. However, it seems reasonable to assume that this error is negligible.

Example 1.3.4. Indeed, in Figure 1.8, when pricing Bermudan options that can be exercised annually for maturities of 2, 5 or 10 years, considering market parameters for σ_d and σ_f, the price difference between the two methods is negligible. The payoff considered is that of the Example 1.3.1. For the example considered in the Figure 1.8, the correlations are assumed to be zero $\rho_{Sd} = \rho_{Sf} = \rho_{df} = 0$, $S_0 = 88.17$, $\sigma_S = 50\%$, $\sigma_d = \sigma_f = 50bp$ $(1bp = 0.01\%)$, $P_d(0,t) = \exp(-r_d t)$ with $r_d = 1.5\%$ and $P_f(0,t) = \exp(-r_f t)$ with $r_f = 1\%$.

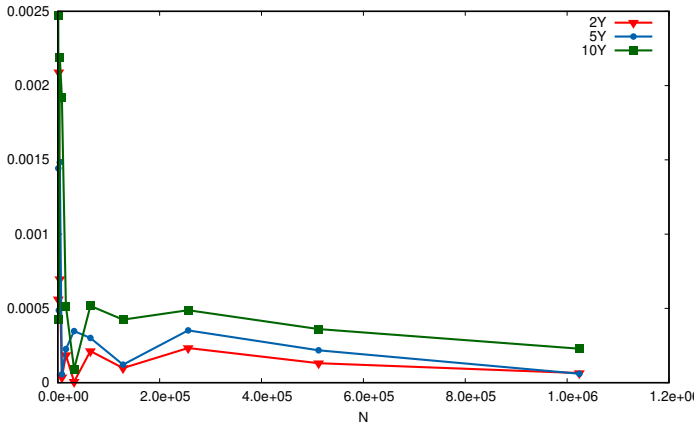


Fig. 1.8 *Relative difference between the prices given by the two methods for Bermudan options yearly exercisable and maturing at* 2, 5 *or* 10 *years.*

Chapter 2

Introduction - Français

2.1 Quantification Optimale

Cette thèse est consacrée à divers aspects théoriques de la quantification optimale en lien avec l'intégration numérique ainsi que diverses applications à la finance. La quantification optimale a initialement été introduite par Sheppard en 1897 dans [She97]. Ses travaux ont porté sur la quantification optimale de la distribution uniforme sur les hypercubes unités. Elle a ensuite été étendue à des lois plus générales à support compact ou non, motivé par des applications à la transmission du signal dans le Laboratoire Bell dans les années 50 (voir [GG82]). La quantification optimale est également liée à une méthode de statistique computationnelle d'apprentissage non-supervisé. En effet la méthode des "k-means" qui est une méthode de classification automatique non paramétrique consistant, étant donné un ensemble de points et un entier k, à diviser les points en k classes ("clusters") se base sur le même algorithme que la méthode de Lloyd utilisée pour construire un quantifieur optimal. La problème des "k-means" fût formulé par Steinhaus dans [Ste56] puis reprise quelques années plus tard par MacQueen dans [Mac67]. Dans les années 90, la quantification optimale fût d'abord utilisée à des fins d'intégration numérique pour l'approximation d'espérances, voir [Pag98], et plus tard utilisée pour l'approximation d'espérances conditionnelles : voir [BPP01; BP03; BPP05] pour des problèmes d'arrêt optimal appliqué à l'évaluation d'options américaines, [PP05; PRS05] pour des problèmes de filtrage non-linéaire, [BDD13; PCR09; PPP04a; PPP04b] pour des problèmes de contrôle stochastique, [Gob+05] pour la discrétisation et la simulation d'équations de Zakai et de McKean-Vlasov et [BSD12; DD12] en présence de processus de Markov déterministe par morceaux (PDMP).

2.1.1 Définitions et principaux résultats

Soit X un vecteur aléatoire à valeurs dans $\mathbb{R}^d$ muni d'une norme $|\cdot|$, ici toujours euclidienne, avec distribution $\mathbb{P}_X$, défini sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ tel que $X \in L^2_{\mathbb{R}^d}(\Omega, \mathcal{A}, \mathbb{P})$. La quantification de X consiste à approcher X par un vecteur aléatoire $q(X)$ où q est une

fonction borélienne à valeurs dans $\Gamma_N = \{x_1^N, \dots, x_N^N\} \subset \mathbb{R}^d$. De plus, on peut remarquer que $\mathrm{dist}(X, q(X)) \geqslant \mathrm{dist}(X, \Gamma_N)$ avec égalité si et seulement si q est une projection au plus proche voisin, notée $q = \mathrm{Proj}_{\Gamma_N}$. Cette projection au plus proche voisin Proj_{Γ_N} est associé biunivoquement à une partition borélienne de Voronoï $\big(C_i(\Gamma_N)\big)_{1\leqslant i\leqslant N}$ de $\mathbb{R}^d$ vérifiant

$$C_i(\Gamma_N) \subset \big\{\xi \in \mathbb{R}, |\xi - x_i^N| \leqslant \min_{j\neq i} |\xi - x_j^N|\big\}.$$

Ainsi, la projection au plus proche voisin associée est définie par

$$\mathrm{Proj}_{\Gamma_N}(\xi) = \sum_{i=1}^{N} x_i^N \, \mathbb{1}_{\xi\in C_i(\Gamma_N)}\,.$$

Une telle quantification est dite "*Voronoï*". Nous noterons $\widehat{X}^{\Gamma_N}$ la projection au plus proche voisin de X sur $\Gamma_N = \{x_1^N, \dots, x_N^N\}$, ainsi

$$\widehat{X}^{\Gamma_N} = \mathrm{Proj}_{\Gamma_N}(X).$$

On allégera la notation de $\widehat{X}^{\Gamma_N}$ en $\widehat{X}^N$ pour plus de clarté.

Ainsi, la loi d'un quantifieur $\widehat{X}^N$ est entièrement caractérisée par la grille des centroïdes $\Gamma_N = \{x_i^N,\ 1 \leqslant i \leqslant N\}$ dans laquelle le quantifieur prend ses valeurs et le N-uplet des poids p_i^N qui représentent la probabilité que $\widehat{X}^N$ soit égal à x_i^N ou, de façon équivalente, que X appartienne à la cellule de Voronoï i, i.e.

$$p_i^N = \mathbb{P}\left(\widehat{X}^N = x_i^N\right) = \mathbb{P}\left(X \in C_i(\Gamma_N)\right), \quad i = 1, \dots, N.$$

Dans cette thèse, nous travaillerons essentiellement avec de la quantification *optimale quadratique*. Le terme *optimale* provient du fait que l'on cherche la meilleure approximation de X dans le sens où l'on va vouloir minimiser la distance entre les vecteurs aléatoires X et $\widehat{X}^N$ en optimisant la grille Γ_N pour une taille N donnée. Cette distance est mesurée en norme L^2, d'où le terme *quadratique*. La distance entre X et $\widehat{X}^N$, que l'on note $\|X - \widehat{X}^N\|_2$, est appelée erreur de quantification moyenne. Mais on raisonne souvent en terme de distorsion qui n'est autre que le carré de l'erreur de quantification moyenne. Pour un N-uplet, elle est définie par

$$\mathcal{Q}_{2,\mathrm{N}} : x = (x_1^N, \dots, x_N^N) \longmapsto \mathbb{E}\left[\min_{i=1,\dots,N} |X - x_i^N|^2\right] = \|X - \widehat{X}^N\|_2^2.$$

Ainsi, nous cherchons la grille Γ_N de cardinal au plus N tel que le quantifeur $\widehat{X}^N = \mathrm{Proj}_{\Gamma_N}(X)$ minimise

$$\min_{\Gamma_N\subset\mathbb{R}^d, |\Gamma_N|\leqslant N} \|X - \widehat{X}^N\|_2^2.$$

Une telle grille existe toujours lorsque $X \in L^2$ (voir théorème 2.1.1 ci-après). Dans la Figure 2.1, nous présentons deux quantifieurs de taille $N = 100$ d'un vecteur gaussien centré et de matrice de covariance-variance unitaire. A gauche, nous représentons un échantillon i.i.d. du vecteur gaussien et à droite un quantifieur optimal. La couleur de chaque cellule représente la probabilité p_i^N associée à la cellule $C_i(\Gamma_N)$ de centroïde x_i^N (point rouge).

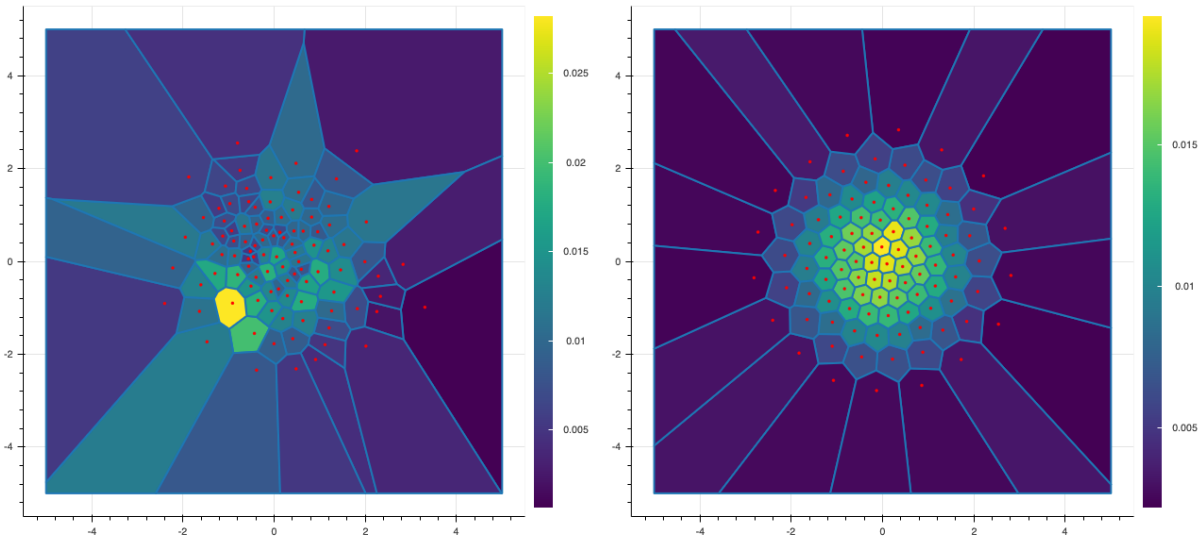


Fig. 2.1 *Deux quantifications de taille $N = 100$ d'un vecteur gaussien centré et de matrice de variance-covariance unitaire.*

Le problème de minimisation étant posé, plusieurs résultats ont été démontrés dans la littérature, voir par exemple les deux livres [GL00; Pag18] pour plus de détails sur la théorie de la quantification optimale. Signalons au passage que cette théorie peut être entièrement développée dans un cadre L^p et l'on parle alors de L^p-quantification optimale. Nous citons en premier un résultat assurant l'existence d'un quantifieur optimal.

Theorem 2.1.1. *(Existence d'un N-quantifieur optimal) Soit $X \in L^2_{\mathbb{R}^d}(\mathbb{P})$ et $N \in \mathbb{N}^\star$.*

(a) La fonction de distorsion quadratique $\mathcal{Q}_{2,\mathrm{N}}$ au niveau N atteint un minimum en (au moins) un N-uplet $x^\star = (x_1^N, \ldots, x_N^N)$ et la grille associée $\Gamma_N^\star = \{x_i^N, i = 1, \ldots, N\}$ est appelé un N-quantifieur quadratique optimal.

(b) Si le support de la distribution $\mathbb{P}_X$ de X a au moins N éléments, alors $x^\star = (x_1^N, \ldots, x_N^N)$ a des composantes deux à deux distinctes et $\mathbb{P}_X\left(C_i(\Gamma_N^\star)\right) > 0, i = 1, \ldots, N$. De plus, la suite $N \mapsto \inf_{x \in (\mathbb{R}^d)^N} \mathcal{Q}_{2,\mathrm{N}}(x)$ converge vers 0 et décroît strictement tant qu'elle est strictement positive.

En plus de savoir que la distorsion quadratique décroît vers 0, la vitesse exacte de convergence a été établi au travers des contributions de plusieurs auteurs [Zad82; BW82; GL00]. Le théorème a été démontré dans le cas L^p et donc caractérise l'erreur de quantification L^p.

Theorem 2.1.2. *(Théorème de Zador) Soit $d \in \mathbb{N}^\star$ et $p \in (0, +\infty)$.*

(a) Sharp rate. *Soit $X \in L^{p+\delta}_{\mathbb{R}^d}(\mathbb{P})$ avec $\delta > 0$. Soit $\mathbb{P}_X(d\xi) = \varphi(\xi) \cdot \lambda_d(d\xi) + \nu(d\xi)$, où $\nu \perp \lambda_d$ i.e. ν est singulière par rapport à la mesure de Lebesgue λ_d sur $\mathbb{R}^d$. Alors, il existe une constante $\widetilde{J}_{p,d} \in (0, +\infty)$ tel que*

$$\lim_{N \to +\infty} N^{1/d} \min_{\Gamma_N \subset \mathbb{R}^d, |\Gamma_N| \leqslant N} \|X - \widehat{X}^N\|_p = \widetilde{J}_{p,d} \left[\int_{\mathbb{R}^d} \varphi^{\frac{d}{d+p}} d\lambda_d \right]^{\frac{1}{p} + \frac{1}{d}}$$

où $\widehat{X}^N$ est un quantifieur L^p-optimal de X.

(b) Borne supérieur non-asymptotique [GL00; Pag18]. *Soit $\delta > 0$. Il existe une constante réelle $C_{d,p,\delta} \in (0, +\infty)$ tel que, pour tout vecteur aléatoire X à valeurs dans $\mathbb{R}^d$,*

$$\forall N \geqslant 1, \qquad \min_{\Gamma_N \subset \mathbb{R}^d, |\Gamma_N| \leqslant N} \|X - \widehat{X}^N\|_p \leqslant C_{d,p,\delta} \sigma_{\delta+p}(X) N^{-1/d}$$

où, pour $r \in (0, +\infty)$, $\sigma_r(X) = \min_{a \in \mathbb{R}^d} \|X - a\|_r < +\infty$.

2.1.2 Construction d'un quantifieur optimal

Il existe de nombreuses méthodes pour construire un quantifieur optimal. Dans certains cas très rares, les centroïdes sont donnés explicitement, par exemple lorsque $X \sim \mathcal{U}([a, b])$ où $a, b \in \mathbb{R}$, la grille Γ_N est donnée par

$$\Gamma_N = \{x_1^N, \ldots, x_N^N\} = \left\{ \frac{2i-1}{2N} : i = 1, \ldots, N \right\}.$$

Nous nous référons également à [GL00] pour la loi de Laplace et à [FP02] pour des formules semi-fermées pour la loi exponentielle, la loi puissance et la loi puissance inverse. Néanmoins, la plupart du temps, ce n'est pas le cas donc nous devons utiliser des méthodes itératives pour construire les grilles et les poids associés à chacun des centroïdes. Ces méthodes itératives se divisent en deux grandes familles : les méthodes déterministes (algorithme de Lloyd, algorithme de Newton-Raphson et leurs variantes, ...) qui se basent sur la connaissance explicite de la densité et la fonction de répartition de la loi de X et les méthodes à base d'optimisation stochastique (Competitive Learning Vector Quantization (CLVQ), randomisation de l'algorithme de Lloyd, ...) nécessitant seulement de pouvoir simuler X. Ces méthodes sont détaillées dans le Chapitre 3.

Cas d'une variable aléatoire réelle - $d = 1$. Dans le cas unidimensionnel, nous avons un résultat d'unicité du quantifieur optimal lorsque la densité de X est log-concave. Ce théorème a été démontré par Kieffer dans [Kie82] (voir aussi [Pag98]).

Si X est une variable aléatoire ($d = 1$) dont on connait le premier moment partiel $K_X(\cdot)$ et la fonction de répartition F_X de X

$$K_X(x) := \mathbb{E}[X\,\mathbb{1}_{X\leqslant x}] \qquad \text{et} \qquad F_X(x) := \mathbb{P}(X \leqslant x),$$

alors on utilise en priorité les méthodes déterministes qui permettent de construire très rapidement un quantifieur optimal de X, tel que l'algorithme de Lloyd introduit dans [Llo82] qui est un algorithme de recherche de point fixe. Il est également possible d'appliquer l'algorithme de Newton-Raphson en calculant la Hessienne de la fonction de distorsion quadratique (voir [PP03] pour un exemple détaillé appliqué à une variable aléatoire normale). D'autres descentes de gradient déterministes peuvent être utilisées tel que Levenberg-Marquardt ou des méthodes de quasi-Newton. Sinon, on utilise les méthodes à base d'optimisation stochastique telles que la version stochastique de l'algorithme de Lloyd ou une descente de gradient stochastique (voir [Pag98]).

Example 2.1.3. Dans la Figure 2.2, nous représentons en bleu la densité d'une variable aléatoire gaussienne unidimensionnelle et en rouge les centroïdes du quantifieur optimale de taille $N = 11$ de cette même variable aléatoire. Nous illustrons également ce que représentent les poids p_i^N associés aux centroïdes x_i^N. De plus, nous pouvons approcher la densité (si elle existe) en chaque point de la grille par la relation suivante

$$f(x_i^N) \approx \frac{2p_i^N}{x_{i+1/2}^N - x_{i-1/2}^N}.$$

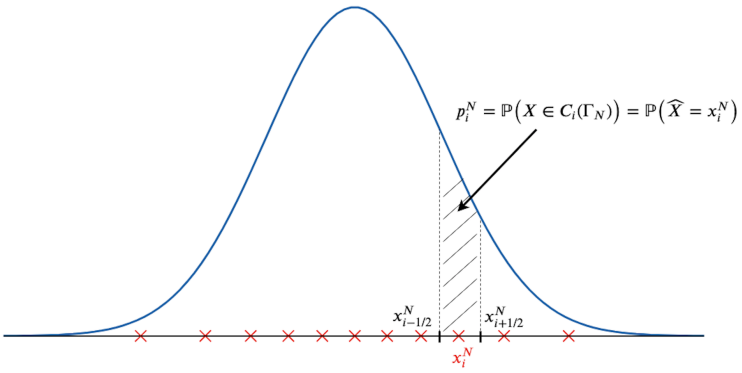


Fig. 2.2 *Densité d'une gaussienne centrée réduite* $\mathcal{N}(0,1)$ *en bleu et les centroïdes d'un quantifieur optimale de taille* $N = 11$ *en rouge.*

Cas d'un vecteur aléatoire - $d \geqslant 2$. Maintenant, considérons un vecteur aléatoire X à valeurs dans $\mathbb{R}^d$ ($d \geqslant 2$). Deux approches existent pour construire un quantifieur optimal de la loi de X.

La première approche consiste à appliquer la méthodologie développée dans le cas scalaire directement au cas vectoriel et obtenir ainsi une quantification optimale de X. Si l'on connait la densité de X alors il est encore possible en dimension 2 ou 3 d'appliquer les méthodes déterministes (cf. Chapitre 3). Cependant dès $d \geqslant 4$, nous ne pouvons plus guère compter que sur des méthodes d'optimisation stochastique fondées sur la simulation d'échantillons de la loi de X.

La seconde, la quantification produit, consiste à construire un quantifieur optimal de chacune des composantes du vecteur aléatoire et ensuite de construire le quantifieur en considérant le produit cartésien entre toutes les composantes quantifiées optimalement. Plus précisément, soit $X = (X^\ell)_{\ell=1:d}$, un vecteur aléatoire à valeurs dans $\mathbb{R}^d$. On considère les d quantifieurs optimaux unidimensionnels $\widehat{X}^\ell$ de taille N^ℓ de chacune des marginales X^ℓ. Chaque quantifieur $\widehat{X}^\ell$ prend ses valeurs dans la grille $\Gamma_\ell^{N_\ell} = \{z_{i_\ell}^\ell, i_\ell \in \{1, \cdots, N_\ell\}\}$. Ainsi, le quantifieur produit de X prend ses valeurs dans la grille Γ^N qui est le produit cartésien des grilles unidimensionnelles, i.e. $\Gamma^N = \prod_{\ell=1}^d \Gamma_\ell^{N_\ell}$ de taille $N = N^1 \times \cdots \times N^d$ ou, de façon équivalente,

$$\Gamma^N = \{(x_{i_1}^1, \cdots, x_{i_\ell}^\ell, \cdots, x_{i_d}^d), \quad i_\ell \in \{1, \cdots, N_\ell\}, \quad \ell \in \{1, \cdots, d\}\}.$$

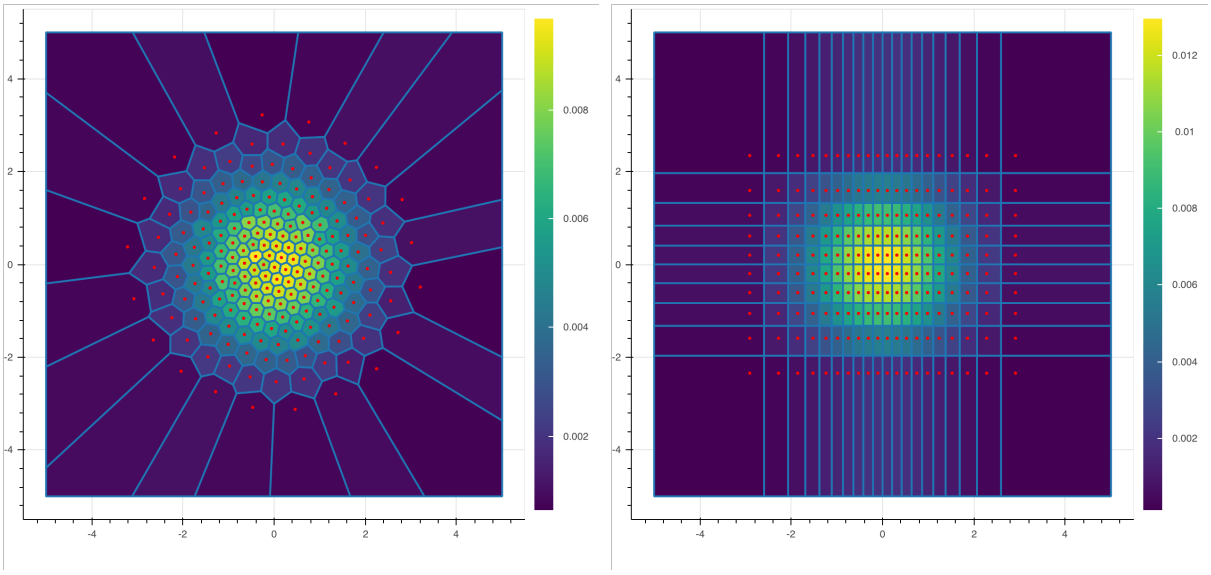


Fig. 2.3 *Deux quantifications de taille $N = 200$ d'un vecteur gaussien centré et de matrice de variance-covariance unitaire. Quantification optimale à gauche et quantification produit à droite.*

Dans la Figure 2.3, nous comparons la quantification optimale et la quantification produit d'un vecteur gaussien centré et de matrice de variance-covariance unitaire. Les deux méthodes

ont leurs avantages et leurs inconvénients, la première méthode produit une meilleure quantification du vecteur aléatoire X comparée à la quantification-produit mais le coût numérique induit pour la construction d'un quantifieur optimal est souvent beaucoup plus élevé.

Cas des diffusions. Si maintenant, au lieu de considérer un vecteur aléatoire, nous nous intéressons aux diffusions, i.e.

$$dX_t = b(t, X_t)dt + \sigma(t, W_t)dW_t$$

alors il existe, là encore, plusieurs solutions pour quantifier X_t. Plus précisément, étant donné une discrétisation en temps à n-pas $(t_k)_{0\leqslant k\leqslant n}$, nous cherchons les quantifieurs $\widehat{X}_{t_k}^{N_k}$ de taille N_k de X_{t_k} que nous noterons $\widehat{X}_k^{N_k}$ et X_k afin d'alléger les notations. L'objet que l'on cherche à construire est dénommé *arbre de quantification*. Un arbre est caractérisé par la connaissance des lois $\left(\Gamma_k, (p_i^k)_{1\leqslant i\leqslant N_k}\right)$ des quantifeurs $(\widehat{X}_k)_{0\leqslant k\leqslant n}$ et des probabilités de transition $p_{i,j}^k$

$$\mathbb{P}\left(\widehat{X}_{k+1} = x_j^{k+1} \mid \widehat{X}_k = x_i^k\right).$$

Nous ne présenterons pas toutes les approches existantes qui permettent d'aborder le problème de schémas quantifiés de discrétisation d'une diffusion mais seulement celles qui nous permettent d'utiliser des méthodes numériques déterministes d'optimisation des grilles. Pour les autres approches, basées sur des algorithmes stochastiques nous renvoyons à la série de papiers [BPP01; BP03].

Quantification des lois marginales. Le problème de la quantification d'une diffusion a été initié et développé dans une série d'articles [PPP04b; BPP05; BBP09; BBP10; CFG19]. Si X_k peut être simulé de façon exacte, c'est à dire sans l'aide d'un schéma de discrétisation en temps, et que nous connaissons la loi marginale de X_k, à chaque instant t_k, alors nous sommes ramenés au cas de la quantification d'un vecteur aléatoire. En effet, nous pouvons quantifier optimalement chaque vecteur aléatoire X_k à l'aide de méthodes numériques déterministes si $d \leqslant 2$, ce qui produit un arbre de quantification optimal, ou quantifier optimalement chacune de ses composantes pour ensuite construire une quantification produit des X_k, produisant un arbre de quantification produit.

Example 2.1.4. Si l'on considère un modèle de Black-Scholes à volatilité constante σ et avec taux d'intérêts constants r

$$dS_t = S_t(rdt + \sigma dW_t), \quad \text{avec} \quad S_0 = s_0,$$

alors nous avons une forme explicite pour S_t

$$S_t = S_0\, \mathrm{e}^{(r-\sigma^2/2)t+\sigma W_t}$$

donc pour un instant donné t, $\log(S_t/S_0) \sim \mathcal{N}\left((r-\sigma^2/2)t, \sigma^2 t\right)$ donc nous pouvons quantifier optimalement S_t à chaque instant qui nous intéresse à l'aide de méthodes déterministes (cf. Chapitre 3). Nous pouvons également quantifier le Brownien W_t qui est "plus universel".

Quantification récursive. Dans le cas où nous ne connaissons pas la loi marginale de X_k et que sommes obligés d'utiliser un schéma de discrétisation (type Euler-Maruyama, Milstein, ...), nous allons utiliser une méthode appelée *quantification récursive*. La quantification récursive (aussi appelée quantification Markovienne) a d'abord été introduite dans [PPP04b] puis étudiée en profondeur dans [PS15] pour le cas d'une diffusion unidimensionnelle discrétisée par un schéma d'Euler-Maruyama. Un algorithme rapide fondé sur des méthodes déterministes pour construire l'arbre de quantification y est développé et analysé. Par la suite, la quantification récursive rapide a été étendue à des schémas unidimensionnels d'ordre supérieur par [McW+18] et à des dimensions supérieures par quantification de produit (voir [PS18b; FSP18; Rud+17; CFG18; CFG17]). Cette méthode consiste à construire récursivement en k les quantifeurs $\widehat{X}_k^{N_k}$ via la récursion

$$\widehat{X}_k^{N_k} = \mathrm{Proj}_{\Gamma_{N_k}}\left(\widetilde{X}_k\right) \quad \text{avec} \quad \widetilde{X}_k = \mathcal{E}_{k-1}\left(\widehat{X}_{k-1}^{N_{k-1}}, Z_k\right)$$

où $\mathcal{E}_{k-1}$ est un schéma de discrétisation.

2.2 Intégration numérique

Un problème courant en pratique est de calculer l'espérance d'une fonction de X lorsque X est une variable ou un vecteur aléatoire, c'est à dire $\mathbb{E}\left[f(X)\right]$. Or, sauf dans des cas très particuliers, il n'est pas possible de calculer explicitement cette quantité, c'est le cas par exemple si $X = X_T$ la valeur d'une diffusion à la date T. C'est pourquoi il est nécessaire de faire appel à des méthodes d'intégration numérique. [Pag98] introduit une méthode de cubature à base de quantification optimale afin de pouvoir approcher des espérances de la forme $\mathbb{E}\left[f(X)\right]$. Considérons $\widehat{X}^N$ un quantifieur optimale de X, le fait que $\widehat{X}^N$ soit discret nous permet de définir facilement la formule de cubature suivante

$$\mathbb{E}\left[f(\widehat{X}^N)\right] = \sum_{i=1}^{N} p_i^N f(x_i^N). \tag{2.1}$$

De plus, étant donné que $\widehat{X}^N$ a été construite comme étant la meilleur approximation discrète de X de cardinal au plus N alors il nous semble raisonnable de penser que $\mathbb{E}\left[f(\widehat{X}^N)\right]$ est une bonne approximation de $\mathbb{E}\left[f(X)\right]$.

Dans le Chapitre 4, tiré de l'article "New Weak Error bounds and expansions for Optimal Quantization" publié dans *Journal of Computational and Applied Mathematics*, voir [LMP19], nous présentons de nouveaux résultats dans le cas réel concernant l'erreur induite par l'approximation à base de quantification de l'espérance $\mathbb{E}\left[f(X)\right]$. Ce travail est un travail

commun avec Vincent Lemaire et Gilles Pagès et il est accessible sur arXiv ou HAL. Ces résultats “faible” sont résumés ci-après.

2.2.1 Convergence faible

Dans la première partie du Chapitre 4, nous nous intéressons à la vitesse de convergence de $\mathbb{E}\big[f(\widehat{X}^N)\big]$ vers $\mathbb{E}\big[f(X)\big]$ en fonction de N pour différentes classes de fonctions f lorsque X est une variable aléatoire à valeurs dans $\mathbb{R}$, i.e nous recherchons le plus grand $\alpha > 0$ tel que, pour toute fonction f dans cette classe $\mathcal{F}$,

$$\overline{\lim_N} N^\alpha \big|\mathbb{E}\big[f(X)\big] - \mathbb{E}\big[f(\widehat{X}^N)\big]\big| \leqslant C_{f,X} < +\infty.$$

Si l'on majore de façon naïve l'erreur faible par l'erreur forte le long des fonctions lipschitziennes, on obtient la majoration suivante (avec $\alpha = 1$) pour une suite de N-quantifieurs L^2-optimaux

$$N\big|\mathbb{E}\big[f(X)\big] - \mathbb{E}\big[f(\widehat{X}^N)\big]\big| \leqslant N[f]_{Lip}\|X - \widehat{X}^N\|_1 \leqslant N[f]_{Lip}\|X - \widehat{X}^N\|_2 \xrightarrow{N\to+\infty} C_f < +\infty$$

où le Théorème de Zador (théorème 2.1.2) a été utilisé. De plus, si nous considérons $f(x) = \mathrm{dist}(x, \Gamma_N)$ alors f est une fonction lipschitzienne et nous avons

$$N\big|\mathbb{E}\big[f(X)\big] - \mathbb{E}\big[f(\widehat{X}^N)\big]\big| = N\|X - \widehat{X}^N\|_1 \leqslant N\|X - \widehat{X}^N\|_2 \xrightarrow{N\to+\infty} C_f < +\infty.$$

Pour certaines classes de fonctions nous pouvons démontrer que la formule de cubature induit une erreur faible d'ordre 2 ($\alpha = 2$). Par exemple, si nous considérons les fonctions dérivables avec une dérivée lipschitzienne alors nous avons une erreur d'ordre 2, voir [Pag98]. En effet, nous utilisons un développement de Taylor avec reste intégral de la forme

$$f(x) = f(y) + f'(y)(x-y) + \int_0^1 \big(f'(tx + (1-t)y) - f'(y)\big)(x-y)dt$$

et la propriété de stationnarité d'un quantifieur quadratique optimale suivante

$$\mathbb{E}\big[X \mid \widehat{X}^N\big] = \widehat{X}^N.$$

Le premier terme du développement de Taylor vaut zéro car

$$\mathbb{E}\big[f'(\widehat{X}^N)(X - \widehat{X}^N)\big] = \mathbb{E}\Big[f'(\widehat{X}^N)\,\mathbb{E}\big[X - \widehat{X}^N \mid \widehat{X}^N\big]\Big] = \mathbb{E}\Big[f'(\widehat{X}^N)\big(\mathbb{E}\big[X \mid \widehat{X}^N\big] - \widehat{X}^N\big)\Big] = 0.$$

Ainsi, en utilisant la propriété de Lipschitz de la dérivée et le théorème de Zador, nous obtenons une erreur faible d'ordre 2 comme convenu

$$N^2\big|\,\mathbb{E}\big[f(X)\big]-\mathbb{E}\big[f(\widehat{X}^N)\big]\big| \leqslant N^2\int_0^1 \mathbb{E}\big[\big|f'(tX+(1-t)\widehat{X}^N)-f'(\widehat{X}^N)\big||X-\widehat{X}^N|\big]dt$$
$$\leqslant \frac{[f']_{Lip}}{2}N^2\|X-\widehat{X}^N\|_2^2 \xrightarrow{N\to+\infty} C_f<+\infty.$$

Dans la première partie du Chapitre 4, nous étendons ces résultats concernant la vitesse de convergence de l'erreur faible d'ordre supérieur à 1 à une plus vaste classe de fonctions ayant moins de régularité, plus précisément, les fonctions qui sont soit :

- continues et affines par morceaux avec un nombre fini de ruptures d'affinités,
- Lipschitz convexe,
- dérivables avec une dérivée définie par morceaux (nombre fini de morceaux K aux points $\{a_1,\dots,a_K\}$ tel que $-\infty=a_0<a_1<\cdots<a_K<a_{K+1}=+\infty$) qui soit localement lipschitzienne, c'est à dire

  $$\forall k=0,\dots,K,\quad \forall x,y\in(a_k,a_{k+1})\quad |f'(x)-f'(y)|\leqslant [f']_{k,Lip,loc}|x-y|\big(g_k(x)+g_k(y)\big)$$

  où $g_k:(a_k,a_{k+1})\to\mathbb{R}_+$ sont des fonctions boréliennes à valeurs positives,
- dérivables avec une dérivée définie par morceaux (nombre fini de morceaux K aux points $\{a_1,\dots,a_K\}$ tel que $-\infty=a_0<a_1<\cdots<a_K<a_{K+1}=+\infty$) qui soit localement α-Hölder, c'est à dire

  $$\forall k=0,\dots,K,\quad \forall x,y\in(a_k,a_{k+1}),\quad |f'(x)-f'(y)|\leqslant [f']_{k,\alpha,loc}|x-y|^\alpha\big(g_k(x)+g_k(y)\big)$$

  où $g_k:(a_k,a_{k+1})\to\mathbb{R}_+$ sont des fonctions boréliennes à valeurs positives.

Pour les trois premières classes de fonctions, nous démontrons que l'erreur faible est d'ordre 2 et pour la dernière, d'ordre $1+\alpha$.

Dans la partie numérique, nous illustrons ce résultat en évaluant le prix d'un *Call* européen dans un modèle de Black-Scholes donné par

$$I_0:=\mathbb{E}\big[\,\mathrm{e}^{-rT}(S_T-K)_+\big]$$

où $S_t=S_0\,\mathrm{e}^{(r-\sigma^2/2)t+\sigma W_t}$ avec $(W_t)_{t\in[0,T]}$ un mouvement brownien. Afin d'approcher, à l'aide de la quantification, le prix du *Call* européen nous pouvons réécrire I_0 de deux façon différentes

$$I_0=\mathbb{E}\big[\varphi(S_T)\big]=\mathbb{E}\big[f(W_T)\big]$$

où φ est une fonction affine par morceaux avec une rupture d'affinité et f est une fonction dérivable avec une dérivée définie par morceaux localement lipschitzienne. Ainsi, en considérant des quantifieurs de S_T ou W_T et en utilisant la formule de cubature, nous observons, pour les deux approximations, une erreur faible d'ordre 2.

2.2.2 Développement d'erreur faible d'ordre supérieur

Dans la seconde partie du Chapitre 4, nous nous intéressons au développement d'erreur faible de l'approximation de $\mathbb{E}\big[f(X)\big]$ par $\mathbb{E}\big[f(\widehat{X}^N)\big]$. C'est à dire que nous cherchons à obtenir un développement de la forme

$$\mathbb{E}\big[f(X)\big] = \mathbb{E}\big[f(\widehat{X}^N)\big] + \frac{c_2}{N^2} + O(N^{-(2+\beta)})$$

où $\beta \in (0, 1]$. Dans la section précédente, nous avons déjà montré que l'approximation par formule de cubature à base de quantification optimale induit un terme d'erreur d'ordre $O(N^{-2})$ dans le meilleur des cas. Ici, nous cherchons à raffiner les résultats précédents afin d'obtenir un développement d'erreur à l'ordre 2 "contrôlé" et non une simple vitesse de convergence à l'ordre 2.

Dans la Section 4.3, nous démontrons que ce développement existe si la fonction $f : \mathbb{R} \to \mathbb{R}$ est deux fois dérivables avec une dérivée seconde lipschitzienne. Ce résultat utilise un développement de Taylor d'ordre 2 avec reste intégral de la forme

$$f(x) = f(y) + f'(y)(x-y) + \frac{1}{2}f''(y)(x-y)^2 + \int_0^1 (1-t)\big(f''(tx+(1-t)y) - f''(y)\big)(x-y)^2 dt$$

où l'on prend l'espérance de chaque côté de l'égalité et on remplace x et y par X et $\widehat{X}^N$, respectivement. Le deuxième terme à droite est annulé en utilisant la propriété de stationnarité du quantifieur quadratique optimal. Pour le troisième terme, nous nous appuyons sur [Del+04] (Théorème 6) qui stipule que $\forall g : \mathbb{R} \to \mathbb{R}$ tel que $\mathbb{E}\big[g(X)\big] < +\infty$

$$\lim_N N^2\, \mathbb{E}\big[g(\widehat{X}^N)|X - \widehat{X}^N|^2\big] = Q_2(\mathbb{P}_X) \int g(\xi)\, \mathbb{P}_X(d\xi)$$

que l'on applique à $g = f''$ où $Q_2(\mathbb{P}_X)$ est la constante de Zador. Ainsi, nous avons déjà les deux premiers termes dans le développement de l'erreur. Pour le dernier terme, on utilise la propriété de Lipschitz de la dérivée seconde et le reste de la preuve se fonde principalement sur un résultat initialement établi dans [GLP08] puis récemment étendu dans [PS18a], connu sous le nom de "L^r-L^s distorsion mismatch", qui se formule ainsi : que peut-on dire du taux de convergence de $\mathbb{E}\big[|X - \widehat{X}^N|^s\big]$ sachant que $\widehat{X}^N$ est un quantifieur L^r-optimal lorsque $s > r$ et $X \in L^s$? Nous citons ce théorème pour $d = 1$, qui est le cas qui nous intéresse.

Theorem 2.2.1 (L^r-L^s-distorsion mismatch)**.** *Soit $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ une variable aléatoire et $r \in (0, +\infty)$. Soit $\mathbb{P}_X(d\xi) = \varphi(\xi) \cdot \lambda(d\xi) + \nu(d\xi)$, où $\nu \perp \lambda$ i.e. ν est singulier par rapport à la mesure de Lebesgue λ sur $\mathbb{R}$ et φ est non-identiquement nul. Soit $(\Gamma_N)_{N\geqslant 1}$ une suite de grilles L^r-optimales et $s \in (r, r+1)$. Si*

$$X \in L^{\frac{s}{1+r-s}+\delta}(\mathbb{P})$$

pour un $\delta > 0$, alors

$$\limsup_N N\|X - \widehat{X}^N\|_s < +\infty.$$

Ainsi, en appliquant ce théorème avec $r = 2$ et $s = 2 + \beta$, nous obtenons un $O(N^{-(2+\beta)})$ pour le dernier terme et $\forall\, \beta \in (0, 1)$, nous avons l'expansion suivante

$$\mathbb{E}\big[f(X)\big] = \mathbb{E}\big[f(\widehat{X}^N)\big] + \frac{c_2}{N^2} + O(N^{-(2+\beta)}).$$

Ce développement d'erreur nous permet de justifier théoriquement l'usage d'extrapolation de Richardson-Romberg qui a pour but de *tuer* le premier terme d'erreur du développement en combinant linéairement deux formules de cubature par quantification, respectivement à N et M points, i.e.

$$\mathbb{E}\big[f(X)\big] = \mathbb{E}\left[\frac{M^2 f(\widehat{X}^M) - N^2 f(\widehat{X}^N)}{M^2 - N^2}\right] + O(N^{-(2+\beta)})$$

pour $M = kN$ avec $k > 1$.

Nous illustrons ce résultat dans la partie numérique en évaluant une option européenne sur *spread* dans un modèle de Black-Scholes en dimension 2 dont le prix est donné par

$$I_0 := \mathbb{E}\big[\,\mathrm{e}^{-rT}(S_T^1 - S_T^2 - K)_+\big].$$

En pré-conditionnant, nous exprimons I_0 comme suit

$$I_0 = \mathbb{E}\big[\varphi(Z_2)\big]$$

où Z_2 est une gaussienne centrée réduite et φ est une fonction deux fois dérivables avec une dérivée seconde lipschitzienne. Ainsi, en considérant des N-quantifieurs optimaux $\widehat{Z}^N$ de $Z_2 \sim \mathcal{N}(0, 1)$, nous approchons I_0 à l'aide de la formule de cubature à base de quantification optimale (2.1) et observons une erreur faible d'ordre 2. De plus, en utilisant l'extrapolation de Richardson-Romberg, nous atteignons une erreur faible d'ordre 3.

Cependant, l'intérêt de la méthode de cubature par quantification optimale lorsque $d = 1$ reste limité car elle est notamment en compétition avec les méthodes utilisant des points de Gauss. Une extension multi-dimensionelle est en revanche très utile dès que $d \geqslant 3$. On considère

une fonction deux fois différentiables $f : \mathbb{R}^d \mapsto \mathbb{R}$ avec une Hessienne bornée et lipschitzienne. De plus, nous supposons que $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}^d$ a des composantes indépendantes X_k, $k = 1, \dots, d$ et que le quantifieur $\widehat{X}^N$ est un quantifieur produit de X à d composantes $(\widehat{X}_k^{N_k})_{k=1,\dots,d}$ tel que $N_1 \times \cdots \times N_d = N$. Ainsi, nous avons

$$\mathbb{E}\big[f(X)\big] = \mathbb{E}\big[f(\widehat{X}^N)\big] + \sum_{k=1}^{d} \frac{c_k}{N_k^2} + O\bigg(\Big(\min_{k=1:d} N_k\Big)^{-(2+\beta)}\bigg).$$

2.2.3 Réduction de variance

Dans la dernière partie du Chapitre 4, nous présentons une nouvelle méthode de réduction de variance d'un estimateur Monte Carlo avec des variables de contrôle à base de quantification optimale unidimensionnelle. D'autres méthodes de réduction de variance à base de quantification optimale ont été développées, voir par exemple [CP15; Pag18] pour plus de détails. Cette approche est motivée par la vitesse de convergence d'ordre 2 de l'erreur faible induite par la formule de cubature à base de quantification pour diverses classes de fonctions, notamment celles évoquées ci-avant.

Le problème. Soit $(Z_k)_{k=1,\dots,d} = Z \in L^2_{\mathbb{R}^d}(\mathbb{P})$ un vecteur aléatoire et une fonction $f : \mathbb{R}^d \to \mathbb{R}$. Nous nous intéressons à la quantité suivante

$$I := \mathbb{E}\big[f(Z)\big]. \tag{2.2}$$

Bien souvent, nous ne pouvons pas calculer explicitement cette quantité, c'est pourquoi une approche standard est de faire appel à un estimateur Monte Carlo $\bar{I}_M := \frac{1}{M}\sum_{m=1}^{M} f(Z^m)$ en simulant des copies indépendantes Z^m de Z pour approcher I. La convergence de la méthode et sa vitesse sont déterminées par la loi forte des grands nombres et le théorème central limite, respectivement, qui assurent, si Z est de carré intégrable, que

$$\bar{I}_M \xrightarrow{p.s.} \mathbb{E}\big[f(Z)\big] \quad \text{et} \quad \sqrt{M}\Big(\bar{I}_M - \mathbb{E}\big[f(Z)\big]\Big) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \sigma^2_{f(Z)}\right) \quad \text{lorsque} \quad M \to +\infty$$

où $\sigma^2_{f(Z)} = \mathbb{V}\mathrm{ar}\big(f(Z)\big)$. On remarque que, pour une taille de simulation M donnée, le facteur limitant de la méthode est $\sigma^2_{f(X)}$, c'est pourquoi des méthodes de réduction de variance qui consistent à réduire la valeur de $\sigma^2_{f(X)}$ pour accélérer la convergence de l'estimateur Monte Carlo vers I ont été développées. Le lecteur peut se référer à [Pag18; Gla13] pour plus de détails sur la simulation de Monte Carlo et les méthodes de réduction de variance en général telles que les variables de contrôle, la méthode antithétique, la stratification, l'échantillonnage préférentiel, ...

Une nouvelle méthode de réduction de variance par variable de contrôle quantifiée. Soit Ξ^N, un vecteur aléatoire à valeurs dans $\mathbb{R}^d$ défini par

$$\Xi^N := (\Xi_k^N)_{k=1,\dots,d},$$

qui sera notre variable de contrôle d-dimensionnelle, chaque composante Ξ_k^N est donnée par

$$\Xi_k^N := f_k(Z_k) - \mathbb{E}\left[f_k(\widehat{Z}_k^N)\right],$$

où $f_k(z) := f(\mathbb{E}[Z_1], \dots, \mathbb{E}[Z_{k-1}], z, \mathbb{E}[Z_{k+1}], \dots, \mathbb{E}[Z_d])$ et $\widehat{Z}_k^N$ est une quantification optimale de taille N de Z_k. Nous utilisons ici une quantification optimale unidimensionnelle afin de tirer profit des résultats d'erreur faible précédemment démontrés, en effet les fonctions $f_k : \mathbb{R} \to \mathbb{R}$ font parties des classes de fonctions nous permettant d'atteindre une erreur faible d'ordre 2. On introduit $I^{\lambda,N}$ comme approximation pour (2.2)

$$I^{\lambda,N} = \mathbb{E}\left[f(Z) - \langle \lambda, \Xi^N \rangle\right] = \mathbb{E}\left[f(Z) - \sum_{k=1}^{d} \lambda_k f_k(Z_k)\right] + \sum_{k=1}^{d} \lambda_k \, \mathbb{E}\left[f_k(\widehat{Z}_k^N)\right] \tag{2.3}$$

où $\lambda \in \mathbb{R}^d$. Les termes $\mathbb{E}\left[f_k(\widehat{Z}_k^N)\right]$ dans (2.3) peuvent être aisément et rapidement calculés en utilisant le caractère discret des quantifieurs.

À ce stade, on peut définir $\widehat{I}_M^{\lambda,N}$ l'estimateur de Monte Carlo associé à $I^{\lambda,N}$

$$\widehat{I}_M^{\lambda,N} = \frac{1}{M} \sum_{m=1}^{M} \left(f(Z^m) - \sum_{k=1}^{d} \lambda_k f_k(Z_k^m) \right) + \sum_{k=1}^{d} \lambda_k \, \mathbb{E}\left[f_k(\widehat{Z}_k^N)\right].$$

Il est important de remarquer que nous introduisons un biais en utilisant une telle variable de contrôle, en effet pour tout $k \in \{1, \dots, n\}$, $\mathbb{E}[\Xi_k^N] \neq 0$ car $\mathbb{E}\left[f_k(\widehat{Z}_k^N)\right]$ est une approximation de $\mathbb{E}\left[f_k(Z_k)\right]$. Cependant, la quantité qui nous intéresse réellement n'est pas le biais induit par l'estimateur $\widehat{I}_M^{\lambda,N}$ mais plutôt l'erreur quadratique moyenne ou Mean Squared Error (MSE) nous donnant une décomposition biais-variance

$$\mathrm{MSE}(\widehat{I}_M^{\lambda,N}) = \underbrace{\left(\sum_{k=1}^{d} \lambda_k \Big(\mathbb{E}\left[f_k(\widehat{Z}_k^N)\right] - \mathbb{E}\left[f_k(Z_k)\right] \Big) \right)^2}_{biais^2} + \frac{1}{M} \underbrace{\mathbb{V}\mathrm{ar}\left(f(Z) - \sum_{k=1}^{d} \lambda_k f_k(Z_k) \right)}_{Variance\ du\ Monte\ Carlo}.$$

Ainsi, nous pouvons prendre des valeurs de N plus élevées pour rendre le terme de biais négligeable comparé à la variance de l'estimateur tout en contrôlant le coût total induit par l'estimateur Monte Carlo. En pratique, nous n'avons pas besoin de prendre des valeurs très élevées pour N. En effet, le terme de biais converge vers 0 comme N^{-4} si f appartient à la bonne classe de fonctions, donc prendre des quantificateurs optimaux de taille 200 est largement

suffisant pour rendre le biais négligeable comparé à la variance de l'estimateur de Monte Carlo. Nous développons ce point dans la troisième partie du Chapitre 4.

Dans la partie numérique du Chapitre 4, nous appliquons la méthode de réduction de variance à l'évaluation d'une option panier dans un modèle de Black-Scholes en dimension d. La variable de contrôle nous permet de diviser la variance de l'estimateur Monte Carlo par 100 en petite dimension ($d = 2$ ou $d = 3$) et par 6 en plus grande dimension ($d = 10$). Nous observons également que le biais induit par la quantification devient négligeable pour des grilles dont la taille est supérieure à 100 ($N > 100$).

2.3 Exemples d'applications à la finance

2.3.1 Modèle d'Heston Stationnaire

Dans le Chapitre 5, nous nous intéressons au modèle d'Heston stationnaire et plus précisément à l'évaluation d'options européennes, bermudiennes et barrières dans ce modèle ainsi qu'à la calibration du modèle. Le Chapitre 5 est tiré du preprint "Stationary Heston model: Calibration and Pricing of exotics using Product Recursive Quantization" accessible sur arXiv ou HAL (voir [LMP20]). Cet article est un travail commun avec Vincent Lemaire et Gilles Pagès.

Le modèle d'Heston standard fut introduit à l'origine par Heston dans [Hes93]. C'est un modèle à volatilité stochastique où la condition initiale de la volatilité est supposée déterministe. Ce modèle a acquis une forte popularité principalement pour les deux raisons suivantes : c'est un modèle à volatilité stochastique donc il introduit un *smile* dans la surface de la volatilité implicite telle qu'observée dans le marché et la fonction caractéristique de ce modèle est donnée par formule semi-fermée ce qui nous permet d'évaluer les options européennes (*Call* & *Put*) presque instantanément (voir Carr & Madan dans [CM99]). Cependant, une remarque souvent faite sur ce modèle concerne le *smile* de volatilité implicite qui n'est pas assez pentu pour des maturités courtes comparé à ce que l'on observe sur le marché (voir [Gat11]). En remarquant que le processus de volatilité est ergodique avec une distribution invariante unique $\nu = \Gamma(\alpha, \beta)$ où les paramètres α et β dépendent des paramètres de diffusion de la volatilité, il a été proposé par Pagès & Panloup dans [PP09] de considérer directement que le processus évolue sous son régime stationnaire au lieu de le démarrer au temps 0 à partir d'une valeur déterministe. Ce choix a pour effet d'accentuer le *smile* de volatilité des maturités courtes tout en gardant le même comportement que le modèle standard pour les maturités plus longues. Plus tard, le comportement à court et long terme de la volatilité implicite générée par un tel modèle a été étudié par Jacquier & Shi dans [JS17].

Ainsi, la diffusion du couple actif-volatilité $(S_t^{(\nu)}, v_t^\nu)$ dans le modèle d'Heston stationnaire est défini par

$$\begin{cases} \dfrac{dS_t^{(\nu)}}{S_t^{(\nu)}} = (r-q)dt + \sqrt{v_t^\nu}\big(\rho d\widetilde{W}_t + \sqrt{1-\rho^2}dW_t\big) \\ dv_t^\nu = \kappa(\theta - v_t^\nu)dt + \xi\sqrt{v_t^\nu}d\widetilde{W}_t \end{cases}$$

où $v_0^\nu \sim \mathcal{L}(\nu) \sim \Gamma(\alpha,\beta)$ avec $\beta = 2\kappa/\xi^2$, $\alpha = \theta\beta$.

Évaluation d'options européennes Tout d'abord, dans la première partie du Chapitre 5, nous rappelons la méthode utilisée pour l'évaluation d'un *Call* dans le modèle d'Heston standard. À partir de la connaissance de la fonction caractéristique $\psi\big(\lambda(v),u,T\big)$ du logarithme de l'actif à la date T (voir [SST04; Gat11; Alb+07] pour un choix robuste de formule), le prix du *Call* de *strike* K et de maturité T sur l'actif $S_T^{(v)}$ dans le modèle d'Heston standard où la volatilité a pour condition initiale $v \in \mathbb{R}$ est donné par

$$C(\phi(v),K,T) = \mathbb{E}\big[\,\mathrm{e}^{-rT}(S_T^{(v)} - K)_+\big] = s_0\,\mathrm{e}^{-qT}\,P_1\big(\lambda(v),K,T\big) - K\,\mathrm{e}^{-rT}\,P_2\big(\lambda(v),K,T\big)$$

où les quantités $P_1\big(\lambda(v),K,T\big)$ et $P_2\big(\lambda(v),K,T\big)$ sont définies par

$$\begin{aligned} P_1\big(\lambda(v),K,T\big) &= \frac{1}{2} + \frac{1}{\pi}\int_0^{+\infty} \mathrm{Re}\left(\frac{\mathrm{e}^{-\mathbf{i}u\log(K)}}{iu}\frac{\psi\big(\lambda(v),u-\mathbf{i},T\big)}{s_0\,\mathrm{e}^{(r-q)T}}\right)du \\ P_2\big(\lambda(v),K,T\big) &= \frac{1}{2} + \frac{1}{\pi}\int_0^{+\infty} \mathrm{Re}\left(\frac{\mathrm{e}^{-\mathbf{i}u\log(K)}}{\mathbf{i}u}\psi\big(\lambda(v),u,T\big)\right)du \end{aligned}$$

avec $\mathbf{i}$ la base des nombres imaginaires (tel que $\mathbf{i}^2 = -1$).

À partir de cette formule, nous en déduisons une méthode pour calculer le prix I_0 d'un *Call* dans le modèle d'Heston stationnaire. En effet, en préconditionnant par v_0^ν, nous avons

$$I_0 = \mathbb{E}\big[\,\mathrm{e}^{-rT}\,\varphi(S_T^{(\nu)})\big] = \mathbb{E}\big[C(\phi(v_0^\nu),K,T)\big].$$

Ainsi, pour obtenir une approximation de I_0, nous proposons deux méthodes. La première, à base de quantification optimale, consiste à construire un quantifieur optimale de la loi gamma $\Gamma(\alpha,\beta)$ et ensuite d'utiliser la formule de cubature étudiée dans le Chapitre 4. La deuxième méthode consiste à utiliser une formule de quadrature à base de polynômes de Laguerre.

Calibration Une fois que nous sommes en mesure de calculer le prix d'options européennes dans le modèle d'Heston stationnaire, nous calibrons le modèle sur des données de marché pour étudier le comportement de sa volatilité implicite en temps court. Nous calibrons également le modèle d'Heston standard afin de comparer sa surface de volatilité implicite à celle du modèle stationnaire. Les deux modèles sont calibrés sur la surface de volatilité implicite de l'Euro Stoxx 50 (voir Figure 2.4). Étant donné que nous nous intéressons au comportement

court-terme de la surface de volatilité implicite, la calibration des modèles est réalisée sur les options de maturité 50 jours ($T = 50/365$). Nous observons ensuite les volatilités implicites générées par les modèles pour des maturités court-termes.

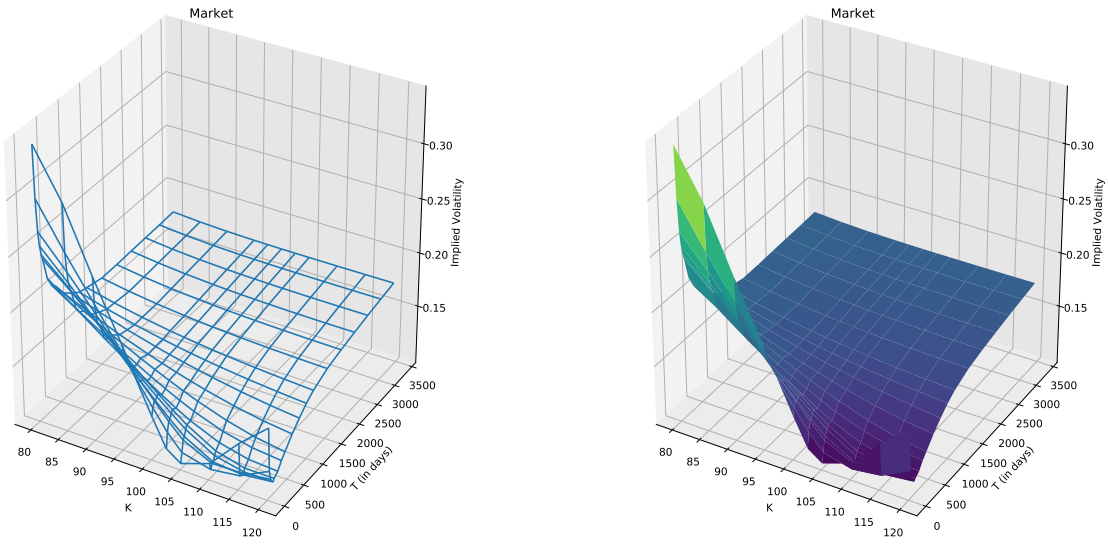


Fig. 2.4 *Surface de volatilité implicite de l'*Euro Stoxx *50 à la date du 26 Septembre 2019. ($S_0 = 3541$, $r = -0.0032$ et $q = 0.00225$)*

Le jeu de 4 paramètres du modèle d'Heston stationnaire devant être calibré est défini par

$$\mathcal{P}_{SH} = \big\{(\theta, \kappa, \xi, \rho) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+ \times [-1, 1]\big\}$$

et celui à 5 paramètres du modèle standard $\mathcal{P}_H$ par

$$\mathcal{P}_H = \big\{(x, \theta, \kappa, \xi, \rho) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+ \times [-1, 1]\big\}.$$

Les autres paramètres sont directement observés dans le marché.

Nous pouvons remarquer que le modèle stationnaire à un paramètre en moins à calibrer par rapport au modèle standard, ce qui rend sa calibration plus robuste que le modèle standard qui est connu pour être sur-paramètré (voir [GR09]). En pratique, nous observons que la calibration du modèle standard est très dépendante du jeu de paramètre utilisé pour initialiser l'algorithme d'optimisation alors que ce n'est pas le cas pour le modèle stationnaire.

Pour la calibration des modèles, la méthode standard consiste à chercher la solution du problème d'optimisation suivant

$$\min_{\phi \in \mathcal{P}} \sum_K \left(\frac{\sigma_{\text{IV}}^{Market}(K,T) - \sigma_{\text{IV}}^{Model}(\phi, K, T)}{\sigma_{\text{IV}}^{Market}(K,T)} \right)^2$$

où les quantités $\sigma_{\text{IV}}^{Market}(K,T)$ et $\sigma_{\text{IV}}^{Model}(\phi,K,T)$ sont, respectivement, les volatilités implicites du marché et celles calculées avec un modèle de Heston de paramètre $\phi = (\theta,\kappa,\xi,\rho)$ ou $\phi = (x,\theta,\kappa,\xi,\rho)$ selon les cas.

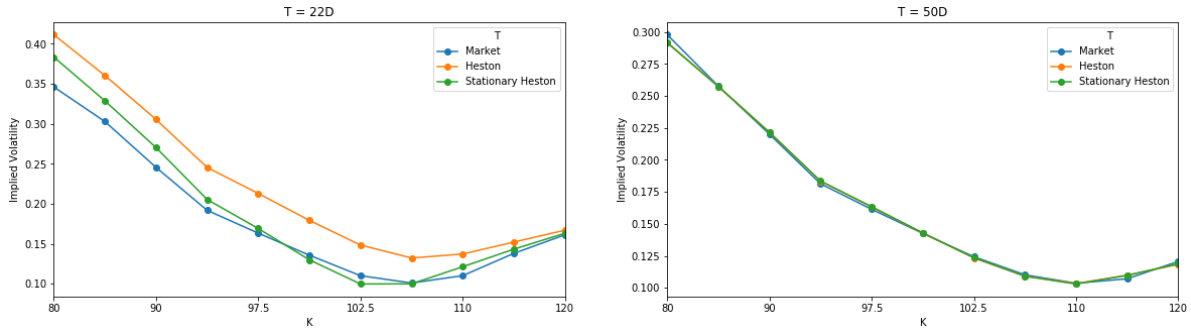


Fig. 2.5 *Volatilité implicite pour des options de maturité* 22 *(gauche) et* 50 *(droite) jours après calibration sans pénalisation.*

Dans la Figure 2.5, nous comparons les courbes de volatilité implicite générées par les deux modèles après calibration à des options européennes de maturité 50 jours. Nous observons en effet que le modèle stationnaire produit un *smile* de volatilité plus pentu que le modèle standard pour des options de maturité 22 jours. Cependant, lorsque l'on effectue la calibration, nous remarquons que les paramètres obtenus ne satisfont pas la condition de Feller

$$\xi^2 \leqslant 2\kappa\theta$$

qui assure la stricte positivité de la volatilité. Cette propriété est importante pour l'évaluation numérique d'options exotiques étudiée dans la dernière partie du chapitre.

Ainsi, pour obtenir des paramètres qui satisfont la condition de Feller, nous contraignons les paramètres en ajoutant une pénalisation dans le problème de minimisation qui devient

$$\min_{\phi\in\mathcal{P}} \sum_K \left(\frac{\sigma_{\text{IV}}^{Market}(K,T) - \sigma_{\text{IV}}^{Model}(\phi,K,T)}{\sigma_{\text{IV}}^{Market}(K,T)} \right)^2 + \lambda \max(\xi^2 - 2\kappa\theta, 0)$$

où λ est le facteur de pénalisation ajusté pendant la procédure.

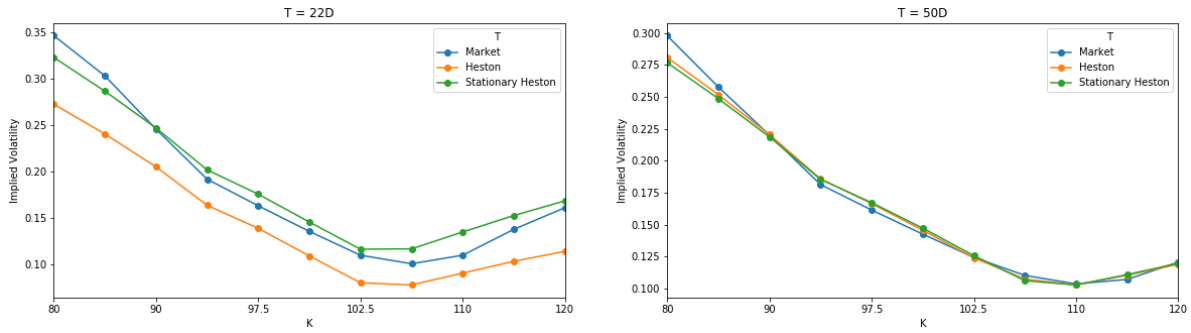


Fig. 2.6 *Volatilité implicite pour des options de maturité* 22 *(gauche) et* 50 *(droite) jours après calibration avec pénalisation.*

Dans la Figure 2.6, nous effectuons la même comparaison que précédement. Nous remarquons que l'ajout de la pénalisation a dégradé la qualité de la calibration à la maturité 50 jours. Pour ce qui est de la maturité 22 jours, nous observons que le modèle stationnaire arrive là encore, à produire un *smile* de volatilité plus proche de celui du marché que le modèle standard.

Évaluation d'options exotiques par quantification produit récursive Dans la dernière partie du Chapitre 5, nous traitons de l'évaluation des options exotiques telles que les options bermudiennes et les options à barrière à l'aide d'un *principe de programmation dynamique*. La méthode numérique que nous proposons est fondée sur de la quantification produit récursive. Nous étendons la méthodologie précédemment développée par [FSP18; CFG18; CFG17] où un schéma d'Euler-Maruyama était considéré pour discrétiser en temps à la fois l'actif et la volatilité.

Discrétisation en temps des diffusions Nous avons fait le choix de considérer un schéma hybride composé d'un schéma d'Euler-Maruyama pour la dynamique du log-actif $X_t = \log(S_t^{(\nu)})$ et d'un schéma de Milstein pour le processus de volatilité *boosté* $Y_t = \mathrm{e}^{\kappa t}\, v_t^{\nu}$. Ainsi, nous avons

$$\begin{cases} \bar{X}_{t_{k+1}} = \mathcal{E}_{b,\sigma}\big(t_k, \bar{X}_{t_k}, \bar{Y}_{t_k}, Z_{k+1}^1\big) \\ \bar{Y}_{t_{k+1}} = \mathcal{M}_{\tilde{b},\tilde{\sigma}}\big(t_k, \bar{Y}_{t_k}, Z_{k+1}^2\big) \end{cases}$$

avec $t_k = \frac{k}{n}$, n le nombre de pas de temps de discrétisation, $Z_{k+1}^1 \sim \mathcal{N}(0,1)$ et $Z_{k+1}^2 \sim \mathcal{N}(0,1)$ tel que $\mathbb{C}\mathrm{orr}(Z_{k+1}^1, Z_{k+1}^2) = \rho$. Le schéma d'Euler-Maruyama est défini par

$$\mathcal{E}_{b,\sigma}(t,x,y,z) = x + b(t,x,y)h + \sigma(t,x,y)\sqrt{h}\, z$$

avec

$$b(t,x,y) = r - q - \frac{\mathrm{e}^{-\kappa t}\, y}{2} \qquad \text{and} \qquad \sigma(t,x,y) = \mathrm{e}^{-\kappa t/2}\sqrt{y},$$

et le schéma de Milstein mis sous sa forme canonique

$$\mathcal{M}_{\widetilde{b},\widetilde{\sigma}}(t,x,z) = x - \frac{\widetilde{\sigma}(t,x)}{2\widetilde{\sigma}'_x(t,x)} + h\left(\widetilde{b}(t,x) - \frac{(\widetilde{\sigma}\widetilde{\sigma}'_x)(t,x)}{2}\right) + \frac{(\widetilde{\sigma}\widetilde{\sigma}'_x)(t,x)h}{2}\left(z + \frac{1}{\sqrt{h}\widetilde{\sigma}'_x(t,x)}\right)^2$$

avec

$$\widetilde{b}(t,x) = \mathrm{e}^{\kappa t}\,\kappa\theta, \qquad \widetilde{\sigma}(t,x) = \xi\sqrt{x}\,\mathrm{e}^{\kappa t/2} \qquad \text{and} \qquad \widetilde{\sigma}'_x(t,x) = \frac{\xi\,\mathrm{e}^{\kappa t/2}}{2\sqrt{x}}.$$

Quantification Markovienne produit recursive Une fois le choix du schéma de discrétisation en temps fait, nous nous intéressons à la discrétisation en espace du couple actif-volatilité.

Pour cela, nous construisons tout d'abord un arbre de quantification Markovien $(\widehat{Y}_{t_k})_{k=0,\dots,n}$. Il est avantageux de remarquer que la volatilité est autonome et donc nous faisons face à un problème unidimensionnel. Ainsi, les quantifieurs $\widehat{Y}_{t_k}$ sont construits récursivement, c'est à dire que $\widehat{Y}_{t_{k+1}}$ est un quantifieur optimal de $\widetilde{Y}_{t_{k+1}}$ défini par

$$\widetilde{Y}_{t_{k+1}} = \mathcal{M}_{\widetilde{b},\widetilde{\sigma}}\big(t_k, \widehat{Y}_{t_k}, Z^2_{k+1}\big), \qquad \widehat{Y}_{t_{k+1}} = \mathrm{Proj}_{\Gamma^Y_{N_{2,k+1}}}\big(\widetilde{Y}_{t_{k+1}}\big).$$

Numériquement, nous utilisons les méthodes à base d'algorithmes déterministes pour la dimension 1 développées dans le Chapitre 3.

Maintenant, en utilisant le fait que Y_t a déjà été quantifié, nous construisons un arbre de quantification Markovien $(\widehat{X}_{t_k})_{k=0,\dots,n}$ de X_t. Là encore nous sommes ramenés à un problème unidimensionnel et nous construisons les quantifieurs $\widehat{X}_{t_k}$ récursivement, c'est à dire que $\widehat{X}_{t_{k+1}}$ est un quantifieur optimal de $\widetilde{X}_{t_{k+1}}$ défini par

$$\widetilde{X}_{t_{k+1}} = \mathcal{E}_{b,\sigma}\big(t_k, \widehat{X}_{t_k}, \widehat{Y}_{t_k}, Z^1_{k+1}\big), \qquad \widehat{X}_{t_{k+1}} = \mathrm{Proj}_{\Gamma^X_{N_{1,k+1}}}\big(\widetilde{X}_{t_{k+1}}\big).$$

Afin de simplifier les notations, nous notons dans la suite $\widehat{X}_k$ et $\widehat{Y}_k$ à la place de $\widehat{X}_{t_k}$ et $\widehat{Y}_{t_k}$.

Maintenant que nous avons calibré le modèle d'Heston stationnaire et que nous sommes capable de construire un arbre de quantification pour le couple actif-volatilité, nous nous intéressons à l'évaluation d'options exotiques et plus précisément des options bermudiennes ou barrières.

Options bermudiennes Le prix à la date t_k d'une option bermudienne pouvant s'exercer aux dates $\{t_k, \cdots, t_n\}$ et de payoff $\psi_{t_k}(X_{t_k}, Y_{t_k})$ à la date t_k est donné par l'enveloppe de Snell V_k

$$V_k = \sup_{\tau\in\mathcal{T}^n_k} \mathbb{E}\Big[\,\mathrm{e}^{-r\tau}\,\psi_\tau(X_\tau, Y_\tau) \mid \mathcal{F}_{t_k}\,\Big],$$

où $\mathcal{T}_k^n$ représente l'ensemble des temps d'arrêt τ à valeurs dans $\{t_k, t_1, \dots, t_n\}$. Le *Principe de Programmation Dynamique* permet de réécrire V_k comme suit

$$\begin{cases} V_n = \mathrm{e}^{-rt_n}\, \psi_n(X_n, Y_n), \\ V_k = \max\left(\mathrm{e}^{-rt_k}\, \psi_k(X_k, Y_k), \mathbb{E}[V_{k+1} \mid \mathcal{F}_k]\right), \qquad 0 \leqslant k \leqslant n-1. \end{cases}$$

Nous appliquons ensuite la méthodologie employée par [BP03; BPP05; Pag18] qui consiste à remplacer X_k et Y_k par les quantifieurs $\widehat{X}_k$ et $\widehat{Y}_k$. Par construction de la quantification récursive, le couple $(\widehat{X}_k, \widehat{Y}_k)$ est Markovien ainsi nous obtenons le *Principe de Programmation Dynamique Quantifié* suivant

$$\begin{cases} \widehat{V}_n = \psi_n(\widehat{X}_n, \widehat{Y}_n), \\ \widehat{V}_k = \max\left(\psi_k(\widehat{X}_k, \widehat{Y}_k), \mathbb{E}[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{Y}_k)]\right), \qquad k = 0, \dots, n-1. \end{cases}$$

Finalement, le prix de l'option bermudienne est donné par $\mathbb{E}\left[\widehat{V}_0\right]$.

Options barrières Le prix à la date t_k d'une option barrière *Up-and-Out* de maturité T, de payoff terminal f et de barrière L est donné par

$$P_{UO} = \mathrm{e}^{-rT}\, \mathbb{E}\left[f(X_T)\, \mathbb{1}_{\sup_{t\in[0,T]} X_t \leqslant L}\right].$$

Pour l'évaluation de l'option barrière, nous appliquons l'algorithme basé sur la loi conditionnelle du schéma d'Euler, voir [Gla13; Sag10; Pag18]. Ainsi, une fois le couple actif-volatilité discrétisé en temps, le prix P_{UO} se réécrit de la façon suivante

$$\bar{P}_{UO} = \mathrm{e}^{-rT}\, \mathbb{E}\left[f(\bar{X}_T)\, \mathbb{1}_{\sup_{t\in[0,T]} \bar{X}_t \leqslant L}\right] = \mathrm{e}^{-rT}\, \mathbb{E}\left[f(\bar{X}_T) \prod_{k=0}^{n-1} G^k_{(\bar{X}_k, \bar{Y}_k), \bar{X}_{k+1}}(L)\right]$$

où

$$G^k_{(x,y),z}(u) = \left(1 - \mathrm{e}^{-2n\frac{(x-u)(z-u)}{T\sigma^2(t_k,x,y)}}\right) \mathbb{1}_{\{u \geqslant \max(x,z)\}}.$$

Finalement, en remplaçant $\bar{X}_k$ et $\bar{Y}_k$ par $\widehat{X}_k$ et $\widehat{Y}_k$ et en utilisant un algorithme récursif afin d'approcher $\bar{P}_{UO}$ par $\mathbb{E}[\widehat{V}_0]$, nous obtenons

$$\begin{cases} \widehat{V}_n = \mathrm{e}^{-rT}\, f(\widehat{X}_n), \\ \widehat{V}_k = \mathbb{E}\left[G^k_{(\widehat{X}_k, \widehat{Y}_k), \widehat{X}_{k+1}}(L)\, \widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{Y}_k)\right], \qquad 0 \leqslant k \leqslant n-1. \end{cases}$$

2.3.2 Évaluation d'options bermudiennes dans un modèle 3 facteurs (PRDC)

Dans le Chapitre 6, nous nous intéressons au problème d'évaluation d'option bermudiennes sur taux de change où l'on considère des taux d'intérêts domestiques et étrangers stochastiques.

Dans ce cas, on fait référence à un modèle 3 facteurs. Le Chapitre 6 correspond à l'article "Quantization-based Bermudan option pricing in the FX world" soumis à *Journal of Computational Finance* et accessible sur arXiv ou HAL (voir [Fay+19]). Cet article est un travail commun avec Jean-Michel Fayolle, Vincent Lemaire et Gilles Pagès.

Le besoin d'évaluer de tels produits est né au Japon à la fin du XXe siècle. En effet la persistance des taux d'intérêt bas durant les dernières décennies du siècle a été l'une des principales raisons qui ont conduit à la création de produits financiers structurés sur taux de change. Ces produits répondaient au besoin des investisseurs souhaitant obtenir des coupons plus élevés que ceux fondés sur le yen. Au fur et à mesure, les produits financiers se sont complexifiés pour en arriver aux produits appelés : power reverse dual currency (PRDC), voir [Wys17].

Même si ces produits ont été émis vers la fin du XXe siècle, ils sont toujours présents dans les portefeuilles des banques et doivent être pris en compte lors de l'évaluation des calculs de risque de contrepartie tels que l'ajustement de valeur de crédit (Credit Valuation Adjustment - CVA), l'ajustement de valeur de la dette (Debt Valuation Adjustment - DVA), l'ajustement de valeur du financement (Funding Valuation Adjustment - FVA), l'ajustement de valeur du capital (Capital Valuation Adjustment - KVA), ..., en bref xVA (voir [BMP13; CBB14; Gre15] pour plus de détails sur le sujet).

Le modèle. On définit $P(t,T)$ comme étant la valeur à l'instant t d'une unité de la devise choisie livrée (c'est-à-dire payée) à l'instant T, également connue sous le nom de prix du zéro coupon ou facteur d'actualisation. Nous noterons le zéro coupon avec d en exposant lorsque nous parlerons de zéro coupon dans la devise domestique ($P^d(t,T)$) et avec f en exposant pour le zéro coupon dans la devise étrangère. Le modèle utilisé pour diffuser les zéro coupons domestique et étranger se place dans la famille des modèles de courbe de rendement Heath-Jarrow-Morton (HJM). Pour plus de détails et de théorie sur ses modèles, on peut se référer aux articles suivants [EFG96; EMV92; HJM92; BS73].

Ainsi la diffusion de la courbe des zéro coupons domestiques sous la probabilité risque-neutre domestique $\mathbb{P}$ est donnée par

$$\frac{dP^d(t,T)}{P^d(t,T)} = r_t^d dt + \sigma_d(T-t)dW_t^d,$$

où W^d est un $\mathbb{P}$-mouvement brownien, r_t^d est le taux instantané domestique au temps t et σ_d est la volatilité. Pour la courbe des zéro coupons étrangers, la dynamique est donnée, sous la probabilité risque-neutre étrangère $\widetilde{\mathbb{P}}$, par la diffusion

$$\frac{dP^f(t,T)}{P^f(t,T)} P^f(t,T) = r_t^f dt + \sigma_f(T-t)d\widetilde{W}_t^f,$$

où $\widetilde{W}^f$ est un $\widetilde{\mathbb{P}}$-mouvement brownien, r_t^f est le taux instantané étranger au moment t et σ_f est la volatilité. Les deux probabilités $\widetilde{\mathbb{P}}$ et $\mathbb{P}$ sont supposées être équivalentes, c'est-à-dire $\widetilde{\mathbb{P}} \sim \mathbb{P}$ et il existe ρ_{df} défini comme limite de la variation quadratique croisée $\langle W^d, \widetilde{W}^f \rangle_t = \rho_{df} t$.

Pour le taux de change (FX), nous désignons par S_t la valeur au temps $t > 0$ d'une unité de monnaie étrangère dans la monnaie domestique. La dynamique de $(S_t)_{t \geqslant 0}$ est de type Black-Scholes sous la forme

$$\frac{dS_t}{S_t} = (r_t^d - r_t^f)dt + \sigma_S dW_t^S,$$

où r_t^d est le taux instantané de la monnaie domestique au temps t, r_t^f est le taux instantané de la monnaie étrangère au temps t, σ_S est la volatilité et W^S est un mouvement brownien standard sous la probabilité risque-neutre.

La problématique. Notre objectif est d'évaluer le prix d'options bermudiennes sur le taux de change S_t pouvant être exercé à $n+1$ dates: $\{t_0, \ldots, t_n\}$. Ainsi, le prix à la date t_k de l'option bermudienne est donné par l'*enveloppe de Snell* V_k de l'obstacle $\left(\mathrm{e}^{-\int_0^{t_k} r_s^d ds} \, \psi_{t_k}(S_{t_k}) \right)_{k=0:n}$

$$V_k = \sup_{\tau \in \mathcal{T}_k^n} \mathbb{E}\left[\mathrm{e}^{-\int_0^\tau r_s^d ds} \, \psi_\tau(S_\tau) \mid \mathcal{F}_{t_k} \right]$$

où τ est un temps d'arrêt à valeurs dans $\{t_k, \ldots, t_n\}$ et $\mathcal{T}_k^n$ représente l'ensemble de ces temps d'arrêt.

Example 2.3.1. Le payoff que nous considérons dans le Chapitre 6 est un celui d'un coupon de PRDC (voir l'exemple dans la Figure 2.7) défini par

$$\psi_{t_k}(x) = \min\left(\max\left(\frac{C_f(t_k)}{S_0} x - C_d(t_k), \mathrm{Floor}(t_k) \right), \mathrm{Cap}(t_k) \right)$$

où $\mathrm{Floor}(t_k)$ et $\mathrm{Cap}(t_k)$ sont les valeurs plancher choisies lors de la création du produit, ainsi que $C_f(t_k)$ et $C_d(t_k)$ qui sont les valeurs des coupon des monnaies étrangères et domestiques auxquels nous souhaitons nous comparer.

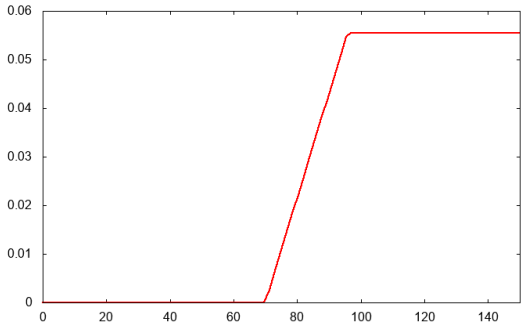


Fig. 2.7 *Exemple de payoff d'un PRDC* $\psi_{t_k}(S_{t_k}) = \min\left(\left(0.189\frac{S_{t_k}}{88.17} - 0.15\right)_+, 0.0555\right)$ *au temps* t_k.

Principe de programmation dynamique. Le *Principe de Programmation Dynamique* permet de réécrire V_k comme suit:

$$\begin{cases} V_n = \mathrm{e}^{-\int_0^{t_n} r_s^d ds}\, \psi_n(S_{t_n}), \\ V_k = \max\left(\mathrm{e}^{-\int_0^{t_k} r_s^d ds}\, \psi_k(S_{t_k}), \mathbb{E}[V_{k+1} \mid \mathcal{F}_{t_k}]\right), \qquad 0 \leqslant k \leqslant n-1. \end{cases}$$

De plus, on remarque que l'obstacle $\mathrm{e}^{-\int_0^t r_s^d ds}\, \psi_t(S_t)$ peut se réécrire comme une fonction h_t de deux processus X_t et Y_t

$$\mathrm{e}^{-\int_0^t r_s^d ds}\, \psi_t(S_t) = h_t(X_t, Y_t)$$

où le couple (X, Y) est défini par

$$(X_t, Y_t) = \left(\sigma_S W_t^S + \sigma_f \int_0^t (t-s) dW_s^f, -\sigma_d \int_0^t (t-s) dW_s^d\right).$$

Ainsi, cette nouvelle expression pour l'obstacle nous permet de réécrire la problème de l'enveloppe de Snell sous la forme

$$V_k = \sup_{\tau \in \mathcal{T}_k^n} \mathbb{E}\left[h_\tau(X_\tau, Y_\tau) \mid \mathcal{F}_{t_k}\right].$$

Cependant, le couple (X_k, Y_k) n'est pas Markovien et cela pose problème dans le *Principe de Programmation Dynamique* car le conditionnement qui apparait dans l'espérance conditionnelle ne pourra pas être remplacé par (X_k, Y_k). C'est pourquoi nous sommes amenés à considérer le vecteur aléatoire (X, W^f, Y, W^d) qui, lui, est Markovien.

Ainsi le *Principe de Programmation Dynamique* peut se réécrire de la façon suivante

$$\begin{cases} V_n = h_n(X_n, Y_n), \\ V_k = \max\left(h_k(X_k, Y_k), \mathbb{E}\left[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\right]\right), \qquad 0 \leqslant k \leqslant n-1. \end{cases} \tag{2.4}$$

Résolution numérique par quantification. Nous nous intéressons maintenant à la partie pratique qui consiste à calculer numériquement les valeurs de V_k. Dans le Chapitre 6, nous avons opté pour une méthode numérique à base de quantification optimale telle qu'introduite dans [BPP01] et développée dans [BP03; PPP04b; BPP05] pour l'évaluation d'options bermudiennes mais avec la variante consistant à utiliser un arbre de quantification optimale produit. Cette approche a pour avantage d'être rapide, stable et précise en petite dimension. Cependant lorsque la dimension croît, elle peut être très coûteuse en temps de calcul et la vitesse de convergence de la méthode se dégrade à cause de la "malédiction de la dimension" qui touche la quantification optimale.

La première idée que nous présentons, lorsque l'on souhaite discrétiser (2.4) par quantification optimale, est la plus naturelle. Nous remplaçons les variables aléatoires X_k, W_k^f, Y_k and W_k^d par leur quantification optimale $\widehat{X}_k$, $\widehat{W}_k^f$, $\widehat{Y}_k$ et $\widehat{W}_k^d$, de taille N_k^X, $N_k^{W^f}$, $N_k^{W^f}$, N_k^Y et $N_k^{W^d}$ respectivement, et nous "forçons", en un certains sens, la propriété de Markov en introduisant le *Principe de Programmation Dynamique Quantifié* "forcé" défini par

$$\begin{cases} \widehat{V}_n = h_n(\widehat{X}_n, \widehat{Y}_n), \\ \widehat{V}_k = \max\left(h_k(\widehat{X}_k, \widehat{Y}_k), \mathbb{E}\left[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)\right]\right), \qquad 0 \leqslant k \leqslant n-1. \end{cases}$$

Le terme "forcé" se justifie car $(\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)_k$ n'est pas une chaine de Markov donc ce principe de programmation dynamique n'est pas naturellement associé à une enveloppe de Snell. Nous désignons par $N_k = N_k^X \times N_k^{W^f} \times N_k^Y \times N_k^{W^d}$ la taille globale de la grille de quantification produit.

Pour cette approximation, nous fournissons une erreur quadratique à priori pour $\|V_k - \widehat{V}_k\|_2$, $k = 0, \ldots, n$.

Theorem 2.3.2. *Si les fonctions $(\psi_{t_k})_{k=0:n}$ sont dérivables à droite avec une dérivée bornée à support compact. Alors l'erreur quadratique induite par l'approximation par quantification $(\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)$ est bornée par*

$$\left\|V_k - \widehat{V}_k\right\|_2 \leqslant \left(\sum_{l=k}^{n} C_{X_l}\left\|X_l - \widehat{X}_l\right\|_{2p}^2 + C_{Y_l}\left\|Y_l - \widehat{Y}_l\right\|_{2p}^2 + C_{W_l^d}\left\|W_l^d - \widehat{W}_l^d\right\|_{2p}^2 + C_{W_l^f}\left\|W_l^f - \widehat{W}_l^f\right\|_{2p}^2\right)^{1/2},$$

où $1 < p < 3/2$ et $q \geqslant 1$ tel que $\frac{1}{p} + \frac{1}{q} = 1$ et les constantes $C_{X_l}, C_{W_l^d}, C_{Y_l}, C_{W_l^f}$ sont finies. Ainsi, en prenant $\bar{N} = \min_k N_k$, nous avons

$$\lim_{\bar{N} \to +\infty} \left\|V_k - \widehat{V}_k\right\|_2^2 = 0.$$

Le problème majeur de l'approche que nous venons de présenter est la complexité algorithmique associée à cette méthode due à la taille des grilles de quantification-produit. Cette

complexité rend très coûteux le calcul des espérances conditionnelles apparaissant dans le principe de programmation dynamique. Notre objectif est donc de réduire la dimension du problème. Pour cela, nous enlevons les processus W^d et W^f de l'arbre de quantification-produit pour ne garder que X et Y. Ce faisant, nous perdons la propriété de markovianité du vecteur aléatoire que nous considérons mais nous réduisons considérablement la complexité numérique du problème. Dans ce cadre, (2.4) est approchée par

$$\begin{cases} \widehat{V}_n = h_n(\widehat{X}_n, \widehat{Y}_n), \\ \widehat{V}_k = \max\left(h_k(\widehat{X}_k, \widehat{Y}_k), \mathbb{E}\left[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{Y}_k)\right]\right), \qquad 0 \leqslant k \leqslant n-1. \end{cases}$$

Nous notons $N_k = N_k^X \times N_k^Y$ la taille de la grille de quantification.

Là encore, nous fournissons une erreur quadratique à priori pour $\|V_k - \widehat{V}_k\|_2$, $k = 0, \dots, n$ basé sur les erreurs de quantification moyennes $\|X_l - \widehat{X}_l\|_{2p}$ et $\|Y_l - \widehat{Y}_l\|_{2p}$ mais également sur les erreurs que nous faisons en ne prenant pas en compte les mouvements browniens dans le conditionnement.

Theorem 2.3.3. *Si les fonctions $(\psi_{t_k})_{k=0:n}$ sont dérivables à droite avec une dérivée bornée à support compact alors l'erreur quadratique induite par l'approximation par quantification $(\widehat{X}_k, \widehat{Y}_k)$ est bornée par*

$$\left\|V_k - \widehat{V}_k\right\|_2 \leqslant \left(\sum_{l=k}^{n-1} C_{W_{l+1}^f} \left\|W_{l+1}^f - \mathbb{E}[W_{l+1}^f \mid (X_l, Y_l)]\right\|_{2p}^2 + C_{W_{l+1}^d} \left\|W_{l+1}^d - \mathbb{E}[W_{l+1}^d \mid (X_l, Y_l)]\right\|_{2p}^2 \right.$$
$$\left. + C_{X_l} \left\|X_l - \widehat{X}_l\right\|_{2p}^2 + C_{Y_l} \left\|Y_l - \widehat{Y}_l\right\|_{2p}^2 \right)^{1/2}$$

où $1 < p < 3/2$ et $q \geqslant 1$ tel que $\frac{1}{p} + \frac{1}{q} = 1$ et les constantes $C_{X_l}, C_{W_{l+1}^d}, C_{Y_l}, C_{W_{l+1}^f}$ sont finies. Ainsi, en prenant $\bar{N} = \min N_k$, nous avons

$$\lim_{\bar{N} \to +\infty} \left\|V_k - \widehat{V}_k\right\|_2^2 = \sum_{l=k}^{n-1} C_{W_{l+1}^f} \left\|W_{l+1}^f - \mathbb{E}[W_{l+1}^f \mid (X_l, Y_l)]\right\|_{2p}^2 + C_{W_{l+1}^d} \left\|W_{l+1}^d - \mathbb{E}[W_{l+1}^d \mid (X_l, Y_l)]\right\|_{2p}^2.$$

Nous pouvons ainsi remarquer que l'approximation que nous avons faite en remplaçant le préconditionnement en (X_k, W_k^f, Y_k, W_k^d) par (X_k, Y_k), même s'il réduit considérablement la complexité du problème induit une erreur systématique. Cependant, il semble raisonnable de penser que cette erreur est négligeable.

Example 2.3.4. En effet, dans la Figure 2.8, lors de l'évaluation d'options bermudiennes pouvant être exercées annuellement pour des maturités de 2, 5 ou 10 ans, en considérant des paramètres de marché pour σ_d et σ_f, la différence de prix entre les deux méthodes est négligeable. Le payoff considéré est celui de l'exemple 2.3.1. Pour l'exemple considéré dans la Figure 2.8, les corrélations sont supposées nulles $\rho_{Sd} = \rho_{Sf} = \rho_{df} = 0$, $S_0 = 88.17$, $\sigma_S = 50\%$,

$\sigma_d = \sigma_f = 50bp$ ($1bp = 0.01\%$), $P_d(0,t) = \exp(-r_d t)$ avec $r_d = 1.5\%$ et $P_f(0,t) = \exp(-r_f t)$ avec $r_f = 1\%$.

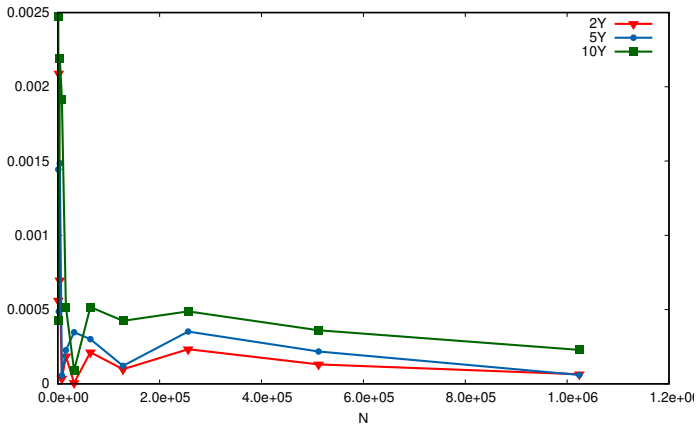


Fig. 2.8 *Différence relative des prix données par les deux méthodes pour des options bermudiennes exerçables annuellement et de maturité* 2, 5 *ou* 10 *ans.*

Chapter 3

Optimization of Optimal Quantizers

Let X be an $\mathbb{R}^d$-valued random vector with distribution $\mu = \mathbb{P}_X$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that $X \in L^2_{\mathbb{R}^d}(\Omega, \mathcal{A}, \mathbb{P})$. Let $|\cdot|$ be the euclidean norm in $\mathbb{R}^d$. In this chapter, we describe existing procedures to build the optimal quantizations of X, by which we mean: the best approximation of X by a discrete random vector $\widehat{X}^N$ with cardinality at most N.

3.1 Theoretical foundations

Definition 3.1.1. Let $\Gamma_N = \{x_1^N, \dots, x_N^N\} \subset \mathbb{R}^d$ be a subset of size N, called N-quantizer. A Borel partition $\big(C_i(\Gamma_N)\big)_{i=1,\dots,N}$ of $\mathbb{R}^d$ is a Voronoï partition of $\mathbb{R}^d$ induced by the N-quantizer Γ_N if, for every $i \in \{1, \dots, N\}$,

$$C_i(\Gamma_N) \subset \big\{\xi \in \mathbb{R}^d, |\xi - x_i^N| \leqslant \min_{j \neq i} |\xi - x_j^N|\big\}.$$

The Borel sets $C_i(\Gamma_N)$ are called Voronoï cells of the partition induced by Γ_N.

Remark. Any such N-quantizer is in correspondence with the N-tuple $x = (x_1^N, \dots, x_N^N) \in (\mathbb{R}^d)^N$ as well as with all N-tuples obtained by a permutation of the components of x. This is why we sometimes replace Γ_N by x.

Definition 3.1.2. A Voronoï quantization of X by Γ_N, $\widehat{X}^N$, is defined as a Borel nearest neighbor projection of X onto Γ_N associated to a Voronoï partition $\big(C_i(\Gamma_N)\big)_{i=1,\dots,N}$ for the euclidean norm

$$\widehat{X}^N := \mathrm{Proj}_{\Gamma_N}(X) = \sum_{i=1}^N x_i^N \mathbb{1}_{X \in C_i(\Gamma_N)}$$

and its associated probabilities, also called weights, are given by

$$\mathbb{P}\left(\widehat{X}^N = x_i^N\right) = \mathbb{P}_X\left(C_i(\Gamma_N)\right) = \mathbb{P}\left(X \in C_i(\Gamma_N)\right).$$

Figure 3.1 shows two Voronoï quantizations of a 2-dimensional centered Gaussian vector. The red dots represent the centroids, the cells are the Voronoï cells associated to each centroid and the color of each cell represent the weight of the cell.

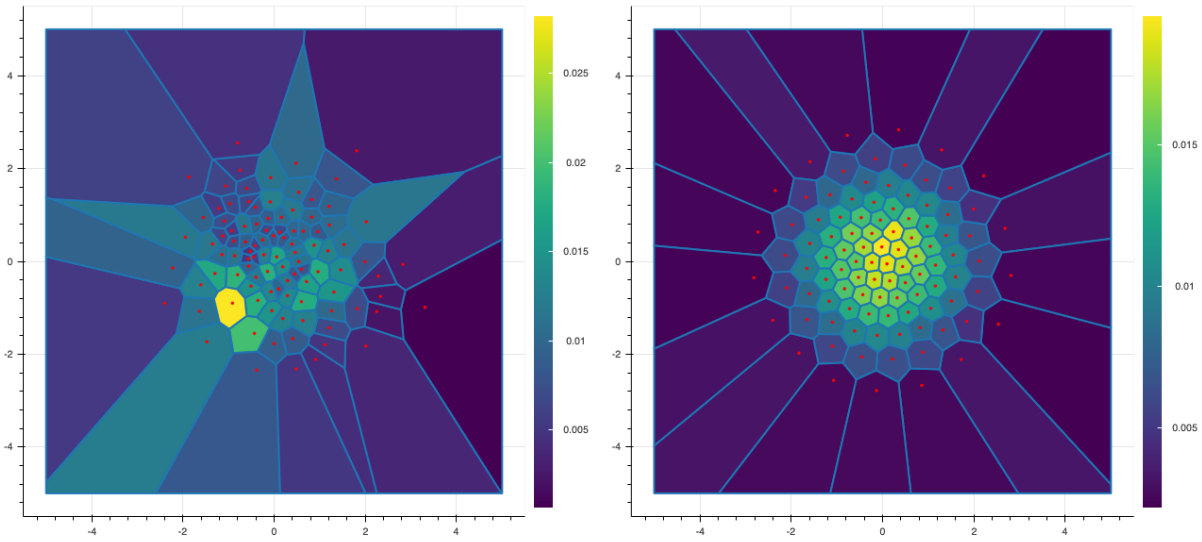


Fig. 3.1 *Two quantizations of size* $N = 100$ *of a* 2*-dimensional standard Gaussian vector.*

We are looking for the best approximation of X in the sense that we want to minimize the distance between X and $\widehat{X}^N$. This distance is measured by the standard L^2 norm, denoted as $\|X - \widehat{X}^N\|_2$, is called the mean quantization error. But we often use the quadratic distortion defined as half of the square of the mean quantization error.

Definition 3.1.3. The quadratic distortion function at level N induced by an N-tuple $x := (x_1^N, \dots, x_N^N)$ is given by

$$\mathcal{Q}_{2,\mathrm{N}} : x \longmapsto \frac{1}{2}\,\mathbb{E}\left[\min_{i=1,\dots,N} |X - x_i^N|^2\right] = \frac{1}{2}\,\mathbb{E}\left[\operatorname{dist}(X, \Gamma_N)^2\right] = \frac{1}{2}\|X - \widehat{X}^N\|_2^2.$$

Of course, the above result can be extended to the L^p case by considering the L^p-mean quantization error in place of the quadratic one.

Thus, we are looking for quantizers $\widehat{X}^N$ taking value in grids Γ_N of size N which minimize the quadratic distortion

$$\min_{\Gamma_N \subset \mathbb{R}^d, |\Gamma_N| \leqslant N} \|X - \widehat{X}^N\|_2^2.$$

We briefly recall some classical theoretical results on optimal quantizer, see [GL00; Pag18] for further details. The first ones deals with of the existence and the uniqueness of optimal quantizers.

Theorem 3.1.4. *(Existence of optimal* N*-quantizers) Let* $X \in L^2_{\mathbb{R}^d}(\mathbb{P})$ *and* $N \in \mathbb{N}^\star$.

(a) The quadratic distortion function $\mathcal{Q}_{2,\mathrm{N}}$ at level N attains a minimum at a N-tuple $x^\star = (x_1^N, \ldots, x_N^N)$ and $\Gamma_N^\star = \{x_i^N, i = 1, \ldots, N\}$ is a quadratic optimal quantizer at level N.

(b) If the support of the distribution $\mathbb{P}_X$ of X has at least N elements, then $x^\star = (x_1^N, \ldots, x_N^N)$ has pairwise distinct components, $\mathbb{P}_X\left(C_i(\Gamma_N^\star)\right) > 0, i = 1, \ldots, N$. Furthermore, the sequence $N \mapsto \inf_{x \in (\mathbb{R}^d)^N} \mathcal{Q}_{2,\mathrm{N}}(x)$ converges to 0 and is decreasing as long as it is positive.

The next results deal with the asymptotic behavior of the distortion. We saw in Theorem 3.1.4 that the infimum of the quadratic distortion converges to 0 as N goes to infinity. The next theorem, known as Zador's Theorem, analyzes the sharp rate of convergence of the quantization error. This result has been proved in the case of the L^p-optimal quantization.

Theorem 3.1.5. *(Zador's Theorem) Let $p \in (0, +\infty)$.*

(a) Sharp rate [Zad82; GL00]. *Let $X \in L^{p+\delta}_{\mathbb{R}^d}(\mathbb{P})$ for some $\delta > 0$. Let $\mathbb{P}_X(d\xi) = \varphi(\xi) \cdot \lambda(d\xi) + \nu(d\xi)$, where $\nu \perp \lambda$ i.e. ν is singular with respect to the Lebesgue measure λ on $\mathbb{R}^d$. Then, there is a constant $\widetilde{J}_{p,d} \in (0, +\infty)$ such that*

$$\lim_{N \to +\infty} N^{1/d} \min_{\Gamma_N \subset \mathbb{R}^d, |\Gamma_N| \leqslant N} \|X - \widehat{X}^N\|_p = \widetilde{J}_{p,d} \left[\int_{\mathbb{R}^d} \varphi^{\frac{d}{d+p}} d\lambda_d \right]^{\frac{1}{p} + \frac{1}{d}}$$

where $\widehat{X}^N$ is an L^p-optimal quantization of X.

(b) Non asymptotic upper-bound [GL00; Pag18]. *Let $\delta > 0$. There exists a real constant $C_{d,p,\delta} \in (0, +\infty)$ such that, for every $\mathbb{R}^d$-valued random vector X,*

$$\forall N \geqslant 1, \qquad \min_{\Gamma_N \subset \mathbb{R}^d, |\Gamma_N| \leqslant N} \|X - \widehat{X}^N\|_p \leqslant C_{d,p,\delta} \sigma_{\delta+p}(X) N^{-1/d}$$

where, for $r \in (0, +\infty), \sigma_r(X) = \min_{a \in \mathbb{R}^d} \|X - a\|_r < +\infty$ is the L^r-pseudo-standard deviation.

Another really interesting property concerning quadratic optimal quantizers is the stationarity property which is closely linked to the Lloyd method defined later in Section 3.2.1.1.

Proposition 3.1.6. *(Stationarity) Assume that the support of $\mathbb{P}_X$ has at least N elements. Any L^2-optimal N-quantizer $\Gamma_N \in (\mathbb{R}^d)^N$ is stationary in the following sense: for every Voronoï quantization $\widehat{X}^N$ of X,*

$$\mathbb{E}\left[X \mid \widehat{X}^N\right] = \widehat{X}^N.$$

Moreover $\mathbb{P}\left(X \in \bigcup_{i=1,\ldots,N} \partial C_i(\Gamma_N)\right) = 0$, so all optimal quantization induced by Γ_N a.s. coincide.

3.2 How to build an optimal quantizer?

In this part, we focus our efforts on the following minimization problem

$$\arg\min_{(\mathbb{R}^d)^N} \mathcal{Q}_{2,\mathrm{N}} \tag{3.1}$$

and more exactly, how to build an optimal quadratic quantizer? For that, we differentiate the L^2-distortion function $\mathcal{Q}_{2,\mathrm{N}}$ at level N. The approaches for solving the above minimization problem can be divided in two families: the fixed-point methods and the gradient descent methods. Both are linked to the distortion's gradient that we define below.

Proposition 3.2.1 ([Pag18]). *The distortion function $\mathcal{Q}_{2,\mathrm{N}}$ is continuously differentiable at N-tuples $x \in (\mathbb{R}^d)^N$ satisfying*

$$x \text{ has pairwise distinct components and } \mathbb{P}\left(X \in \bigcup_{i=1,\dots,N} \partial C_i(\Gamma_N)\right) = 0$$

with a gradient $\nabla\, \mathcal{Q}_{2,\mathrm{N}} = \left(\frac{\partial\, \mathcal{Q}_{2,\mathrm{N}}}{\partial x_i^N}\right)_{1\leqslant i\leqslant N}$ given by

$$\frac{\partial\, \mathcal{Q}_{2,\mathrm{N}}}{\partial x_i^N}(x) = \mathbb{E}\left[\frac{\partial\, q_{2,\mathrm{N}}}{\partial x_i^N}(x, X)\right] = \int_{\mathbb{R}^d} \frac{\partial\, q_{2,\mathrm{N}}}{\partial x_i^N}(x, \xi)\, \mathbb{P}_X(d\xi),$$

the local gradient being given by

$$\frac{\partial\, q_{2,\mathrm{N}}}{\partial x_i^N}(x, \xi) = 2(x_i^N - \xi)\, \mathbb{1}_{\mathrm{Proj}_{\Gamma_N}(\xi)=x_i^N}, \qquad 1 \leqslant i \leqslant N. \tag{3.2}$$

Equivalently, the gradient can also be written as

$$\nabla\, \mathcal{Q}_{2,\mathrm{N}}(x) = 2\left[\int_{C_i(\Gamma_N)} (x_i^N - \xi)\, \mathbb{P}_X(d\xi)\right]_{i=1,\dots,N} = 2\Big[\, \mathbb{E}\left[\, \mathbb{1}_{X\in C_i(\Gamma_N)}(x_i^N - X)\right]\Big]_{i=1,\dots,N}. \tag{3.3}$$

The latter expression is useful for numerical methods based on deterministic procedures while the former featuring a local gradient is handy when we work with stochastic algorithms.

3.2.1 Real valued random variables: $d = 1$

In the first part of this section, we focus on the scalar case, when X is a random variable taking values in $\mathbb{R}$. Hence the Voronoï cells are intervals in $\mathbb{R}$ and if we consider that the quantizers $(x_i^N)_i$ are ordered: $x_1^N < x_2^N < \cdots < x_{N-1}^N < x_N^N$, then the Voronoï cells are given by

$$C_i(\Gamma_N) = \big(x_{i-1/2}^N, x_{i+1/2}^N\big], \qquad i = 1, \dots, N-1, \qquad C_N(\Gamma_N) = \big(x_{N-1/2}^N, x_{N+1/2}^N\big)$$

where $\forall i = 2, \dots, N, \quad x^N_{i-1/2} := (x^N_{i-1} + x^N_i)/2$ and $x^N_{1/2} := \inf(\mathrm{supp}(\mathbb{P}_X))$ and $x^N_{N+1/2} := \sup(\mathrm{supp}(\mathbb{P}_X))$.

In Figure 3.2, we represent in red the optimal quantizer of a standard normal distribution and the vertices of the cell $C_i(\Gamma_N)$ are represented by black lines on the real axis. The probability associated to the quantizer x^N_i is the integral on the cell $C_i(\Gamma_N)$ of the normal density, as represented in the Figure.

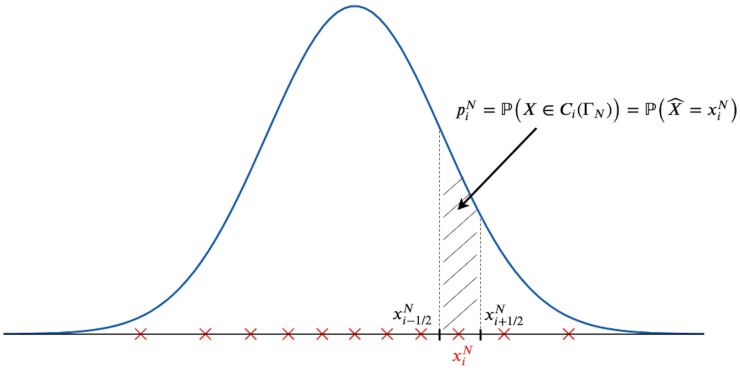


Fig. 3.2 *Optimal quantization of size* $N = 11$ *of a standard normal distribution* $\mathcal{N}(0, 1)$.

Moreover, in dimension 1, Kieffer (see [Kie83]) showed the uniqueness of the optimal quantizer if the density of X is log-concave with respect to the Lebesgue measure.

Theorem 3.2.2. *(Uniqueness of optimal N-quantizers see [Kie83]) If $\mathbb{P}_X(d\xi) = \varphi_X(\xi)d\xi$ with $\log \varphi_X$ concave, then for every $N \geqslant 1$, there is exactly one stationary N-quantizer (up to the permutations of the N-tuple). This unique stationary quantizer is a global (local) minimum of the distortion function, i.e.*

$$\forall N \geqslant 1, \qquad \arg\min_{\mathbb{R}^N} \mathcal{Q}_{2,\mathrm{N}} = \{x^\star\}.$$

In what follows, we will forget the star notation ($\star$) when speaking of optimal quantizers, $x^\star$ and $\Gamma^\star_N$ will be replaced by x and Γ_N.

Now, we focus on the algorithmic aspects of the optimization of optimal quantizer. If we know the density of X, then we can devise fast deterministic minimization procedures. For that purpose, we rewrite Equation (3.3) with the expression of the first partial moment and the cumulative distribution function of X

$$\nabla \mathcal{Q}_{2,\mathrm{N}}(x) = 2\Big[x_i\Big(F_X\left(x^N_{i+1/2}\right) - F_X\left(x^N_{i-1/2}\right)\Big) - \Big(K_X\left(x^N_{i+1/2}\right) - K_X\left(x^N_{i-1/2}\right)\Big)\Big]_{i=1,\dots,N} \tag{3.4}$$

where $K_X(\cdot)$ and $F_X(\cdot)$ are, respectively, the first partial moment and the cumulative distribution function of X

$$K_X(x) := \mathbb{E}[X\,\mathbb{1}_{X\leqslant x}] \qquad \text{and} \qquad F_X(x) := \mathbb{P}(X \leqslant x).$$

In the one dimensional case, we have access to a closed-form formula (or efficient numerical implementation) of the density function, the cumulative distribution function and partial first moment for a lot of random variables. We summarize below, for several random variables X, $K_X(\cdot)$, $F_X(\cdot)$ and $\varphi_X(\cdot)$, the first partial moment, the cumulative distribution function and the density of X, respectively.

- **Standard normal distribution**: $X \sim \mathcal{N}(0,1)$

$$\varphi_X(\xi) = \frac{\mathrm{e}^{-\xi^2/2}}{\sqrt{2\pi}}, \qquad F_X(\xi) = \mathcal{N}(\xi), \qquad K_X(\xi) = -\varphi_X(\xi).$$

- **Log-normal distribution**: $X = \exp(\mu + \sigma Z)$ with $\mu \in \mathbb{R}$ and $\sigma > 0$ where $Z \sim \mathcal{N}(0,1)$

$$\varphi_X(\xi) = \frac{1}{\xi\sigma}\varphi_Z\Big(\frac{\log(\xi)-\mu}{\sigma}\Big), \qquad F_X(\xi) = \mathcal{N}\left(\frac{\log(\xi)-\mu}{\sigma}\right),$$
$$K_X(\xi) = \mathrm{e}^{\mu+\sigma^2/2}\,\mathcal{N}\left(\frac{\log(\xi)-\mu-\sigma^2}{\sigma}\right)$$

with φ_Z the density of Z.

- **Exponential distribution**: $X \sim \mathcal{E}(\lambda)$ with $\lambda > 0$

$$\varphi_X(\xi) = \lambda\,\mathrm{e}^{-\lambda\xi}, \qquad F_X(\xi) = 1 - \mathrm{e}^{-\lambda\xi}, \qquad K_X(\xi) = -\,\mathrm{e}^{-\lambda\xi}\left(\xi + \frac{1}{\lambda}\right) + \frac{1}{\lambda}.$$

- **Gamma distribution**: $X \sim \Gamma(\alpha,\beta)$ with $\alpha,\beta > 0$

$$\varphi_X(\xi) = \frac{\beta^\alpha}{\Gamma(\alpha)}\xi^{\alpha-1}\,\mathrm{e}^{-\beta\xi}, \qquad F_X(\xi) = \frac{\gamma(\alpha,\beta\xi)}{\Gamma(\alpha)}, \qquad K_X(\xi) = F_X(\xi) - \frac{\xi}{\beta}\varphi_X(\xi),$$

where $\Gamma(\cdot)$ is the gamma function and $\gamma(s,x) = \int_0^x t^{s-1}e^{-t}dt$ is the lower incomplete gamma function. Optimized numerical implementations for both functions can easily be find in any programming language.

- **Non-central $\chi^2(1)$ distribution**: $X \sim \chi^2(1) = (Z+m)^2$ with $m \in \mathbb{R}$ where $Z \sim \mathcal{N}(0,1)$

$$\varphi_X(\xi) = \frac{\varphi_Z\big(m+\sqrt{\xi}\big) + \varphi_Z\big(m-\sqrt{\xi}\big)}{2\sqrt{\xi}}, \qquad F_X(\xi) = \mathcal{N}\left(m+\sqrt{\xi}\right) - \mathcal{N}\left(m-\sqrt{\xi}\right),$$
$$K_X(\xi) = \big(m-\sqrt{\xi}\big)\,\mathcal{N}\left(m+\sqrt{\xi}\right) - \big(m+\sqrt{\xi}\big)\,\mathcal{N}\left(m-\sqrt{\xi}\right) + \big(1+m^2\big)F_X(\xi).$$

• **Supremum of the Brownian bridge**: $X = \sup_{t\in[0,1]} |W_t - tW_1|$. This distribution is also known as the Kolmogorov-Smirnov distribution.

$$\varphi_X(\xi) = 8\xi \sum_{k\geqslant 1} (-1)^{k-1} k^2 \, \mathrm{e}^{-2k^2\xi^2}, \qquad F_X(\xi) = 1 - 2\sum_{k\geqslant 1} (-1)^{k-1} \, \mathrm{e}^{-2k^2\xi^2},$$
$$K_X(\xi) = \sqrt{2\pi} \sum_{k\geqslant 1} \frac{(-1)^{k-1}}{k} \Big(\mathcal{N}(2k\xi) - \frac{1}{2} \Big) - \xi(1 - F_X(\xi)),$$

where $\mathcal{N}(x)$ denotes the cumulative distribution function of the normal distribution. The proof of the formulas above are given in Appendix 3.A.

• **Symmetric random variable** For some random variables X, we have no access to closed-form formulas for φ_X, F_X and K_X but if X is symmetric and we have an explicit expression for its characteristic function $\chi(u) = \mathbb{E}\left[\, \mathrm{e}^{\mathbf{i}uX} \,\right]$, where $\mathbf{i}$ is the imaginary number, s.t. $\mathbf{i}^2 = -1$, then the functions φ_X, F_X and K_X can be written as alternate series using Fourier transform. This method was introduced in chapter 5 of [Pag18]. The proof of the formulas below are given in Appendix 3.A. For $\xi \geqslant 0$, we have

$$\varphi_X(\xi) = \frac{1}{\pi\xi} \sum_{k\geqslant 0} (-1)^k \int_0^\pi \cos(u) \chi\Big(\frac{u + k\pi}{\xi} \Big) du,$$
$$F_X(\xi) = \frac{1}{2} + \frac{1}{\pi} \sum_{k\geqslant 0} (-1)^k \int_0^\pi \frac{\sin(u)}{u + k\pi} \chi\Big(\frac{u + k\pi}{\xi} \Big) du,$$
$$K_X(\xi) = -C + \xi\Big(F_X(\xi) - \frac{1}{2} \Big) + \frac{\xi}{\pi} \sum_{k\geqslant 0} \int_0^\pi \frac{1 - (-1)^k \cos(u)}{(u + k\pi)^2} \chi\Big(\frac{u + k\pi}{\xi} \Big) du,$$

where $C = \mathbb{E}[X_+]$ and for $\xi < 0$

$$\varphi_X(\xi) = \varphi_X(-\xi), \qquad F_X(\xi) = 1 - F_X(-\xi), \qquad K_X(\xi) = K_X(-\xi).$$

Example 3.2.3. We give some examples of symmetric random variables where we can use the above formulas based on Fourier in order to obtain the functions φ_X, F_X and K_X.

∗ **One-sided Lévy's area**: $X \sim \int_0^1 W_s^1 dW_s^2$ where (W^1, W^2) is a 2-dimensional standard Brownian motion. The characteristic function of the Lévy's area is given by

$$\chi(u) = \frac{1}{\sqrt{\cosh(u)}} \quad \text{and} \quad C = 0.24852267 \pm 2.033 \times 10^{-7},$$

where C has been computed using a ML2R estimator.

∗ **Standard normal distribution**: $X \sim \mathcal{N}(0, 1)$. Although we have explicit formulas for the desired functions, we can still use the above formulas based on alternating series to the

Gaussian case in order to validate the methodology. For the normal distribution, we have

$$\chi(u) = \mathrm{e}^{-u^2/2} \quad \text{and} \quad C = \frac{1}{\sqrt{2\pi}}.$$

• **Closed-form formula of the characteristic function** Another method, introduced in [CFG19] for the quantization of a positive diffusion $(S_t)_{t\in[0,T]}$ at time T, is based on Fourier inversion in order to determine a computable expression of the density and the cumulative distribution function. They use the fact that the conditional characteristic function of $X = \log(S_T)$ is explicitly known or can be computed efficiently and denoted

$$\chi(u) = \mathbb{E}\left[\mathrm{e}^{\mathbf{i}u\log(S_T)}\right], \qquad u \in \mathbb{R}.$$

Using the knowledge of the characteristic function of X, they obtain

$$\mathbb{P}(S_T \in dz) = \left(\frac{1}{\pi}\frac{1}{z}\int_0^{+\infty} \mathrm{Re}\big(\mathrm{e}^{-\mathbf{i}\log(z)\xi}\,\chi(u)\big)du\right)dz$$
$$\mathbb{P}(S_T \leqslant z) = \frac{1}{2} - \frac{1}{\pi}\int_0^{+\infty} \mathrm{Re}\left(\frac{\mathrm{e}^{-\mathbf{i}u\log(z)}\,\chi(u)}{\mathbf{i}u}\right)du, \qquad z \in (0,+\infty).$$

Hence, based on these formulas, they devise a Newton-Raphson algorithm (as detailed in Algorithm 4) for the optimization of an optimal quantizer of S_T.

Remark. Let $\Gamma_N = \{x_1^N, \ldots, x_N^N\}$ be an N-quantizer of X. In the one-dimensional case, when we know the cumulative distribution function of X as detailed above, we can deduce directly the probabilities $p_i^N = \mathbb{P}\left(\widehat{X}^N = x_i^N\right)$, indeed

$$p_i^N = \mathbb{P}\left(\widehat{X}^N = x_i^N\right) = \mathbb{P}\left(X \in C_i(\Gamma_N)\right) = F_X\left(x_{i+1/2}^N\right) - F_X\left(x_{i-1/2}^N\right).$$

3.2.1.1 Fixed-point search (Lloyd method)

Starting from Equation (3.4), when we search a zero of the gradient, we derive a fixed-point problem. Let $\Lambda_i : \mathbb{R}^N \mapsto \mathbb{R}$ defined by

$$\Lambda_i(x) = \frac{K_X\left(x_{i+1/2}^N\right) - K_X\left(x_{i-1/2}^N\right)}{F_X\left(x_{i+1/2}^N\right) - F_X\left(x_{i-1/2}^N\right)} \tag{3.5}$$

then

$$\nabla\, \mathcal{Q}_{2,\mathrm{N}}(x) = 0 \quad \Longleftrightarrow \quad \forall i = 1, \ldots, N \qquad x_i = \Lambda_i(x).$$

Hence, from this equality, we deduce a fixed-point search algorithm. This method, known as the Lloyd method, was first devised by Lloyd in [Llo82]. The convergence at an exponential rate of the algorithm was shown by Kieffer in [Kie82].

Algorithm 1: *Lloyd method.*

Data: x^0 an initial guess for the N-quantizer.
Initialization: $x \leftarrow x^0$;
`/* Stopping criteria defined in` (3.6) `*/`
while *not converged* **do**
  $x \leftarrow \Lambda(x)$;
end
Result: x

Let $\Lambda : \mathbb{R}^N \mapsto \mathbb{R}^N$ such that $\Lambda = (\Lambda_i)_{1 \leqslant i \leqslant N}$, the Lloyd method with initial condition x^0 is defined as follows

$$x^{[n+1]} = \Lambda\big(x^{[n]}\big)$$

where the i-th component of the map $\Lambda : \mathbb{R}^N \to \mathbb{R}^N$ is equal to Λ_i and $x^{[n]}$ is the quantizer obtain after n iterations of the algorithm. The pseudo-algorithm of the Lloyd method written on the vector x starting from a given quantizer x^0 is outlined in Algorithm 1.

Remark. The stopping criteria of the algorithm is arbitrary. A first idea could be to compute the gradient at each step and stop the iterations when its norm or all its components are lower than a chosen ϵ. However, the computation of the gradient would increase the computation time of each iteration. Hence, in all the algorithms we present, we use the following stopping criteria. Let $\epsilon \in \mathbb{R}$ chosen before the optimization process, we stop if the following is verified

$$\frac{\left|x^{[n+1]} - x^{[n]}\right|}{\left|x^{[n+1]}\right|} \leqslant \epsilon. \tag{3.6}$$

Moreover as shown in the next proposition, an interesting feature of the Lloyd algorithm is that it decreases the distortion at each iteration.

Proposition 3.2.4. *see e.g. [Pag18] The Lloyd algorithm makes the quadratic distortion decrease, i.e.*

$$n \mapsto \|X - \widehat{X}^{N,[n]}\|_2, \qquad \textit{is non-increasing.}$$

Remark. When no closed-form exists for the first partial moment $K_X(\cdot)$ and the cumulative distribution function $F_X(\cdot)$, we cannot rely anymore on the deterministic version of the Lloyd method and we have to use a stochastic version of the Lloyd method detailed in Section 3.2.2.1.

Acceleration procedures When working with fixed-points, it is useful to work with acceleration procedures. Indeed methods to accelerate fixed-point search procedures have been extensively studied since the 1960's and a wide range of methods is readily available today. We refer to [RH15; BZ13] for a review on the literature. As for optimal quantization, we tested many methods and retained Anderson's acceleration as the most efficient one. It is introduced in [And65] and detailed in [WN11].

The Anderson acceleration consists in updating the quantizer $x^{[n+1]}$ not only by applying the map Λ to the current step of the quantizer: $\Lambda\big(x^{[n]}\big)$ but to select a linear combination of the $m_n = \min(n, m)$ previous steps $\Lambda\big(x^{[n-k]}\big)$ for $k = 1, \dots, m_n$ and $\Lambda\big(x^{[n]}\big)$ yielding

$$x^{[n+1]} = \sum_{k=0}^{m_n} \alpha_k \Lambda\big(x^{[n-k]}\big).$$

The α_k's are chosen in such a way that the residual $\Lambda\big(x^{[n]}\big) - x^{[n]}$ decreases as much as possible, then the α_k's are solution to the following minimization problem

$$\min_{\alpha_k, k=0,\dots,m_n} \left\| \sum_{k=0}^{m_n} \alpha_k \Big(\Lambda\big(x^{[n-k]}\big) - x^{[n-k]}\Big) \right\|_2 \qquad \text{s.t.} \qquad \sum_{k=0}^{m_n} \alpha_k = 1. \tag{3.7}$$

This minimization problem cannot be solved directly hence we use the equivalent form of the least-squares problem (3.7) recalled in [WN11]

$$\min_{\gamma=(\gamma_0,\dots,\gamma_{m_n-1})^T} \|f_n - \mathcal{F}_n\,\gamma\|_2 \tag{3.8}$$

where $f_n = \Lambda\big(x^{[n]}\big) - x^{[n]}$ and $\mathcal{F}_n = (\Delta f_{n-m_n}, \dots, \Delta f_{n-1})$ is a matrix of size $N \times m_n$ with $\Delta f_i = f_{i+1} - f_i$ and now $x^{[n+1]}$ is updated using this formula

$$x^{[n+1]} = \Lambda\big(x^{[n]}\big) + (\mathcal{X}_n + \mathcal{F}_n)\gamma^{[n]}.$$

where $\gamma^{[n]}$ is the solution of (3.8) and $\mathcal{X}_n = (\Delta x^{[n-m_n]}, \dots, \Delta x^{[n-1]})$ is a matrix of size $N \times m_n$ with $\Delta x^{[i]} = x^{[i+1]} - x^{[i]}$. Anderson acceleration's pseudo-algorithm applied to Lloyd method for building optimal quantizers is detailed in Algorithm 2.

Remark. Even-though the Anderson acceleration reduce drastically the computation time for building optimal quantizers, it may suffer, in some cases, instability and produce centroids that are not in the support of the distribution we wish to quantize and in that case is not able to produce a quantizer. This is the case of log-normal of chi-squared distributions for example.

3.2.1.2 Gradient descent

Another approach for building an optimal quantizer consists in minimizing directly the problem (3.1) using a gradient descent. Several gradient descent algorithms applied in the search of an optimal quantizer exist and we detail them below.

Mean-field CLVQ The first idea is to use a first-order gradient descent. This is the deterministic or batch version of the Competitive Learning Vector Quantization (CLVQ) algorithm, which is a stochastic gradient descent introduced for the cases where we cannot numerically compute the gradient. In the literature on stochastic approximation, it is common

Algorithm 2: *Anderson acceleration applied to Lloyd method.*

Data: x^0 an initial guess for the N-quantizer, m the deepth of the memory.
Initialization: $x \leftarrow x^0$;
/* First, we apply one step of Standard Lloyd method */
$g \leftarrow \Lambda(x)$;
$f \leftarrow g - x$;
$\widetilde{x} \leftarrow x$; /* We keep in memory the previous iteration */
$x \leftarrow g$; /* Standard Lloyd */
$\widetilde{f} \leftarrow f$; /* We keep in memory the previous residual */
/* Then, we apply Anderson acceleration */
$n \leftarrow 1$;
/* Stopping criteria defined in (3.6) */
while *not converged* **do**
  $m_n \leftarrow \min(n, m)$;
  $g \leftarrow \Lambda(x)$;
  $f \leftarrow g - x$;
  $\Delta x \leftarrow x - \widetilde{x}$;
  $\Delta f \leftarrow f - \widetilde{f}$;
  Add a column to $\mathcal{F}$ with value Δf ; /* Size of $\mathcal{F}$: $N \times (\min(n-1, m) + 1)$ */
  Add a column to $\mathcal{X}$ with value Δx ; /* Size of $\mathcal{X}$: $N \times (\min(n-1, m) + 1)$ */
  if $n > m$ **then**
    Delete first column of $\mathcal{F}$; /* Size of $\mathcal{F}$: $N \times m_n$ */
    Delete first column of $\mathcal{X}$; /* Size of $\mathcal{X}$: $N \times m_n$ */
  end
  Find γ solution of $\min_\gamma \|f - \mathcal{F}\gamma\|_2$; /* QR decomposition */
  $\widetilde{x} \leftarrow x$; /* We keep in memory the previous iteration */
  $x \leftarrow g - (\mathcal{X} + \mathcal{F})\gamma$; /* We update x using the acceleration */
  $\widetilde{f} \leftarrow f$; /* We keep in memory the previous residual */
  $n \rightarrow n + 1$;
end
Result: x

Algorithm 3: *Mean-field CLVQ.*

Data: x^0 an initial guess for the N-quantizer.
Initialization: $x \leftarrow x^0$;
`/* Stopping criteria defined in` (3.6) `*/`
while *not converged* **do**
  $\gamma \leftarrow update\_step(\gamma)$;
  $x \leftarrow x - \gamma \nabla \mathcal{Q}_{2,\mathrm{N}}(x)$;
end
Result: x

to name the vector field $h(\theta) = H(\theta, Z)$ the mean vector field of the algorithm and, by extension, the deterministic recursive algorithm the mean-field algorithm or the mean algorithm. It was far before the emergence of "mean field games". In the one dimensional case, the gradient is easily computable using the expression of F_X and K_X hence we devise a gradient descent directly on the distortion. Starting from a given initial condition x^0, the quantizer after $n+1$ iterations is given by

$$x^{[n+1]} = x^{[n]} - \gamma_{n+1} \nabla \mathcal{Q}_{2,\mathrm{N}}\left(x^{[n]}\right)$$

where $\gamma_{n+1} \in (0,1)$ is either taken constant ($\gamma_{n+1} = \gamma$) or updated at each step using a line search (see [Bon+06; Swa69]) or using the Barzilai–Borwein method (see [BB88]). We consider $\gamma_{n+1} \in (0,1)$ in order to preserve the non-decreasing order of the quantizer after each iteration.

The pseudo-algorithm of the mean-field CLVQ is detailed in Algorithm 3 and the *update_step* function is chosen before the optimization.

Newton Raphson method One can optimize the algorithm defined above using a second-order method where the step γ_{n+1} is chosen optimally at each step and is set as the inverse of the Hessian matrix of the distortion function. Again, starting from a initial condition x^0 at step 0, we have

$$x^{[n+1]} = x^{[n]} - \left(\nabla^2 \mathcal{Q}_{2,\mathrm{N}}\left(x^{[n]}\right)\right)^{-1} \left(\nabla \mathcal{Q}_{2,\mathrm{N}}\left(x^{[n]}\right)\right) \tag{3.9}$$

with $\nabla^2 \mathcal{Q}_{2,\mathrm{N}}(x)$ the Hessian matrix for $x = (x_1, \ldots, x_N)$

$$\nabla^2 \mathcal{Q}_{2,\mathrm{N}}(x) = \left[\frac{\partial^2 \mathcal{Q}_{2,\mathrm{N}}}{\partial x_i \partial x_j}(x)\right]_{1 \leqslant i,j \leqslant N}.$$

The Hessian matrix tridiagonal and since we have access to X's density and cumulative distribution functions, each component of the matrix can be computed efficiently with the

Algorithm 4: *Newton Raphson algorithm.*

Data: x^0 an initial guess for the N-quantizer.
Initialization: $x \leftarrow x^0$;
`/* Stopping criteria defined in (3.6) */`
while *not converged* **do**
  $H \leftarrow \nabla^2 \mathcal{Q}_{2,\mathrm{N}}(x)$;
  $G \leftarrow \nabla \mathcal{Q}_{2,\mathrm{N}}(x)$;
  Find u solution of $Hu = G$; `/* SVD decomposition */`
  $x \leftarrow x - u$;
end
Result: x

following expression

$$\frac{\partial^2 \mathcal{Q}_{2,\mathrm{N}}}{\partial x_i^2}(x) = 2\left[F_X\left(x^N_{i+1/2}\right) - F_X\left(x^N_{i-1/2}\right)\right] - \frac{x_{i+1}-x_i}{2}\varphi_X\left(x^N_{i+1/2}\right) - \frac{x_i - x_{i-1}}{2}\varphi_X\left(x^N_{i-1/2}\right),$$
$$\frac{\partial^2 \mathcal{Q}_{2,\mathrm{N}}}{\partial x_i \partial x_{i+1}}(x) = -\frac{x_{i+1}-x_i}{2}\varphi_X\left(x^N_{i+1/2}\right),$$
$$\frac{\partial^2 \mathcal{Q}_{2,\mathrm{N}}}{\partial x_i \partial x_{i-1}}(x) = -\frac{x_i - x_{i-1}}{2}\varphi_X\left(x^N_{i-1/2}\right),$$
$$\frac{\partial^2 \mathcal{Q}_{2,\mathrm{N}}}{\partial x_i \partial x_j}(x) = 0 \text{ otherwise.}$$

The pseudo-algorithm of the Newton-Raphson method is detailed in the Algorithm 4, where in place of computing the inverse of the Hessian matrix, we solve the following linear system: we search for $u \in \mathbb{R}^N$ solution of

$$Hu = G$$

where $H = \nabla^2 \mathcal{Q}_{2,\mathrm{N}}(x)$ and $G = \nabla \mathcal{Q}_{2,\mathrm{N}}(x)$. In practice for solving the linear system, we use an SVD decomposition (of course, other decompositions can be used such that LU of QR).

The problem with the Newton-Raphson algorithm is that it suffers from high instability when the Hessian is almost non-invertible. Indeed, when the Hessian is almost non-invertible, the inverted Hessian is numerically unstable, the algorithm does not preserve the non-decreasing order of the quantizer and the solution can diverge. The problem is said to be ill-conditioned. It often arises when we wish to build optimal quantizers for high values of N. Moreover, when N is large, even though Newton-Raphson algorithm needs few iterations to converge (when it converges), each iteration requires a lengthy amount of time which results in slow convergence.

From Newton Raphson algorithm to Levenberg-Marquart method In this part, we propose a solution to the instability of the standard Newton Raphson algorithm. To obtain a well-conditioned problem, we replace the Hessian matrix $\nabla^2 \mathcal{Q}_{2,\mathrm{N}}(x)$ for a given x in (3.9) by $\nabla^2 \mathcal{Q}_{2,\mathrm{N}}(x) + \lambda I$ where I is the identity matrix and λ is a chosen parameter. Hence, starting

Algorithm 5: *Newton Raphson algorithm with Levenberg-Marquart method.*

Data: x^0 an initial guess for the N-quantizer.
Initialization: $x \leftarrow x^0$, $cost \leftarrow \mathcal{Q}_{2,\mathrm{N}}(x^0)$, $\lambda \leftarrow 1$;
`/* Stopping criteria defined in (3.6)  */`
while *not converged* **do**
  $previous\ cost \leftarrow cost$;
  $H \leftarrow \nabla^2\, \mathcal{Q}_{2,\mathrm{N}}(x)$;
  $G \leftarrow \nabla\, \mathcal{Q}_{2,\mathrm{N}}(x)$;
  while $cost \geqslant previous\ cost$ **do**
    $\widetilde{H} \leftarrow H + \lambda I$; `/* Or replace` I `by` $\mathrm{diag}(H)I$ `*/`
    Find u solution of $\widetilde{H}u = G$; `/* SVD decomposition */`
    $\widetilde{x} \leftarrow x - u$;
    $cost \leftarrow \mathcal{Q}_{2,\mathrm{N}}(\widetilde{x})$;
    if $cost \geqslant previous\ cost$ **then**
      $\lambda \leftarrow 10 \times \lambda$;
    end
  end
  $x \leftarrow \widetilde{x}$;
  $\lambda \leftarrow \lambda/10$;
end
Result: x

from a initial condition x^0 at step 0, we have

$$x^{[n+1]} = x^{[n]} - \left(\nabla^2\, \mathcal{Q}_{2,\mathrm{N}}\left(x^{[n]}\right) + \lambda_n I\right)^{-1}\left(\nabla\, \mathcal{Q}_{2,\mathrm{N}}\left(x^{[n]}\right)\right).$$

The parameter λ is called the damping parameter. Several approaches have been proposed for the choice of its value. In our case we consider the following heuristic: we start with a λ_0 (in our case we choose $\lambda_0 = 1$), choose a multiplicative factor $\nu > 1$ and compute the distortion after one iteration of the algorithm with λ. If this choice of λ reduce the distortion then we keep the obtained quantizer x, divide λ by ν and restart the procedure. Otherwise we discard the obtained quantizer and retry with $\lambda \times \nu$. An other possibility is to replace the identity matrix I by the matrix containing only the diagonal entries of the Hessian H.

The Newton-Raphson pseudo-algorithm with its Levenberg-Marquart variant is detailed in Algorithm 5.

Quasi-Newton algorithms Another solution for solving both the instability problem and Newton-Raphson algorithm's cost is to use a Quasi-Newton algorithm. Quasi-Newton algorithm does not require to compute the Hessian matrix. The Hessian at iteration $n+1$ is approximated by a matrix A_{n+1} which is a function of the previous step of the approximation A_n, the quantizers and the gradients at the previous iterations. For the update of the matrix A_n several formulas exist such that BFGS, Broyden, DFP or SR1, among the most popular.

3.2.1.3 Numerical examples

In this section, we compare the algorithms based on fixed-point search or gradient descent for building optimal quantization of a chosen distribution.

Fixed-point iterations First, in Table 3.1, we compare the optimal quantization of the Gaussian distribution (with $\mu = 0$ and $\sigma = 1$) using the standard Lloyd method or the Lloyd method with Anderson acceleration denoted AA-Lloyd (where we consider $m = 10$). We clearly notice that the Anderson acceleration reduces drastically the computation time when we wish to quantize the Gaussian distribution.

N	Lloyd	AA-Lloyd
10	182 (1 ms)	11 (1 ms)
50	3430 (4 ms)	67 (1 ms)
100	12105 (27 ms)	170 (2 ms)
200	42381 (148 ms)	461 (10 ms)
500	216973 (1731 ms)	1442 (78 ms)

Table 3.1 *Optimal quantization of the Gaussian distribution (with $\mu = 0$ and $\sigma = 1$). We display the number of iterations needed, with in parenthesis the computation time, in order to satisfy the stopping criteria* (3.6) *with $\epsilon = 1e-9$.*

However, in some cases, the Anderson acceleration procedure fails to produce an optimal quantizer, as noticed in Remark 3.2.1.1. For example, if we consider a log-normal distribution, Lloyd method converges in 14548 (61 ms) and 50589 (390 ms) iterations for the optimization grids of size $N = 50$ and $N = 100$, respectively, whereas the acceleration procedure *explodes* and is not able to output a grid (see Table 3.3).

Gradient descent In Table 3.2, we compare gradient descent-based methods for building optimal quantization grids of the Gaussian distribution. In the table, MF-CLVQ stands for Mean-Field CLVQ, NR for Newton-Raphson and NR-LM for Newton-Raphson with Levenberg-Marquart method. We notice that NR and NR-LM need few iterations in order to converge but each iteration takes a significant amount of time compare to the naive gradient descent MF-CLVQ. It is important to notice that, if we compare the fixed-point methods (Table 3.1) and gradient descent methods (Table 3.2), it is preferable to use fixed-point methods in the Gaussian case.

However, if we want to build optimal quantizers of the log-normal distribution with $\mu = 0$ and $\sigma = 1$, the Newton-Raphson algorithm fails to build quantizers of size $N = 50$, $N = 100$ and $N = 200$ when we do not use the Levenberg-Marquart method but when use it, we build it in less than a second (see Table 3.3).

N	MF-CLVQ	NR	NR-LM
10	4859 (4 ms)	7 (1 ms)	7 (1 ms)
50	215295 (436 ms)	11 (15 ms)	10 (13 ms)
100	1057302 (3458 ms)	13 (127 ms)	12 (117 ms)
200	4798625 (30 s)	14 (1.3 s)	14 (1.3 s)
500	27468462 (417 s)	17 (30 s)	16 (28 s)

Table 3.2 *Optimal quantization of the Gaussian distribution (with $\mu = 0$ and $\sigma = 1$). We display the number of iterations needed, with in parenthesis the computation time, in order to satisfy the stopping criteria* (3.6) *with $\epsilon = 1e-9$.*

Remark. For the numerical test, the stopping criterion ϵ has been set equal to $\epsilon = 1\text{e}-9$. Of course, higher values could have been used. For example, we tried with $\epsilon = 1\text{e}-6$ but when doing so, we noticed that the Mean-Field CLVQ algorithm stopped prematurely. Indeed, the algorithm converge really slowly, hence each iteration as a really small impact on the grid and if the update is too small, the stopping criterion is triggered. Concerning the other gradient descend based methods (NR and NR-LM), this change of value for ϵ has no impact because, when they converged, they converged really fast. For the fixed-point based methods, it depends on the size of the grid. For small grids ($N = 10, 50, 100$), it has almost no impact in terms of computation times as we are talking of computation times of few milliseconds. For bigger grids, the computation time is reduced, when $N = 500$, with Lloyd it takes 41516 iterations (344 ms) with $\epsilon = 1\text{e}-6$ compared to 216973 iterations (1731 ms) when $\epsilon = 1\text{e}-9$ and with AA-Lloyd it takes 704 iterations (39 ms) with $\epsilon = 1\text{e}-6$ compared to 1442 iterations (78 ms) when $\epsilon = 1\text{e}-9$.

Fixed-point search vs Gradient descent In this paragraph, first, we consider the log-normal distribution with $\mu = 0$ and $\sigma = 1$. In Table 3.3, we compare all the methods for building optimal quantizers of the log-normal distribution. The Lloyd method and the Mean-Field CLVQ always succeed to build a quantizer, even-though it can take several minutes. The Lloyd method with Anderson acceleration and the Newton-Raphson algorithm fail to build a quantizer for some values of N. And we notice that the Levenberg-Marquart procedure applied to the Newton-Raphson algorithm solves the instability problem and makes it the most competitive method for building optimal quantizers of the log-normal distribution.

In Table 3.4, we display the numerical results for the optimal quantization of the exponential distribution with parameter $\lambda = 1$. For this distribution, the only method failing to build optimal quantizer is the Newton-Raphson algorithm and the fastest method is the Lloyd method with Anderson acceleration.

N	Lloyd	AA-Lloyd	MF-CLVQ	NR	NR-LM
10	845 (1 ms)	28 (1 ms)	1.9e7 (23 s)	11 (1 ms)	11 (1 ms)
50	1.5e4 (59 ms)	Not converged	3.3e7 (130 s)	Not converged	13 (14 ms)
100	5.1e4 (363 ms)	Not converged	4.4e7 (328 s)	Not converged	14 (118 ms)
200	1.9e5 (2649 ms)	2.3e4 (803 ms)	6.5e7 (898 s)	Not converged	15 (1278 ms)

Table 3.3 *Optimal quantization of the log-normal distribution (with $\mu = 0$ and $\sigma = 1$). We display the number of iterations needed, with in parenthesis the computation time, in order to satisfy the stopping criteria* (3.6) *with $\epsilon = 1e-9$.*

N	Lloyd	AA-Lloyd	MF-CLVQ	NR	NR-LM
10	728 (1 ms)	19 (1 ms)	2.0e6 (1.6 s)	8 (1 ms)	8 (1 ms)
50	1.4e4 (14 ms)	189 (1 ms)	9.3e6 (12 s)	Not converged	13 (16 ms)
100	4.8e4 (79 ms)	610 (8 ms)	1.7e7 (33 s)	Not converged	13 (108 ms)
200	1.7e5 (538 ms)	1513 (36 ms)	3.4e7 (113 s)	Not converged	15 (1260 ms)

Table 3.4 *Optimal quantization of the exponential distribution (with $\lambda = 1$). We display the number of iterations needed, with in parenthesis the computation time, in order to satisfy the stopping criteria* (3.6) *with $\epsilon = 1e-9$.*

3.2.2 Higher dimension: $d \geqslant 2$

In this section, we consider the general case of a random vector X taking values in $\mathbb{R}^d$. We recall the expression of the gradient of the quadratic distortion

$$\nabla \mathcal{Q}_{2,\mathrm{N}}(x) = 2\left[\int_{C_i(\Gamma_N)} (x_i^N - \xi)\, \mathbb{P}_X(d\xi)\right]_{i=1,\dots,N}.$$

Even if we have access to the density of X, it is no longer feasible to compute numerically the integrals inside (3.2.2) over the cells (the Voronoï cells are polyhedral convex sets of dimension d) except in dimension 2 where it is possible to build the Voronoï tesselation of a quantizer and use two-dimensional quadrature formulas for computing the integrals effectively. We detail the possible numerical procedures for building an optimal quantizer in dimension 2 in Section 3.2.2.2.

Hence, in the generic case, we cannot rely anymore on deterministic procedures because of the computation of the integral. Instead, we can use stochastic algorithm that we detail in Section 3.2.2.1.

3.2.2.1 Stochastic procedures

Two main stochastic algorithms exist for building an optimal quantizer in $\mathbb{R}^d$. The first is the stochastic version of the fixed-point search also called Lloyd method and the second is a stochastic gradient descent.

Randomized Lloyd method The first method is based on the same idea as the Lloyd Algorithm 1 and in absence of deterministic methods, the expectations and probabilities are computed using Monte Carlo simulation. First, we recall (3.5)

$$\Lambda_i(x) = \frac{\mathbb{E}\left[X\,\mathbb{1}_{X\in C_i(\Gamma_N)}\right]}{\mathbb{P}\left(X \in C_i(\Gamma_N)\right)}.$$

Let $\xi_1, \ldots, \xi_M$ be independent copies of X and $\Lambda_i^M : \mathbb{R}^N \mapsto \mathbb{R}$, the stochastic version of Λ_i defined by

$$\Lambda_i^M(x) = \frac{\sum_{m=1}^{M} \xi_m\, \mathbb{1}_{\left\{\mathrm{Proj}_{\Gamma_N}(\xi_m)=x_i^N\right\}}}{\sum_{m=1}^{M} \mathbb{1}_{\left\{\mathrm{Proj}_{\Gamma_N}(\xi_m)=x_i^N\right\}}} \qquad \text{with} \qquad \Gamma_N = \{x_1^N, \ldots, x_N^N\}.$$

Hence, let $\Lambda = (\Lambda_i)_{1\leqslant i\leqslant N}$, the $n+1$ iteration of the Randomized Lloyd method is given by

$$x^{[n+1]} = \Lambda^M\left(x^{[n]}\right). \tag{3.10}$$

During the optimization of the quantizer it is possible to compute the weight p_i^N and the local distortion q_i^N associated to a centroid defined by

$$p_i^N = \mathbb{P}\left(X \in C_i(\Gamma_N)\right) \quad \text{and} \quad q_i^N = \mathbb{E}\left[(X - x_i^N)^2\,\mathbb{1}_{X\in C_i(\Gamma_N)}\right]. \tag{3.11}$$

We detail the pseudo-algorithm of the Randomized Lloyd method and the computation of the weights and the local-distortion (3.11) in Algorithm 6.

Remark. In the pseudo-algorithm 6, we use new random numbers, independent copies of X, for each batch of size M. However, it is also possible to generate only once a set of size M of independent copies of X and then in the loop that iterates from 1 to M we use them for every batch, as suggested in subsection 6.3.5 of [Pag18]. This amounts to consider the M-sample of the distribution of X as the distribution to be quantized.

Competitive Learning Vector Quantization The second algorithm is a stochastic gradient descent called Competitive Learning Vector Quantization (CLVQ) algorithm. Since in higher dimensions, the gradient cannot be computed, the idea is to replace it in the gradient

Algorithm 6: *Randomized Lloyd method.*

Data: x^0 an initial guess for the N-quantizer, M: number of copies of X to generate.
Initialization: $x \leftarrow x^0$;
`/* Vector of size` N `of probabilities init with` $0 \in \mathbb{R}$ `*/`
Initialization: $p \leftarrow 0$;
`/* Vector of size` N `of local-distortions init with` $0 \in \mathbb{R}$ `*/`
Initialization: $q \leftarrow 0$;
`/* Vector of size` N `for the sum in numerator of` (3.10) `init with` $0 \in \mathbb{R}^d$ `*/`
Initialization: $sum\_nearest \leftarrow 0$;
`/* Stopping criteria defined in` (3.6) `*/`
while *not converged* **do**
  for $m = 1$ **to** M **by** 1 **do**
    `/* Generate an iid copy of` X `*/`
    $\xi \leftarrow randomGeneration()$;
    `/* Return the index and the distance to the closest centroid of` ξ `*/`
    $i, dist \leftarrow closest\_centroid(x, \xi, dist, i)$;
    $p(i) \leftarrow p(i) + 1$;
    $sum\_nearest(i) \leftarrow sum\_nearest(i) + \xi$;
    $q(i) \leftarrow q(i) + (dist)^2$;
  end
  for $i = 1$ **to** N **by** 1 **do**
    $x(i) \leftarrow sum\_nearest(i)/p(i)$;
    $p(i) \leftarrow p(i)/M$;
    $q(i) \leftarrow q(i)/M$;
  end
end
Result: x, p, q

Algorithm 7: *Competitive Learning Vector Quantization (CLVQ) algorithm.*

Data: x^0 an initial guess for the N-quantizer, M: number of copies of X to generate.
Initialization: $x \leftarrow x^0$;
`/* Vector of size` N `of probabilities init with` $0 \in \mathbb{R}$ `*/`
Initialization: $p \leftarrow 0$;
`/* Vector of size` N `of local-distortions init with` $0 \in \mathbb{R}$ `*/`
Initialization: $q \leftarrow 0$;
Initialization: $n \leftarrow 0$; `/* Counter */`
`/* Stopping criteria defined in (3.6) */`
while *not converged* **do**
  `/* Generate an iid copy of` X `*/`
  $\xi \leftarrow randomGeneration()$;
  $\gamma \leftarrow update\_step(\gamma)$;
  `/* Return the index and the distance to the closest centroid of` ξ `*/`
  $i, dist \leftarrow closest\_centroid(x, \xi, dist, index)$;
  $x(i) \leftarrow (1-\gamma)x(i) + \gamma\,\xi$;
  $p(i) \leftarrow p(i) + 1$;
  $q(i) \leftarrow q(i) + (dist)^2$;
  $n \leftarrow n + 1$;
end
for $i = 1$ **to** N **by** 1 **do**
  $p(i) \leftarrow p(i)/n$;
  $q(i) \leftarrow q(i)/n$;
end
Result: x, p, q

descent by the local-gradient defined in (3.2). Let $\xi_1, \ldots, \xi_n, \ldots$ a sequence of independent copies of X, the $n+1$ iterate of the CLVQ algorithm is given by

$$x^{[n+1]} = x^{[n]} - \gamma_{n+1} \nabla\, q_{2,\mathrm{N}}(x^{[n]}, \xi_{n+1})$$

where $\nabla\, q_{2,\mathrm{N}} = \left(\frac{\partial\, q_{2,\mathrm{N}}}{\partial x_i^N}\right)_{1 \leqslant i \leqslant N}$ and for the choice on the learning rate we refer to the section 6.3.5 in [Pag18]. Again, during optimization we can compute the weights p_i^N and the local distortions q_i^N associated to the centroids. We detail the pseudo-algorithm of the CLVQ algorithm and the computation of the weights and the local-distortion in Algorithm 7.

Remark. Two developments of the CLVQ algorithm can be considered.

The first one consists in using the averaging algorithm of Rupper and Polyak, yielding the averaged quantizer $\widetilde{x}^{[n+1]}$ defined by

$$\begin{cases} x^{[n+1]} = x^{[n]} - \gamma_{n+1} \nabla\, q_{2,\mathrm{N}}(x^{[n]}, \xi_{n+1}) \\ \widetilde{x}^{[n+1]} = \dfrac{1}{n+1} \displaystyle\sum_{i=1}^{n+1} x^{[i]}. \end{cases}$$

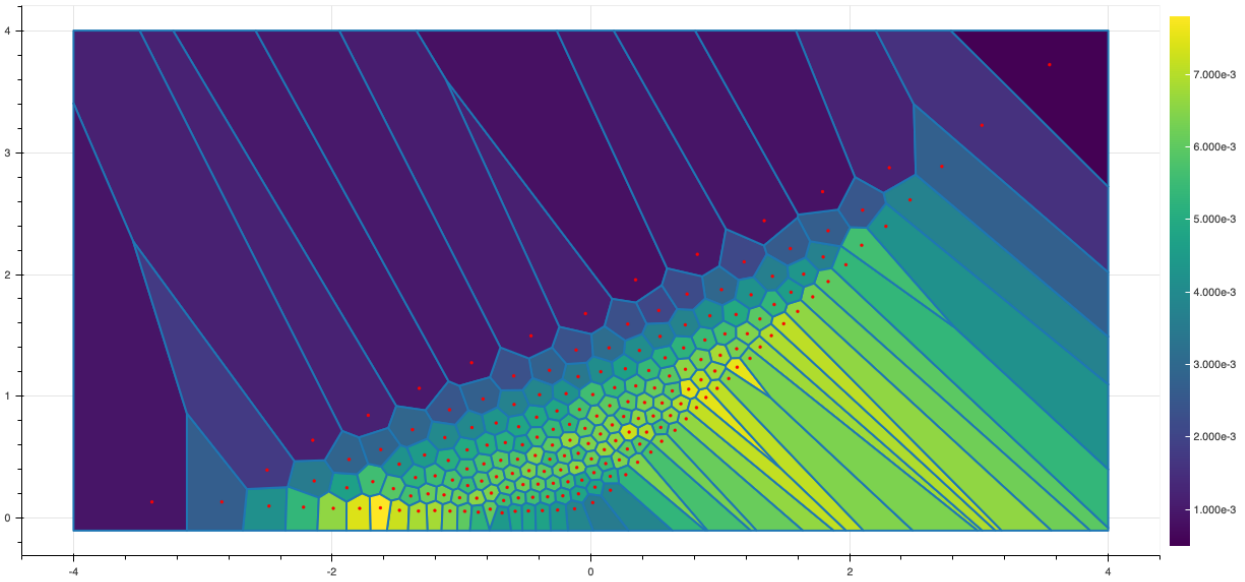


Fig. 3.3 *Optimal quantization of size* $N = 200$ *of* $(W_1, \sup_{t\in[0,1]W_t})$ *using the randomized Lloyd method* ($n = 50$ *and* $M = 5e6$). W *is the standard Brownian motion.*

The second, consists in considering a batch version of the stochastic algorithm in order to approximate the gradient at each step, yielding

$$x^{[n+1]} = x^{[n]} - \gamma_{n+1} \frac{1}{M} \sum_{m=1}^{M} \nabla\, q_{2,\mathrm{N}}(x^{[n]}, \xi_{n+1}^{m}).$$

This algorithm is the randomized version in dimension d of the mean-field CLVQ introduced in the one-dimensional setting.

Numerical example In Figure 3.3, we display the optimal quantization of size 200 of $(W_1, \sup_{t\in[0,1]W_t})$ where W is the standard Brownian motion. The optimal quantizer is obtained using the randomized Lloyd method with $n = 50$ iterations and $M = 5e6$, the size of the Monte-Carlo at each-step. The red dots represent the centroids, the cells are the Voronoï cells associated to each centroid and the color of each cell represent the weight of the cell.

3.2.2.2 Two dimensional setting: deterministic optimization

Let X a random vector taking values in $\mathbb{R}^2$ with distribution μ having a second moment. We consider in this section absolutely continuous distributions with density φ. The Voronoï cells can no longer be expressed as intervals but by as polyhedral convex sets. We then we go back

to the original form of the distortion gradient of (3.3)

$$\begin{aligned}\nabla \mathcal{Q}_{2,N}(x) &= 2\left[x_i^N \, \mathbb{P}\left(X \in C_i(\Gamma_N)\right) - \mathbb{E}\left[X \, \mathbb{1}_{X \in C_i(\Gamma_N)}\right]\right]_{i=1,\dots,N} \\ &= 2\left[x_i^N \int_{C_i(\Gamma_N)} \varphi(\xi) d\xi - \int_{C_i(\Gamma_N)} \xi \varphi(\xi) d\xi\right]_{i=1,\dots,N}.\end{aligned}$$

So if we are able to compute the two-dimensional integrals above, we can use deterministic optimization algorithms as in the one-dimensional case for optimizing the optimal quantizers.

In practice, these integrals cannot be computed exactly but they can be approximated numerically in a very effective way. For that, we build the Voronoï tessellation of the quantizer Γ_N and use quadrature formulas. Indeed, we detail below the steps for the numerical approximation of integrals of the type

$$\int_{C_i(\Gamma_N)} f(\xi) d\xi = \int\int_{C_i(\Gamma_N)} f(x, y) dx dy$$

with $\xi \in \mathbb{R}^2$ or $x, y \in \mathbb{R}$ where f can either be defined as a function with values in $\mathbb{R}$ (i.e. $f : \mathbb{R}^2 \to \mathbb{R}$) or in $\mathbb{R}^2$ (i.e. $f : \mathbb{R}^2 \to \mathbb{R}^2$).

Domain decomposition

1. First, if the support of X is not compact, we make a first approximation truncating the definition domain of X. For example, if X is a centered Gaussian vector with identity variance-covariance matrix, its support is $\mathbb{R}_2$, then we truncate the support and work on the squared domain (for the Gaussian distribution, otherwise on a rectangular domain) with coordinates

   $$\big((-M, -M), (-M, M), (M, M), (M, -M)\big)$$

   (in practice we consider $M = 15$). This choice seems reasonable because the Gaussian density vanishes for values far from zero, i.e. $\varphi(\xi) \approx 0$ for $|\xi| \geqslant M$. This truncation will impact the estimation of the integrals on the border of the Voronoï tesselation.

2. Then, we build the Voronoï tesselation of the grid Γ_N. Each $C_i(\Gamma_N)$ is a convex polygon with vertices $(z_0, \dots, z_{m-1})$ with $m \geqslant 3$, see Figure 3.4 as an example of Voronoï cell. For convenience we set $z_m = z_0$. The vertices of all Voronoï cells have finite coordinates because, after truncation, we work on a compact set. Open source C++ libraries able to build the Voronoï tesselation of a set of points are available online such as the QHull library [Bar+96] or the Boost Voronoi Diagram Library[1].

[1] https://www.boost.org/doc/libs/1_67_0/libs/polygon/doc/voronoi_diagram.htm

3. Once the coordinates of the Voronoï cell are obtained, we divide the cell into m triangles denoted $(T_\ell)_{\ell=1,\dots,m}$, see Figure 3.4, yielding

$$C_i(\Gamma_N) = \bigcup_{\ell=1}^{m} T_\ell \qquad \text{where} \qquad T_\ell = (x_i^N, z_{\ell-1}, z_\ell).$$

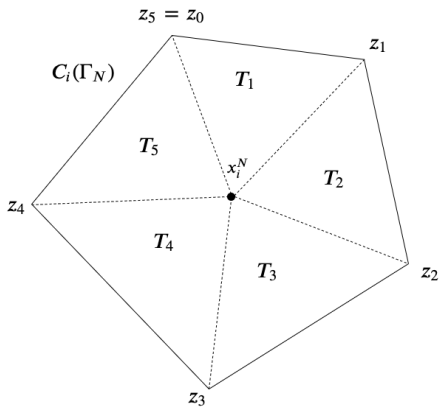


Fig. 3.4 *Example of division of a Voronoï cell* $C_i(\Gamma_N)$ *into* 5 *triangles.*

4. Hence, the integral over the Voronoï cell $C_i(\Gamma_N)$ is equal the sum of the integral over the triangles

$$\int\int_{C_i(\Gamma_N)} f(x,y)dxdy = \sum_{\ell=1}^{m} \int\int_{T_\ell} f(x,y)dxdy.$$

Now, we only need to approximate the integral on each triangle.

Integration over a triangle

1. For a given triangle $T \in \{T_\ell,\ \ell = 1,\dots,m\}$, we transform the integral over that triangle with coordinates $\big((a_x,a_y),(b_x,b_y),(c_x,c_y)\big)$ to an integral over the 2-simplex, denoted S_2 with coordinates $\big((0,0),(1,0),(0,1)\big)$, as detailed in [HKA12]. We use the nodal shape functions for triangles defined by

$$N_1(u,v) = 1-u-v, \qquad N_2(u,v) = u, \qquad N_3(u,v) = v.$$

Hence, if we set

$$\begin{aligned} x = P(u,v) &= a_x N_1(u,v) + b_x N_2(u,v) + c_x N_3(u,v) \\ y = Q(u,v) &= a_y N_1(u,v) + b_y N_2(u,v) + c_y N_3(u,v) \end{aligned}$$

we have

$$\int\int_T f(x,y)dxdy = 2A_T \int\int_{S_2} f\big(P(u,v),Q(u,v)\big)dudv$$

where A_T is the area of the triangle T

$$A_T = \big|(b_x - a_x)(c_y - a_y) - (c_x - a_x)(b_y - a_y)\big|.$$

2. Now, we use a quadrature formula of degree n for general triangular elements which yields

$$\int\int_T f(x,y)dxdy \approx A_T \sum_k \omega_k f\big(P(u_k,v_k),Q(u_k,v_k)\big)$$

where the points (u_k, v_k) and the weights ω_k are provided by the quadrature of order n for the standard triangles as suggested in [Den10] (we considered Gaussian quadrature points for the optimal quantization of Gaussian random vector but any quadrature formula can be used in order to build an optimal quantizer of a chosen random vector X).

Fixed-point search Now, that we are able to evaluate the quantities inside the gradient of the distortion, as in the scalar case, we can implement a fixed-point search using the Lloyd Algorithm 1 with this time the fixed-point operator $\Lambda : (\mathbb{R}^2)^N \to (\mathbb{R}^2)^N$ defined as $\Lambda = (\Lambda_i)_{1\leqslant i\leqslant N}$ defined by

$$\Lambda_i(x) = \frac{\mathbb{E}\left[X\,\mathbb{1}_{X\in C_i(\Gamma_N)}\right]}{\mathbb{P}\left(X \in C_i(\Gamma_N)\right)} = \frac{\displaystyle\int_{C_i(\Gamma_N)} \xi\varphi(\xi)d\xi}{\displaystyle\int_{C_i(\Gamma_N)} \varphi(\xi)d\xi}.$$

Hence, starting from an initial condition x^0 we can apply the Algorithm 1 with Λ defined above where the integrals are approximated using the methodology detailed above.

Gradient descent Again, using that for a given quantizer $x = (x_1^N, \dots, x_N^N)$ we are able to compute the gradient of the distortion, we can apply the mean-field CLVQ gradient descent Algorithm 3 using the following expression for the gradient

$$\nabla\, \mathcal{Q}_{2,\mathrm{N}}(x) = 2\left[x_i^N \int_{C_i(\Gamma_N)} \varphi(\xi)d\xi - \int_{C_i(\Gamma_N)} \xi\varphi(\xi)d\xi\right]_{i=1,\dots,N}.$$

Numerical example We apply this methodology for building optimal quantizers of the two-dimensional Gaussian random vector $\mathcal{N}(0, I_2)$ where I_2 is the identity matrix. We use the knowledge of the density, set $M = 15$ (the value of the squared domain coordinates) and apply the Lloyd method. We display two optimal quantizations of size 100 and 200.

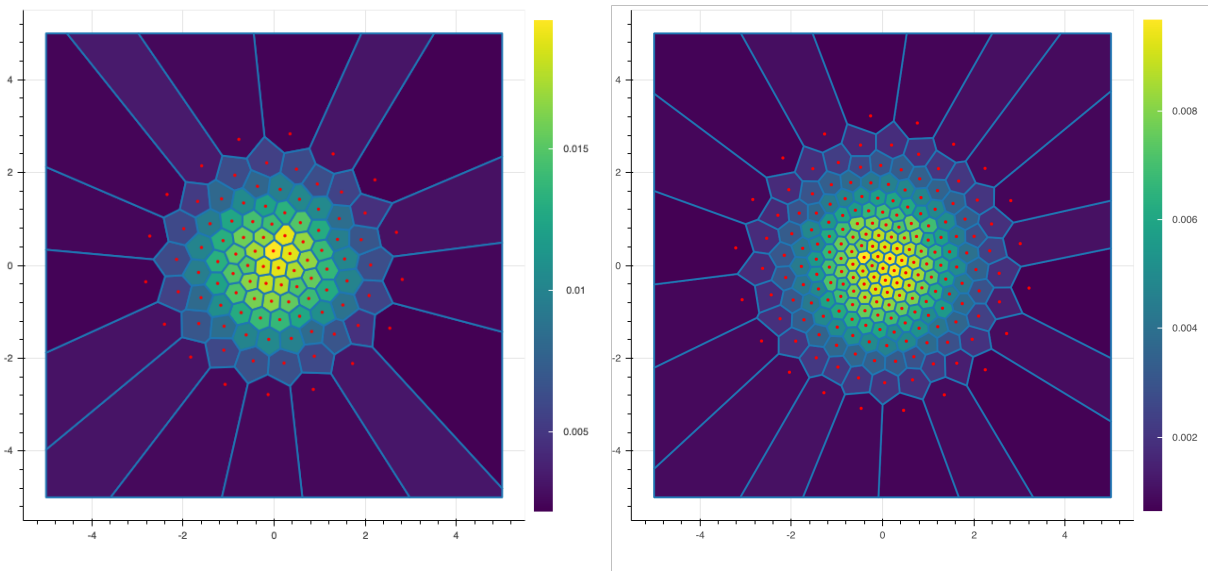


Fig. 3.5 *Two optimal quantizations of size* $N = 100$ *and* $N = 200$ *of a 2-dimensional standard Gaussian vector.*

Appendix 3.A Proof for the formulas of F_X and K_X

Supremum of the Brownian bridge Let $X = \sup_{t\in[0,1]} |W_t - tW_1|$ denotes the supremum of the standard Brownian bridge. This distribution is also known as the Kolmogorov-Smirnov distribution. Let $x \geqslant 0$, the distribution is characterized by its survival function

$$\mathbb{P}(X \geqslant x) = 2\sum_{k\geqslant 1}(-1)^{k-1}\,\mathrm{e}^{-2k^2x^2}\,.$$

The cumulative distribution function is given by $F_X(x) = 1 - \mathbb{P}(X \geqslant x)$, then

$$F_X(x) = 1 - 2\sum_{k\geqslant 1}(-1)^{k-1}\,\mathrm{e}^{-2k^2x^2}\,.$$

In order to obtain the density φ_X, we compute the derivative of F_X

$$\varphi_X(x) = \frac{\partial F_X(x)}{\partial x} = 8x\sum_{k\geqslant 1}(-1)^{k-1}k^2\,\mathrm{e}^{-2k^2x^2},$$

yielding the desired formula.

Now, using the definition of the cumulated partial first moment function K_X of X

$$\begin{aligned}
K_X(x) &= \mathbb{E}\left[X\,\mathbb{1}_{X\leqslant x}\right] = \int_0^x \xi\varphi_X(\xi)d\xi \\
&= xF_X(x) - \int_0^x F_X(\xi)d\xi \\
&= x\Big(1 - 2\sum_{k\geqslant 1}(-1)^{k-1}\,\mathrm{e}^{-2k^2x^2}\Big) - \int_0^x \Big(1 - 2\sum_{k\geqslant 1}(-1)^{k-1}\,\mathrm{e}^{-2k^2\xi^2}\Big)d\xi \\
&= -x\big(1 - F_X(x)\big) + 2\sum_{k\geqslant 1}(-1)^{k-1}\int_0^x \mathrm{e}^{-2k^2\xi^2}\,d\xi.
\end{aligned}$$

Moreover, as

$$\int_0^x \mathrm{e}^{-2k^2\xi^2}\,d\xi = \int_0^{2kx} \mathrm{e}^{-\xi^2/2}\,\frac{d\xi}{2k},$$

we have

$$\begin{aligned}
K_X(x) &= 2\sum_{k\geqslant 1}(-1)^{k-1}\int_0^{2kx} \mathrm{e}^{-\xi^2/2}\,\frac{d\xi}{2k} - x\big(1 - F_X(x)\big) \\
&= \sqrt{2\pi}\sum_{k\geqslant 1}\frac{(-1)^{k-1}}{k}\Big(\mathcal{N}(2kx) - \frac{1}{2}\Big) - x\big(1 - F_X(x)\big),
\end{aligned}$$

which concludes the proof.

Symmetric random variable Let X be a symmetric random variable with characteristic function $\chi(u) = \mathbb{E}\left[\,\mathrm{e}^{\mathbf{i}uX}\,\right]$, where $\mathbf{i}$ is the imaginary number, s.t. $\mathbf{i}^2 = -1$. As X is symmetric,

the characteristic function is χ is real-valued and even: $\chi(u) = \mathbb{E}\left[\,\mathrm{e}^{\mathbf{i}uX}\,\right] = \mathbb{E}\left[\,\mathrm{e}^{-\mathbf{i}uX}\,\right] = \chi(-u)$. Let $x > 0$.

• First, we express the density of X as an alternate series and show that it is even. By definition, $\varphi_X(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{e}^{\mathbf{i}xu}\,\chi(u)du$ and we deduce that the density is even

$$\begin{aligned}\varphi_X(-x) &= \frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{e}^{\mathbf{i}xu}\,\chi(u)du \\ &= -\frac{1}{2\pi}\int_{+\infty}^{-\infty} \mathrm{e}^{-\mathbf{i}xv}\,\chi(-v)dv \qquad\qquad (v=-u) \\ &= \frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{e}^{-\mathbf{i}xv}\,\chi(v)dv = \varphi_X(x).\end{aligned}$$

Then

$$\varphi_X(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{e}^{\mathbf{i}xu}\,\chi(u)du = \frac{1}{\pi}\int_0^{+\infty} \mathrm{e}^{\mathbf{i}xu}\,\chi(u)du$$

and

$$\varphi_X(0) = \frac{1}{\pi}\int_0^{+\infty} \chi(u)du.$$

Hence, we deduce the first desired expression of φ_X

$$\begin{aligned}\varphi_X(x) = \frac{\varphi_X(x)+\varphi_X(-x)}{2} &= \frac{1}{2\pi}\int_0^{+\infty} \big(\,\mathrm{e}^{\mathbf{i}xu} + \mathrm{e}^{-\mathbf{i}xu}\,\big)\chi(u)du \\ &= \frac{1}{\pi}\int_0^{+\infty} \cos(xu)\chi(u)du \\ &= \frac{1}{\pi x}\int_0^{+\infty} \cos(v)\chi\Big(\frac{v}{x}\Big)dv. \qquad\qquad (v=ux)\end{aligned}$$

From this expression, we express φ_X as an alternate series

$$\begin{aligned}\varphi_X(x) &= \frac{1}{\pi x}\int_0^{+\infty} \cos(u)\chi\Big(\frac{u}{x}\Big)du \\ &= \frac{1}{\pi x}\sum_{k\geqslant 0}\int_{\pi k}^{\pi(k+1)} \cos(u)\chi\Big(\frac{u}{x}\Big)du \\ &= \frac{1}{\pi x}\sum_{k\geqslant 0}(-1)^k\int_0^{\pi} \cos(u)\chi\Big(\frac{u+k\pi}{x}\Big)du.\end{aligned}$$

• Now, we focus on the cumulative distribution function F_X. As X is symmetric, then $F_X(0) = \frac{1}{2}$ and for every $x > 0$

$$\begin{aligned}
F_X(x) &= \frac{1}{2} + \int_0^x \varphi_X(\xi)d\xi \\
&= \frac{1}{2} + \int_0^x \frac{1}{\pi}\int_0^{+\infty} \cos(\xi u)\chi(u)du d\xi \\
&= \frac{1}{2} + \frac{1}{\pi}\int_0^{+\infty} \Big(\int_0^x \cos(\xi u)d\xi\Big)\chi(u)du \\
&= \frac{1}{2} + \frac{1}{\pi}\int_0^{+\infty} \frac{\sin(xu)}{u}\chi(u)du \\
&= \frac{1}{2} + \frac{1}{\pi}\int_0^{+\infty} \frac{\sin(v)}{v}\chi\Big(\frac{v}{x}\Big)dv. \qquad (v = ux)
\end{aligned}$$

Then, we deduce $F_X(-x)$

$$\begin{aligned}
F_X(-x) &= \int_{-\infty}^{-x} \varphi_X(\xi)d\xi = 1 - \int_{-x}^{\infty} \varphi_X(\xi)d\xi \\
&= \frac{1}{2} - \int_{-x}^{0} \varphi_X(\xi)d\xi = \frac{1}{2} - \int_0^x \varphi_X(\xi)d\xi \\
&= 1 - F_X(x).
\end{aligned}$$

Finally, we express F_X as an alternate series using the same argument as for the density

$$\begin{aligned}
F_X(x) &= \frac{1}{2} + \frac{1}{\pi}\int_0^{+\infty} \frac{\sin(u)}{u}\chi\Big(\frac{u}{x}\Big)du \\
&= \frac{1}{2} + \frac{1}{\pi}\sum_{k\geqslant 0}(-1)^k \int_0^{\pi} \frac{\sin(u)}{u + k\pi}\chi\Big(\frac{u + k\pi}{x}\Big)du.
\end{aligned}$$

• Next, we focus on the first partial moment function K_X. Let $x > 0$, first, we show that K_X is even

$$\begin{aligned}
K_X(x) &= \int_{-\infty}^{x} \xi\varphi_X(\xi)d\xi \\
&= K_X(-x) + \int_{-x}^{x} \xi\varphi_X(\xi)d\xi \\
&= K_X(-x) + \int_{-x}^{0} u\varphi_X(u)du + \int_0^x v\varphi_X(v)dv \\
&= K_X(-x) - \int_0^{-x} u\varphi_X(u)du + \int_0^x v\varphi_X(v)dv \\
&= K_X(-x) - \int_0^x w\varphi_X(-w)dw + \int_0^x v\varphi_X(v)dv \qquad (w = -u) \\
&= K_X(-x)
\end{aligned}$$

because φ_X is even. Moreover,

$$K_X(0) = \int_{-\infty}^{0} \xi\varphi_X(\xi)d\xi = -\int_{0}^{\infty} \xi\varphi_X(\xi)d\xi = -\int_{\mathbb{R}} \xi\varphi_X(\xi)\,\mathbb{1}_{\xi>0}\,d\xi = -\,\mathbb{E}[X_+].$$

Now, we express K_X as an alternate series

$$\begin{aligned}
K_X(x) &= -\,\mathbb{E}[X_+] + \int_0^x \xi\varphi_X(\xi)d\xi \\
&= -\,\mathbb{E}[X_+] + xF_X(x) - \int_0^x F_X(\xi)d\xi \\
&= -\,\mathbb{E}[X_+] + xF_X(x) - \int_0^x \left(\frac{1}{2} + \frac{1}{\pi}\int_0^{+\infty} \frac{\sin(u\xi)}{u}\chi(u)du\right)d\xi \\
&= -\,\mathbb{E}[X_+] + x\Big(F_X(x) - \frac{1}{2}\Big) - \frac{1}{\pi}\int_0^{+\infty} \frac{-\cos(ux)+1}{u}\frac{\chi(u)}{u}du \\
&= -\,\mathbb{E}[X_+] + x\Big(F_X(x) - \frac{1}{2}\Big) - \frac{x}{\pi}\int_0^{+\infty} \frac{1-\cos(v)}{v^2}\chi\Big(\frac{v}{x}\Big)dv. \qquad (v = ux)
\end{aligned}$$

Finally, we express K_X as an alternate series

$$\begin{aligned}
K_X(x) &= -\,\mathbb{E}[X_+] + x\Big(F_X(x) - \frac{1}{2}\Big) - \frac{x}{\pi}\int_0^{+\infty} \frac{1-\cos(u)}{u^2}\chi\Big(\frac{u}{x}\Big)du \\
&= -\,\mathbb{E}[X_+] + x\Big(F_X(x) - \frac{1}{2}\Big) - \frac{x}{\pi}\sum_{k\geqslant 0}\int_0^{\pi} \frac{1-(-1)^k\cos(u)}{(u+k\pi)^2}\chi\Big(\frac{u+k\pi}{x}\Big)du.
\end{aligned}$$

Chapter 4

New Weak Error bounds and expansions for Optimal Quantization

This chapter corresponds to the article "New Weak Error bounds and expansions for Optimal Quantization" published in *Journal of Computational and Applied Mathematics* (see [LMP19]). This paper is a joint work with Vincent Lemaire and Gilles Pagès and it is accessible in arXiv or HAL.

Abstract We propose new weak error bounds and expansion in dimension one for optimal quantization-based cubature formula for different classes of functions, such that piecewise affine functions, Lipschitz convex functions or differentiable function with piecewise-defined locally Lipschitz or α-Hölder derivatives. These new results rest on the local behaviours of optimal quantizers, the L^r-L^s distribution mismatch problem and Zador's Theorem. This new expansion supports the definition of a Richardson-Romberg extrapolation yielding a better rate of convergence for the cubature formula. An extension of this expansion is then proposed in higher dimension for the first time. We then propose a novel variance reduction method for Monte Carlo estimators, based on one dimensional optimal quantizers.

Introduction

Optimal quantization was first introduced in [She97], Sheppard worked on optimal quantization of the uniform distribution on unit hypercubes. It was then extended to more general distributions with applications to Signal transmission at the Bell Laboratory in the 50's (see [GG82]) and then developed as a numerical method in the early 90's, for expectation approximations (see [Pag98]) and later for conditional expectation approximations (see [PPP04b; BPP01; BP03; BPP05]).

In modern terms, vector quantization consists in finding the projection for the L^p-Wasserstein distance of a probability measure on $\mathbb{R}^d$ with a finite p-th moment on the convex subset of

Γ-supported probability measure, where Γ is a finite subset of $\mathbb{R}^d$ and $0 < p < +\infty$. The aim of Optimal Quantization is to determine the set $\Gamma_N := \{x_1^N, \dots, x_N^N\} \subset \mathbb{R}^d$ with cardinality at most N which minimizes this distance among all such sets Γ. Formally, if we consider a random vector $X \in L^p(\mathbb{P})$, we search for Γ_N, the solution to the following problem

$$\min_{\Gamma_N \subset \mathbb{R}, |\Gamma_N| \leqslant N} \|X - \widehat{X}^{\Gamma_N}\|_p$$

where $\widehat{X}^{\Gamma_N}$ denotes the projection of X onto Γ_N (often $\widehat{X}^{\Gamma_N}$ is denoted by $\widehat{X}^N$ in order to alleviate the notations). The term $\|X - \widehat{X}^{\Gamma_N}\|_p$ is often referred to as the distortion of order p. The existence of an optimal quantizer at a given level N has been shown in [GL00; Pag98] and in the one-dimensional case if the distribution of X is absolutely continuous with a *log-concave* density then there exists a unique optimal quantizer at level N. In the present paper we will consider one dimensional optimal quantizers. Moreover, we are not only interested by the existence of such a quantizer but also in the asymptotic behaviour of the distortion because it is an important feature for the method in order to determine the level of the error introduced by the approximation. The question concerning the sharp rate of convergence of $\|X - \widehat{X}^N\|_p$ as N goes to infinity is answered by Zador's Theorem. For $X \in L^{p+\delta}(\mathbb{P})$, $\delta > 0$, such that $\mathbb{P}_X(d\xi) = \varphi(\xi) \cdot \lambda(d\xi) + \nu(d\xi)$, where $\nu \perp \lambda$ is the singular component of $\mathbb{P}_X$ with respect to the Lebesgue measure λ on $\mathbb{R}^d$, the rate of convergence is given by

$$\lim_{N \to +\infty} N^{\frac{1}{d}} \|X - \widehat{X}^N\|_p = \widetilde{J}_{p,d} \left[\int_{\mathbb{R}^d} \varphi^{\frac{d}{d+p}} d\lambda_d \right]^{\frac{1}{p} + \frac{1}{d}}$$

where φ is the density of X, λ_d is the Lebesgue measure on $\mathbb{R}^d$ and $\widetilde{J}_{p,d} = \inf_{N \geqslant 1} N^{\frac{1}{d}} \|U - \widehat{U}^N\|_p$, $U \overset{\mathcal{L}}{\sim} \mathcal{U}\big((0,1)^d\big)$. For more insights on the mathematical/probabilistic aspects of Optimal quantization theory, we refer to [GL00; Pag15].

The reason for which we are interested in this optimal quantizer is numerical integration. The discrete feature of the optimal quantizer $\widehat{X}^N$ allows us to define, for every continuous function $f : \mathbb{R}^d \longrightarrow \mathbb{R}$, such that $f(X) \in L^2(\mathbb{P})$, the following quantization-based cubature formula

$$\mathbb{E}\left[f(\widehat{X}^N)\right] = \sum_{i=1}^{N} p_i f(x_i^N)$$

where $p_i = \mathbb{P}(\widehat{X}^N = x_i^N)$. Indeed, as $\widehat{X}^N$ is constructed as the best discrete approximation of X in $L^p(\mathbb{P})$, it is reasonable to approximate $\mathbb{E}\left[f(X)\right]$ by $\mathbb{E}\left[f(\widehat{X}^N)\right]$ which is useful for numerical integrations problems.

The problem of numerical integration appears a lot in applied fields, such as Physics, Computer Sciences or Numerical Probability. For example, in Quantitative Finance, many

quantities of interest are of the form

$$\mathbb{E}\left[f(S_t)\right] \qquad \text{for some } t > 0,$$

where $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ is a Borel function and $(S_s)_{s\in[0,t]}$ is a diffusion process solution to a Stochastic Differential Equation (SDE)

$$S_t = S_0 + \int_0^t b(s, S_s)ds + \int_0^t \sigma(s, S_s)dW_s, \qquad S_0 = s_0,$$

where W is a standard Brownian motion living on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and b and σ are Lipschitz continuous in x uniformly with respect to $s \in [0, t]$, which are the standard assumptions in order to ensure existence and uniqueness of a strong solution to the SDE. Since it is often impossible to compute $\mathbb{E}\left[f(S_t)\right]$ directly, it has been proposed in [Pag98] to compute an optimal quantizer $\widehat{X}^N$ of X where X is a random variable having the same distribution as S_t and to use the previously defined quantization-based cubature formula as an approximation.

Another approach, often used in order to approximate $\mathbb{E}\left[f(X)\right]$, is to perform a Monte Carlo simulation $\widehat{I}_M := \sum_{m=1}^M f(X^m)$, where $(X^m)_{m=1,\dots,M}$ is a sequence of independent copies of X. The method's rate of convergence is determined by the strong law of numbers and the central limit theorem, which says that if X is square integrable, then

$$\sqrt{M}\Big(\widehat{I}_M - \mathbb{E}\left[f(X)\right]\Big) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \sigma^2_{f(X)}\right) \quad \text{as} \quad M \to +\infty$$

where $\sigma^2_{f(X)} = \mathbb{V}\text{ar}\left(f(X)\right)$. One notices that, for a given M, the limiting factor of the method is $\sigma^2_{f(X)}$. Hence, a lot of methods have been developed in order to reduce the variance term: antithetic variables, control variates, importance sampling, etc. The reader can refer to [Pag18; Gla13] for more details concerning the Monte Carlo methodology and the variance reduction methods.

In this paper we propose a novel variance reduction method of Monte Carlo estimator through quantization. Our method innovates in that it uses a linear combination of one dimensional control variates to reduce the variance of a higher dimensional problem. More precisely, we introduce a quantization-based control variates Ξ^N_k for $k = 1, \dots, d$. If one considers a function $f : \mathbb{R}^d \mapsto \mathbb{R}$, we approximate $\mathbb{E}\left[f(X)\right]$ by

$$\mathbb{E}\left[f(X) - \langle \lambda, \Xi^N \rangle\right]$$

with $\langle \cdot, \cdot \rangle$ the scalar product in $\mathbb{R}^d$ and $(\Xi^N_k)_{k=1,\dots,d} := f_k(X_k) - \mathbb{E}\left[f_k(\widehat{X}^N_k)\right]$, where X_k is the k-th component of X, $\widehat{X}^N_k$ is an optimal quantizer of X_k of size N and $f_k : \mathbb{R} \mapsto \mathbb{R}$ is designed from f. Looking closely at the introduced control variates, one notices that we introduce a bias in the approximation. However, as since it is closely linked to weak error, this bias can be controlled. The present paper focuses on the weak error's rate of convergence.

First, we place ourselves in the case where X is a random variable in dimension one and we consider a quadratic optimal quantizer. We work on the rate of convergence of the weak error induced by the expectation approximation by an optimal quantization-based cubature formula for different classes of functions f

$$\lim_{N\to+\infty} N^{\alpha}\big|\,\mathbb{E}\left[f(X)\right]-\mathbb{E}\left[f(\widehat{X}^N)\right]\big|\leqslant C_{f,X}<+\infty.$$

The first classical result concerns Lipschitz continuous functions. Using directly the Lipschitz continuity property of f and Zador's Theorem a rate of order $\alpha=1$ can be obtained. Moreover, if we consider the supremum among all functions with a Lipschitz constant upper-bounded by 1, then

$$N\sup_{[f]_{Lip}\leqslant 1}\big|\,\mathbb{E}\left[f(X)\right]-\mathbb{E}\left[f(\widehat{X}^N)\right]\big|=N\|X-\widehat{X}^N\|_1\leqslant N\|X-\widehat{X}^N\|_2\xrightarrow{N\to+\infty}C_f<+\infty.$$

A faster rate ($\alpha=2$) can be attained for differentiable functions with Lipschitz continuous derivative, using a Taylor expansion with integral remainder and the following stationarity property of quadratic optimal quantizers

$$\mathbb{E}\left[X\mid\widehat{X}^N\right]=\widehat{X}^N.$$

Moreover, considering the supremum among all functions where the Lipschitz constant of the derivative is upper-bounded by 1, we have

$$N^2\sup_{[f']_{Lip}\leqslant 1}\big|\,\mathbb{E}\left[f(X)\right]-\mathbb{E}\left[f(\widehat{X}^N)\right]\big|=\frac{1}{2}N^2\|X-\widehat{X}^N\|_2^2\xrightarrow{N\to+\infty}C_f<+\infty$$

where the limit is given by Zador's Theorem. A detailed summary about this results can be found in [Pag18].

In the first part of this paper, we extend this improved rate ($\alpha=2$) to classes of less smooths functions in one dimension. These new results enable us to design efficient variance reduction methods in higher dimensional settings with in view applications to option pricing. The new results concerns the following classes of functions

- Lipschitz continuous piecewise affine functions with finitely many breaks of affinity. We use the stationarity property of the optimal quantizer on the cells where there is no break of affinity and then we control the error on the remaining cells using results on the local behaviour of the quantizer.

- Lipschitz continuous convex functions, using local behaviours results on optimal quantizers. We use a representation formula for convex functions as integrals of Ridge functions combined with the local behaviour result in order to control the error again.

- Differentiable functions with piecewise-defined locally Lipschitz derivative. The functions have K breaks of affinity $\{a_1, \dots, a_K\}$, such that $-\infty = a_0 < a_1 < \cdots < a_K < a_{K+1} = +\infty$ and the locally Lipschitz property of the derivative is defined by

$$\forall k = 0, \dots, K, \quad \forall x, y \in (a_k, a_{k+1}) \quad |f'(x) - f'(y)| \leqslant [f']_{k,Lip,loc}|x - y|\big(g_k(x) + g_k(y)\big)$$

  where $g_k : (a_k, a_{k+1}) \to \mathbb{R}_+$ are non-negative Borel functions. We use the locally Lipschitz property of the derivative combined with the L^r-L^s distortion Theorem and Zador's Theorem on the cells where there is no break of affinity and then we control the error on the remaining cells using results on the local behaviour of the quantizer.

- Differentiable functions with piecewise-defined locally α-Hölder derivative. The functions have K breaks of affinity $\{a_1, \dots, a_K\}$, such that $-\infty = a_0 < a_1 < \cdots < a_K < a_{K+1} = +\infty$ and the locally α-Hölder property of the derivative is defined by

$$\forall k = 0, \dots, K, \quad \forall x, y \in (a_k, a_{k+1}), \quad |f'(x) - f'(y)| \leqslant [f']_{k,\alpha,loc}|x - y|^\alpha\big(g_k(x) + g_k(y)\big)$$

  where $g_k : (a_k, a_{k+1}) \to \mathbb{R}_+$ are non-negative Borel functions. For this class of functions, the rate of convergence is of order $1 + \alpha$. The result is obtained using the same ideas as in the locally Lipschitz case.

Hence, for all this classes of functions, except the last one, we have

$$\lim_{N \to +\infty} N^2 \big| \mathbb{E}\big[f(X)\big] - \mathbb{E}\big[f(\widehat{X}^N)\big] \big| \leqslant C_{f,X} < +\infty.$$

In the second part of the paper we deal with the *weak error expansion* of the approximation of $\mathbb{E}\big[f(X)\big]$ by $\mathbb{E}\big[f(\widehat{X}^N)\big]$. First, we place ourselves in the one dimensional case by considering a twice differentiable function $f : \mathbb{R} \mapsto \mathbb{R}$ with a bounded Lipschitz continuous second derivative and $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$. Through a second order Taylor expansion and with the help of Corollary 4.1.7, Theorem 4.1.12 and the L^r-L^s distortion mismatch Theorem we obtain

$$\mathbb{E}\big[f(X)\big] = \mathbb{E}\big[f(\widehat{X}^N)\big] + \frac{c_2}{N^2} + O(N^{-(2+\beta)})$$

where $\beta \in (0, 1)$. This expression suggests to use a Richardson-Romberg extrapolation in order to *kill* the first term of the expansion which yields

$$\mathbb{E}\big[f(X)\big] = \mathbb{E}\left[\frac{M^2 f(\widehat{X}^M) - N^2 f(\widehat{X}^N)}{M^2 - N^2}\right] + O(N^{-(2+\beta)}).$$

Second, we present a result in higher dimension when considering a twice differentiable function $f : \mathbb{R}^d \mapsto \mathbb{R}$ with a bounded Lipschitz continuous Hessian, $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}^d$ with independent components $(X_k)_{k=1,\dots,d}$ and $\widehat{X}^N$ a product quantizer of X with d components $(\widehat{X}_k^{N_k})_{k=1,\dots,d}$

such that $N_1 \times \cdots \times N_d \simeq N$. Using product quantizer allows us to rely on the one dimensional results for quadratic optimal quantizers and in that case we have

$$\mathbb{E}\left[f(X)\right] = \mathbb{E}\left[f(\widehat{X}^N)\right] + \sum_{k=1}^{d} \frac{c_k}{N_k^2} + O\left(\left(\min_{k=1:d} N_k\right)^{-(2+\beta)}\right).$$

The paper is organized as follows. First we recall some basic facts and deeper results about optimal quantization in Section 4.1. In Section 4.2, we present our new results on weak error for some classes of functions. Then, we see in Section 4.3 how to derive *weak error expansion* allowing us to specify the right hypothesis under which we can use a Richardson-Romberg extrapolation. Finally, we conclude with some applications. The first one is the introduction of our novel variance reduction involving optimal quantizers. The last one illustrates numerically the results shown in Section 4.2 and 4.3, by considering a Black-Scholes model and pricing different types of European Options. We also propose a numerical example for the variance reduction.

4.1 About optimal quantization ($d = 1$)

Let X be a $\mathbb{R}$-valued random variable with distribution $\mathbb{P}_X$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that $X \in L^2(\mathbb{P})$.

Definition 4.1.1. Let $\Gamma_N = \{x_1^N, \ldots, x_N^N\} \subset \mathbb{R}$ be a subset of size N, called N-quantizer. A Borel partition $\left(C_i(\Gamma_N)\right)_{i=1,\ldots,N}$ of $\mathbb{R}$ is a Voronoï partition of $\mathbb{R}$ induced by the N-quantizer Γ_N if, for every $i = 1, \ldots, N$,

$$C_i(\Gamma_N) \subset \left\{\xi \in \mathbb{R}, |\xi - x_i^N| \leqslant \min_{j \neq i} |\xi - x_j^N|\right\}.$$

The Borel sets $C_i(\Gamma_N)$ are called Voronoï cells of the partition induced by Γ_N.

One can always consider that the quantizers are ordered: $x_1^N < x_2^N < \cdots < x_{N-1}^N < x_N^N$ and in that case the Voronoï cells are given by

$$C_k(\Gamma_N) = (x_{k-1/2}^N, x_{k+1/2}^N], \qquad k = 1, \ldots, N-1, \qquad C_N(\Gamma_N) = (x_{N-1/2}^N, x_{N+1/2}^N)$$

where $\forall k = 2, \ldots, N$, $x_{k-1/2}^N := \frac{x_{k-1}^N + x_k^N}{2}$ and $x_{1/2}^N := \inf\left(\mathrm{supp}(\mathbb{P}_X)\right)$ and $x_{N+1/2}^N := \sup\left(\mathrm{supp}(\mathbb{P}_X)\right)$.

Definition 4.1.2. Let $\Gamma_N = \{x_1^N, \ldots, x_N^N\}$ be an N-quantizer. The nearest neighbour projection $\mathrm{Proj}_{\Gamma_N} : \mathbb{R} \to \{x_1^N, \ldots, x_N^N\}$ induced by a Voronoï partition $\left(C_i(\Gamma_N)\right)_{i=1,\ldots,N}$ is defined by

$$\forall \xi \in \mathbb{R}, \qquad \mathrm{Proj}_{\Gamma_N}(\xi) := \sum_{i=1}^{N} x_i^N \, \mathbb{1}_{\xi \in C_i(\Gamma_N)}.$$

We can now define the quantization of X by composing Proj_{Γ_N} and X

$$\widehat{X}^{\Gamma_N} = \mathrm{Proj}_{\Gamma_N}(X) = \sum_{i=1}^{N} x_i^N \mathbb{1}_{X \in C_i(\Gamma_N)}$$

and the point-wise error induced by the replacement of X by $\widehat{X}^{\Gamma_N}$ given by

$$|X - \widehat{X}^{\Gamma_N}| = \mathrm{dist}\left(X, \{x_1^N, \ldots, x_N^N\}\right) = \min_{i=1,\ldots,N} |X - x_i^N|.$$

In order to alleviate the notations, from now on we write $\widehat{X}^N$ in place of $\widehat{X}^{\Gamma_N}$.

Definition 4.1.3. The L^2-mean (or mean quadratic) quantization error induced by the replacement of X by the quantization of X using a N-quantizer $\Gamma_N \subset \mathbb{R}$ is defined as the quadratic norm of the point-wise error previously defined

$$\|X - \widehat{X}^N\|_2 := \left(\mathbb{E}\left[\min_{i=1,\ldots,N} |X - x_i^N|^2\right]\right)^{1/2} = \left(\int_{\mathbb{R}} \min_{i=1,\ldots,N} |\xi - x_i^N|^2 \, \mathbb{P}_X(d\xi)\right)^{1/2}.$$

It is convenient to define the quadratic distortion function at level N as the squared mean quadratic quantization error on $(\mathbb{R})^N$:

$$\mathcal{Q}_{2,\mathrm{N}} : x = (x_1^N, \ldots, x_N^N) \longmapsto \mathbb{E}\left[\min_{i=1,\ldots,N} |X - x_i^N|^2\right] = \|X - \widehat{X}^N\|_2^2.$$

Remark. All these definitions can be extended to the L^p case. For example the L^p-mean quantization error induced by a quantizer of size N is

$$\|X - \widehat{X}^N\|_p := \left(\mathbb{E}\left[\min_{i=1,\ldots,N} |X - x_i^N|^p\right]\right)^{1/p} = \left(\int_{\mathbb{R}} \min_{i=1,\ldots,N} |X - x_i^N|^p \, \mathbb{P}_X(d\xi)\right)^{1/p}.$$

We briefly recall some classical theoretical results, see [GL00; Pag18] for further details.

Theorem 4.1.4. *(Existence of optimal N-quantizers) Let $X \in L^2(\mathbb{P})$ and $N \in \mathbb{N}^*$.*

(a) The quadratic distortion function $\mathcal{Q}_{2,\mathrm{N}}$ at level N attains a minimum at an N-tuple $x^{(N)} = (x_1^N, \ldots, x_N^N)$ and $\Gamma_N = \{x_i^N, i = 1, \ldots, N\}$ is a quadratic optimal quantizer at level N.

(b) If the support of the distribution $\mathbb{P}_X$ of X has at least N elements, then $x^{(N)} = (x_1^N, \ldots, x_N^N)$ has pairwise distinct components, $\mathbb{P}_X\left(C_i(x^{(N)})\right) > 0$, $i = 1, \ldots, N$. Furthermore, the sequence $N \mapsto \inf_{x \in (\mathbb{R})^N} \mathcal{Q}_{2,\mathrm{N}}(x)$ converges to 0 and is decreasing as long as it is positive.

Following the existence of a minimum for $\mathcal{Q}_{2,\mathrm{N}}$ at $x^{(N)}$, we can define an optimal quadratic N-quantizer.

Definition 4.1.5. A grid associated to any N-tuple solution to the above distortion minimization problem is called an optimal quadratic N-quantizer.

A really interesting and useful property concerning quadratic optimal quantizers is the stationarity property.

Proposition 4.1.6. *(Stationarity) Assume that the support of $\mathbb{P}_X$ has at least N elements. Any L^2-optimal N-quantizer $\Gamma_N \in (\mathbb{R})^N$ is stationary in the following sense: for every Voronoï quantization $\widehat{X}^N$ of X,*

$$\mathbb{E}\left[X \mid \widehat{X}^N\right] = \widehat{X}^N.$$

Corollary 4.1.7. *If $\widehat{X}^N$ is a L^2-optimal quantization of X, hence has the above stationarity property, and $f(X) \in L^2(\mathbb{P})$ with $f : \mathbb{R} \to \mathbb{R}$ then*

$$\mathbb{E}\left[f(\widehat{X}^N)(X - \widehat{X}^N)\right] = 0.$$

Proof. The proof is straightforward, indeed

$$\begin{aligned}\mathbb{E}\left[f(\widehat{X}^N)(X - \widehat{X}^N)\right] &= \mathbb{E}\left[\mathbb{E}\left[f(\widehat{X}^N)(X - \widehat{X}^N) \mid \widehat{X}^N\right]\right] = \mathbb{E}\left[f(\widehat{X}^N)\,\mathbb{E}[X - \widehat{X}^N \mid \widehat{X}^N]\right] \\ &= \mathbb{E}\left[f(\widehat{X}^N)\big(\mathbb{E}\left[X \mid \widehat{X}^N\right] - \widehat{X}^N\big)\right] = 0.\end{aligned}$$

□

We now take a look at the asymptotic behaviour in N of the quadratic mean quantization error. We saw in Theorem 4.1.4 that the infimum of the quadratic distortion converges to 0 as N goes to infinity. The next Theorem, known as Zador's Theorem, analyzes the rate of convergence of the L^p-mean quantization error.

Theorem 4.1.8. *(Zador's Theorem) Let $p \in (0, +\infty)$.*

(a) Sharp rate. *Let $X \in L^{p+\delta}(\mathbb{P})$ for some $\delta > 0$. Let $\mathbb{P}_X(d\xi) = \varphi(\xi) \cdot \lambda(d\xi) + \nu(d\xi)$, where $\nu \perp \lambda$ is the singular component of $\mathbb{P}_X$ with respect to the Lebesgue measure λ on $\mathbb{R}$. Then*

$$\lim_{N \to +\infty} N \min_{\Gamma_N \subset \mathbb{R}, |\Gamma_N| \leqslant N} \|X - \widehat{X}^N\|_p = \widetilde{J}_{p,1} \left[\int_{\mathbb{R}} \varphi^{\frac{1}{1+p}} d\lambda\right]^{1+\frac{1}{p}}$$

with $\widetilde{J}_{p,1} = \frac{1}{2^p(p+1)}$.

(b) Non asymptotic upper-bound. *Let $\delta > 0$. There exists a real constant $C_{1,p,\delta} \in (0, +\infty)$ such that, for every $\mathbb{R}$-valued random variable X,*

$$\forall N \geqslant 1, \qquad \min_{\Gamma_N \subset \mathbb{R}, |\Gamma_N| \leqslant N} \|X - \widehat{X}^N\|_p \leqslant C_{1,p,\delta} \sigma_{\delta+p}(X) N^{-1}$$

where, for $r \in (0, +\infty)$, $\sigma_r(X) = \min_{a \in \mathbb{R}} \|X - a\|_r < +\infty$.

Now, we state some intuitive but remarkable results concerning the local behaviour of the optimal quantizers.

Lemma 4.1.9. *Let $\mathbb{P}_X$ be a distribution on the real line with connected support $I_{\mathbb{P}_X} := \mathrm{supp}(\mathbb{P}_X)$. Let $\Gamma_N = \{x_1^N, \ldots, x_N^N\}$ be a sequence of r-optimal quantizers, $r > 0$. Let $[a, b]$, be a closed interval then*

$$\bigcup_N \bigcup_{C_i(\Gamma_N)\cap[a,b]\neq\varnothing} C_i(\Gamma_N) \subset K_0$$

where K_0 is a compact set.

Proof. First, if $+\infty \notin \overline{I_{\mathbb{P}_X}}$ then the upper-bound of K_0 is the upper-bound of $\overline{I_{\mathbb{P}_X}}$ otherwise if $+\infty \in \overline{I_{\mathbb{P}_X}}$, let $b_0 \in I_{\mathbb{P}_X}$ such that $b_0 < b$, as $\mathbb{P}_X$ has a density, then $\mathbb{P}_X\left(\{b_0\}\right) = \mathbb{P}_X\left(\{b\}\right) = 0$. Considering the weighted empirical measure

$$\mathbb{P}_{\widehat{X}^N} := \sum_{i=1}^{N} \mathbb{P}_X\left(C_i(\Gamma_N)\right)\delta_{x_i^N} \xrightarrow{N\to+\infty} \mathbb{P}_X$$

then $\mathbb{P}_{\widehat{X}^N}\left([b_0, b]\right) \xrightarrow{N\to+\infty} \mathbb{P}_X\left([b_0, b]\right) < \mathbb{P}_X\left([b_0, +\infty)\right)$. Moreover, one notices that

$$\mathbb{P}_{\widehat{X}^N}\left([b_0, b]\right) = \mathbb{P}_X\left(\bigcup_{i\in\{i_{b_0},\ldots,i_b\}} C_i(\Gamma_N)\right) = \mathbb{P}_{\widehat{X}^N}\left(\bigcup_{i\in\{i_{b_0},\ldots,i_b\}} C_i(\Gamma_N)\right)$$

where $x_{i_u}^N$ is the centroid of the cell that contains u. Then, as $[b_0, x_{i_b+1/2}^N] \subset \bigcup_{i\in\{i_{b_0},\ldots,i_b\}} C_i(\Gamma_N)$

$$\mathbb{P}_X\left([b_0, x_{i_b+1/2}^N]\right) \leqslant \mathbb{P}_{\widehat{X}^N}\left([b_0, b]\right) \xrightarrow{N\to+\infty} \mathbb{P}_X\left([b_0, b]\right) < \mathbb{P}_X\left([b_0, +\infty)\right)$$

hence, $\limsup_N x_{i_b+1/2}^N < +\infty$ and $\sup_N x_{i_b+1/2}^N < +\infty$, which gives us the upper-bound of K_0: $\sup_N x_{i_b+1/2}^N$.

Finally, if $-\infty \notin \overline{I_{\mathbb{P}_X}}$ then the lower-bound of K_0 is the lower-bound of $\overline{I_{\mathbb{P}_X}}$ otherwise if $-\infty \in \overline{I_{\mathbb{P}_X}}$, then following the same idea as above, we can apply the same deductions in order to show that $\inf_N x_{i_a-1/2}^N > -\infty$ which gives us the lower-bound of K_0: $\inf_N x_{i_a-1/2}^N$. In conclusion, $K_0 := \mathrm{supp}(\mathbb{P}_X) \bigcap [\inf_N x_{i_a-1/2}^N, \sup_N x_{i_b+1/2}^N]$. □

The next result, proved in [DFP04], deals with the local behaviour of optimal quantizer, more precisely it characterises the rate of convergence, in function of N, of the weights and the local distortions associated to an optimal quantizer. This is the key result of the first part of this paper. It allows us to extend the weak error bound of order two to less regular functions than those originally considered in [Pag98], namely differentiable functions with Lipschitz continuous derivative.

Theorem 4.1.10. *(Local behaviour of optimal quantizers) Let $\mathbb{P}_X$ be a distribution on the real line with connected support* $\mathrm{supp}(\mathbb{P}_X)$. *Assume that $\mathbb{P}_X$ has a probability density function*

φ which is positive and Lipschitz continuous on every compact set of the interior $(\underline{m}, \overline{m})$ of $\mathrm{supp}(\mathbb{P}_X)$. *Let $\Gamma_N = \{x_1^N, \dots, x_N^N\}$ be a sequence of stationary and L^r optimal quantizers, $r > 0$.*

(a) The sequence of functions $(\psi_N)_{N\geqslant 1}$ defined by

$$\psi_N(\xi) := N \sum_{i=1}^{N} \mathbb{1}_{C_i(\Gamma_N)}(\xi)\, \mathbb{P}_X\left(C_i(\Gamma_N)\right), \quad N \geqslant 1,$$

converges uniformly on compact sets of $(\underline{m}, \overline{m})$ towards $c_{\varphi,1/(r+1)}\varphi^{\frac{r}{r+1}}$, with $c_{\varphi,1/(r+1)} = \|\varphi\|_{1/(1+r)}^{-1/(1+r)}$ i.e., for every $[a,b] \subset (\underline{m}, \overline{m})$, $a < b$,

$$\sup_{\{i:x_i^N\in[a,b]\}} \left| N\, \mathbb{P}_X\left(C_i(\Gamma_N)\right) - c_{\varphi,1/(r+1)}\varphi^{\frac{r}{r+1}}(x_i^N) \right| \xrightarrow{N\to+\infty} 0. \tag{4.1}$$

The local distortion is asymptotically uniformly distributed i.e., for every $[a,b] \subset (\underline{m}, \overline{m})$,

$$\sup_{\{i:x_i^N\in[a,b]\}} \left| N^{r+1} \int_{C_i(\Gamma_N)} |x_i^N - \xi|^r\, \mathbb{P}_X(d\xi) - \frac{\|\varphi\|_{1/(r+1)}}{2^r(r+1)} \right| \xrightarrow{N\to+\infty} 0. \tag{4.2}$$

(b) Moreover, if $\mathbb{P}_X$ has a compact support $[\underline{m}, \overline{m}]$ and φ is bounded away from 0 on the whole interval $[m, M]$, then all the above convergences hold uniformly on $[\underline{m}, \overline{m}]$.

The next result is a weaker version of Theorem 4.1.10 but it is a really useful tool when dealing with weak error induced by quantization-based cubature formulas.

Corollary 4.1.11. *Under the same hypothesis as in Theorem 4.1.10 and if $1 \leqslant s \leqslant r$, we have the following result, for every $i \in \{1, \dots, N\}$,*

$$\limsup_N N^{s+1} \int_{C_i(\Gamma_N)} |x_i^N - \xi|^s\, \mathbb{P}_X(d\xi) = \limsup_N N^{s+1}\, \mathbb{E}\left[|\widehat{X}^N - X|^s\, \mathbb{1}_{\{\widehat{X}^N = x_i^N\}} \right] < +\infty.$$

Proof. If $s = 1$, using Schwarz's inequality

$$\int_{C_i(\Gamma_N)} |x_i^N - \xi|\, \mathbb{P}_X(d\xi) \leqslant \left(\int_{C_i(\Gamma_N)} |x_i^N - \xi|^2\, \mathbb{P}_X(d\xi) \cdot \mathbb{P}_X\left(C_i(\Gamma_N)\right) \right)^{\frac{1}{2}}$$

$$\iff \quad N^2 \int_{C_i(\Gamma_N)} |x_i^N - \xi|\, \mathbb{P}_X(d\xi) \leqslant \left(N^3 \int_{C_i(\Gamma_N)} |x_i^N - \xi|^2\, \mathbb{P}_X(d\xi) \cdot N\, \mathbb{P}_X\left(C_i(\Gamma_N)\right) \right)^{\frac{1}{2}}.$$

And applying Theorem 4.1.10 with $\mathbb{P}_X = \varphi \cdot \lambda$ and $r = 2$, one derives

$$\limsup_N N^2 \int_{C_i(\Gamma_N)} |x_i^N - \xi|\, \mathbb{P}_X(d\xi) \leqslant \frac{1}{2\sqrt{3}} \left(c_{\varphi,1/3} \|\varphi\|_{1/3} \|\varphi^{2/3}\|_\infty \right)^{\frac{1}{2}} < +\infty.$$

Otherwise, for $1 < s < r$, using Hölder's inequality with $p = \frac{1}{s}$ and $q = \frac{1}{1-s}$

$$\begin{aligned}\int_{C_i(\Gamma_N)} |x_i^N - \xi|^s \, \mathbb{P}_X(d\xi) &\leqslant \left(\int_{C_i(\Gamma_N)} |x_i^N - \xi|^{ps} \, \mathbb{P}_X(d\xi) \right)^{1/p} \left(\int_{C_i(\Gamma_N)} \mathbb{P}_X(d\xi) \right)^{1/q} \\ &\leqslant \left(\int_{C_i(\Gamma_N)} |x_i^N - \xi| \, \mathbb{P}_X(d\xi) \right)^s \left(\mathbb{P}_X\left(C_i(\Gamma_N)\right) \right)^{1-s} \\ \iff N^{s+1} \int_{C_i(\Gamma_N)} |x_i^N - \xi|^s \, \mathbb{P}_X(d\xi) &\leqslant N^{s+1} \left(\int_{C_i(\Gamma_N)} |x_i^N - \xi| \, \mathbb{P}_X(d\xi) \right)^s \left(\mathbb{P}_X\left(C_i(\Gamma_N)\right) \right)^{1-s} \\ &\leqslant \left(N^2 \int_{C_i(\Gamma_N)} |x_i^N - \xi| \, \mathbb{P}_X(d\xi) \right)^s \left(N \, \mathbb{P}_X\left(C_i(\Gamma_N)\right) \right)^{1-s}.\end{aligned}$$

And using the result proved above for $s = 1$ and (4.1), we obtain the desired result

$$\begin{aligned}\limsup_N N^{s+1} \int_{C_i(\Gamma_N)} |x_i^N - \xi|^s \, \mathbb{P}_X(d\xi) &\\ &\leqslant \limsup_N \left(N^2 \int_{C_i(\Gamma_N)} |x_i^N - \xi| \, \mathbb{P}_X(d\xi) \right)^s \left(N \, \mathbb{P}_X\left(C_i(\Gamma_N)\right) \right)^{1-s} \\ &\leqslant \left(\frac{1}{12} \|\varphi\|_{1/3} \right)^{\frac{s}{2}} \left(c_{\varphi,1/3} \|\varphi^{2/3}\|_\infty \right)^{1-\frac{s}{2}} \\ &< +\infty.\end{aligned}$$

□

The following result will be useful in the last part of the paper, which is the Theorem 6 in [Del+04].

Theorem 4.1.12. *Let $(\Gamma_N)_{N \geqslant 1}$ a sequence of optimal quantizers for $\mathbb{P}_X$. Then*

$$\lim_{N \to +\infty} N^2 \, \mathbb{E}\left[g(\widehat{X}^N) |X - \widehat{X}^N|^2 \right] = \mathcal{Q}_2(\mathbb{P}_X) \int g(\xi) \, \mathbb{P}_X(d\xi)$$

for every function $g : \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}\left[g(X)\right] < +\infty$, with $\mathcal{Q}_2(\mathbb{P}_X)$ the Zador's constant.

The last result we state is an answer to the following question: what can we say about the rate of convergence of $\mathbb{E}\left[|X - \widehat{X}^N|^{2+\beta}\right]$ knowing that $\widehat{X}^N$ is a quadratic optimal quantization? This problem is known as the distortion mismatch problem and has been first addressed in [GLP08] and the results have been extended in Theorem 4.3 of [PS18a].

Theorem 4.1.13. *[L^r-L^s-distortion mismatch] Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ be a random variable and let $r \in (0, +\infty)$. Assume that the distribution $\mathbb{P}_X$ of X has a non-zero absolutely continuous component with density φ, i.e. $\mathbb{P}_X(d\xi) = \varphi(\xi) \cdot \lambda(d\xi) + \nu(d\xi)$, where $\nu \perp \lambda$ is the singular component of $\mathbb{P}_X$ with respect to the Lebesgue measure λ on $\mathbb{R}$ and φ is non-identically null.*

Let $(\Gamma_N)_{N\geqslant 1}$ be a sequence of L^r-optimal grids. Let $s \in (r, r+1)$. If

$$X \in L^{\frac{s}{1+r-s}+\delta}(\mathbb{P})$$

for some $\delta > 0$, then

$$\limsup_N N\|X - \widehat{X}^N\|_s < +\infty.$$

4.2 Weak Error bounds for Optimal Quantization ($d = 1$)

Let $X \in L^2(\mathbb{P})$ and $\widehat{X}^N$ a quadratic optimal quantizer of X which takes its values in the finite grid $\Gamma_N = \{x_1^N, \dots, x_N^N\}$ of size N. We consider a function $f : \mathbb{R} \to \mathbb{R}$ with $f(X) \in L^2(\mathbb{P})$. One of the application of the framework developed above is the approximation of expectations of the form $\mathbb{E}\big[f(X)\big]$. Indeed, as $\widehat{X}^N$ is close to X in $L^2(\mathbb{P})$, a natural idea is to replace X by $\widehat{X}^N$ inside the expectation

$$\mathbb{E}\big[f(\widehat{X}^N)\big] = \sum_{i=1}^N f(x_i^N)\,\mathbb{P}_X\big(C_i(\Gamma_N)\big).$$

The above formula is referred as the quantization-based cubature formula to approximate $\mathbb{E}\big[f(X)\big]$. Now, we need to have an idea of the error we make when doing such an approximation and what is its rate of convergence as N tends to infinity? For that, we want to find the largest $\alpha \in \mathbb{R}$, such that the beyond limit is bounded

$$\lim_{N\to+\infty} N^\alpha \big|\,\mathbb{E}\big[f(X)\big] - \mathbb{E}\big[f(\widehat{X}^N)\big]\big| \leqslant C_{f,X} < +\infty. \tag{4.3}$$

The first class of function we consider is the class of Lipschitz continuous functions, more precisely piecewise affine functions and convex Lipschitz continuous functions. Then we deal with differentiable functions with piecewise-defined derivatives.

4.2.1 Piecewise affine functions

We improve the standard rate of convergence which is of order 1 for Lipschitz continuous functions by considering a subclass of the Lipschitz continuous functions, namely piecewise affine functions. This new result shows that the weak error induced is of order 2 ($\alpha = 2$ in (4.3)).

Lemma 4.2.1. *Assume that the distribution $\mathbb{P}_X = \varphi\cdot\lambda$ of X satisfies the conditions of Theorem 4.1.10. Let $f : \mathbb{R} \to \mathbb{R}$ be a Borel function.*

(a) If f is a continuous piecewise affine function with finitely many breaks of affinity, then there exists a real constant $C_{f,X} > 0$ such that

$$\limsup_N N^2 \big|\mathbb{E}\big[f(X)\big] - \mathbb{E}\big[f(\widehat{X}^N)\big]\big| \leqslant C_{f,X} < +\infty.$$

(b) However, if f is not supposed continuous but is still a piecewise affine function with finitely many breaks of affinity, then there exists a real constant $C_{f,X} > 0$ such that

$$\limsup_N N \big|\mathbb{E}\big[f(X)\big] - \mathbb{E}\big[f(\widehat{X}^N)\big]\big| \leqslant C_{f,X} < +\infty.$$

Proof. Let I be a compact interval containing all the affinity breaks of f denoted $a_1, \ldots, a_\ell$.

(a) Let f supposed to be continuous. Note that f is Lipschitz continuous (with coefficient denoted $[f]_{Lip} := \max_{i=1,\ldots,\ell} |a_i|$). Let $\Gamma_N = \{x_1^N, \ldots, x_N^N\}$ be an L^2- optimal quantizer at level $N \geqslant 1$.

$$\begin{aligned}\mathbb{E}\big[f(X)\big] - \mathbb{E}\big[f(\widehat{X}^N)\big] &= \sum_{i=1}^N \int_{C_i(\Gamma_N)} \big(f(\xi) - f(x_i^N)\big)\, \mathbb{P}_X(d\xi) \\ &= \sum_{i \in J_f^N} \int_{C_i(\Gamma_N)} \big(f(\xi) - f(x_i^N)\big)\, \mathbb{P}_X(d\xi) \qquad (4.4)\end{aligned}$$

where $J_f^N = \{i : C_i(\Gamma_N) \text{ contains an affinity break}\}$ since all other terms are 0. Indeed, as $f(\xi) = \alpha_i \xi + \beta_i$ on $C_i(\Gamma_N)$ and using Corollary 4.1.7

$$\int_{C_i(\Gamma_N)} \big(f(\xi) - f(x_i^N)\big)\, \mathbb{P}_X(d\xi) = \alpha_i\, \mathbb{E}\big[(X - \widehat{X}^N)\, \mathbb{1}_{\{\widehat{X}^N = x_i^N\}}\big] = 0.$$

Now, taking the absolute value in (4.4), we have

$$\begin{aligned}\big|\mathbb{E}\big[f(X)\big] - \mathbb{E}\big[f(\widehat{X}^N)\big]\big| &\leqslant \operatorname{card}(J_f^N) \max_{i \in J_f^N} \int_{C_i(\Gamma_N)} |f(\xi) - f(x_i^N)|\, \mathbb{P}_X(d\xi) \\ &\leqslant \operatorname{card}(J_f^N)[f]_{Lip} \max_{i \in J_f^N} \int_{C_i(\Gamma_N)} |\xi - x_i^N|\, \mathbb{P}_X(d\xi) \qquad (4.5)\end{aligned}$$

and using Corollary 4.1.11 with $s = 1$, we have the desired result, with an explicit asymptotic upper bound,

$$\begin{aligned}\limsup_N N^2 \big|\mathbb{E}\big[f(X)\big] - \mathbb{E}\big[f(\widehat{X}^N)\big]\big| &\leqslant [f]_{Lip} \lim_N \operatorname{card}(J_f^N) \max_{i \in J_f^N} N^2 \int_{C_i(\Gamma_N)} |\xi - x_i^N|\, \mathbb{P}_X(d\xi) \\ &< [f]_{Lip} \frac{\ell}{2\sqrt{3}} \big(c_{\varphi,1/3} \|\varphi\|_{1/3} \|\varphi^{1/3}\|_\infty\big)^{\frac{1}{2}} \\ &< +\infty.\end{aligned}$$

(b) The sum in (4.4) in the discontinuous case is still true. However, the bound in (4.5) changes and becomes

$$\left|\mathbb{E}\left[f(X)\right]-\mathbb{E}\left[f(\widehat{X}^N)\right]\right|\leqslant 2\ell\|f\|_{\infty,K_0}\max_{i\in J_f^N}\mathbb{P}_X\left(C_i(\Gamma_N)\right)$$

where $\|f\|_{\infty,K_0}$ denotes the maximum of $|f|$ on K_0 and K_0 is defined as the compact appearing in Lemma 4.1.9 stating that the union over all N of all the cells where their intersection with the interval $[a_1,a_\ell]$ is non empty lies in a compact K_0, namely

$$\bigcup_N\bigcup_{C_i(\Gamma_N)\cap[a_1,a_\ell]\neq\varnothing}C_i(\Gamma_N)\subset K_0.$$

The desired limit is obtained using Theorem 4.1.10.

□

4.2.2 Lipschitz Convex functions

Thanks to the previous result on piecewise-affine functions, we can extend the rate of convergence of order 2 to a bigger class of functions: Lipschitz convex functions.

We recall that a real-valued function f defined on a non-trivial interval $I\subset\mathbb{R}$ is convex if

$$f\big(tx+(1-t)y\big)\leqslant tf(x)+(1-t)f(y),$$

for every $t\in[0,1]$ and $x,y\in I$. If $f:I\to\mathbb{R}$ is supposed to be a convex function, then its right and left derivatives exist, are non-decreasing on $\mathring{I}$ and $\forall x\in\mathring{I}$, $f'_-(x)\leqslant f'_+(x)$. Moreover, as f is supposed to be Lipschitz continuous, then f'_- and f'_+ are bounded on I by $[f]_{Lip}$.

Remark. One of the very interesting properties of convex functions when dealing with stationary quantizers follows from Jensen's inequality. Indeed, for every convex function $f:I\to\mathbb{R}$ such that $f(X)\in L^1(\mathbb{P})$,

$$\mathbb{E}\left[f\big(\mathbb{E}\left[X\mid\widehat{X}^N\right]\big)\right]\leqslant\mathbb{E}\left[\mathbb{E}\left[f(X)\mid\widehat{X}^N\right]\right]$$

so that,

$$\mathbb{E}\left[f(\widehat{X}^N)\right]\leqslant\mathbb{E}\left[f(X)\right].$$

This means that the quantization-based cubature formula used to approximate $\mathbb{E}\left[f(X)\right]$ is a lower-bound of the expectation.

We present, here, a more convenient and general form of the well known Carr-Madan formula representation (see [CM01]).

Proposition 4.2.2. *Let $f:I\to\mathbb{R}$ be a Lipschitz convex function and let I be any interval non trivial ($\neq\varnothing,\{a\}$) with endpoints $a,b\in\overline{\mathbb{R}}$. Then, there exists a unique finite non-negative*

Borel measure $\nu := \nu_f$ on I such that, for every $c \in I$,

$$\forall x \in I, \quad f(x) = f(c) + (x-c)f'_+(c) + \int_{[a,c]\cap I} (u-x)_+ \nu(du) + \int_{(c,b]\cap I} (x-u)_+ \nu(du).$$

Proof. Let $f : I \to \mathbb{R}$ be a Lipschitz convex function. We can define the non-negative finite measure $\nu := \nu_f$ on I by setting

$$\forall x, y \in I, \quad x \leqslant y, \quad \nu\big((x,y]\big) = f'_+(y) - f'_+(x).$$

The finiteness of ν is induced by the Lipschitz continuity of f as the left and right derivatives are bounded by $[f]_{Lip} = \max(\|f'_+\|_\infty, \|f'_-\|_\infty)$. Let $c \in I$, for every $x \geqslant c$, we have the following representation of $f(x)$:

$$\begin{aligned}
f(x) &= f(c) + \int_c^x f'_+(u)du \\
&= f(c) + xf'_+(c) + \int_c^x \nu((c,u])du \\
&= f(c) + xf'_+(c) + \int\int \mathbb{1}_{(c,x]}(u)\, \mathbb{1}_{(c,u]}(v)\nu(dv)\ du \\
&= f(c) + xf'_+(c) + \int_{(c,x]} (x-v)du\ \nu(dv) \\
&= f(c) + xf'_+(c) + \int_{(c,b]\cap I} (x-v)_+ \nu(dv)
\end{aligned}$$

using Fubini's Theorem and noting that $\mathbb{1}_{(c,x]}(u)\,\mathbb{1}_{(c,u]}(v) = \mathbb{1}_{(c,x]}(v)\,\mathbb{1}_{[v,x]}(u)$. Similarly for $x \leqslant c$

$$f(x) = f(c) + xf'_+(c) + \int_{[a,c]\cap I} (u-x)_+ \nu(du).$$

Then,

$$\forall x \in \mathbb{R}, \quad f(x) = f(c) + xf'_+(c) + \int_{[a,c]\cap I} (u-x)_+ \nu(du) + \int_{(c,b]\cap I} (x-u)_+ \nu(du).$$

□

We can now use the representation of convex functions given above and extend the result concerning the weak error of order 2 ($\alpha = 2$ in (4.3)).

Proposition 4.2.3. *We assume that the distribution $\mathbb{P}_X = \varphi \cdot \lambda$ of X satisfies the conditions of Theorem 4.1.10. Let I be any non-trivial interval and let $f : I \to \mathbb{R}$ be a Lipschitz convex function with second derivative ν (see Proposition 4.2.2). If $I_{\mathbb{P}_X} \cap \mathrm{supp}(\nu)$ is compact, with*

$I_{\mathbb{P}_X} := \mathrm{supp}(\mathbb{P}_X)$, *then there exists a real constant* $C_{f,X} > 0$ *such that*

$$\limsup_N N^2 \big| \mathbb{E}\left[f(X)\right] - \mathbb{E}\left[f(\widehat{X}^N)\right] \big| \leqslant C_{f,X} < +\infty.$$

Remark. Assuming that $\mathrm{supp}(\nu)$ is compact actually means that f is affine outside a compact set, namely that there exist $\alpha^{(\pm)}$ and $\beta^{(\pm)}$ such that $f(x) = \alpha^{(+)}x + \beta^{(+)}$, for x large enough ($x \geqslant K_+$) and $f(x) = \alpha^{(-)}x + \beta^{(-)}$, for x small enough ($x \leqslant K_-$). Therefore, this class of functions contains all classical vanilla financial payoffs: call, put, butterfly, saddle, straddle, spread, etc. Moreover, if $I_{\mathbb{P}_X}$ is compact, such as in the uniform distribution, then there is no need for the hypothesis on ν and we could consider any Lipschitz convex functions we want. The hypothesis on the intersection allows us to consider more cases.

Proof. First we decompose the expectations across the Voronoï cells as follows

$$\begin{aligned}
\mathbb{E}\left[f(X) - f(\widehat{X}^N)\right] &= \sum_{i=1}^N \mathbb{E}\left[\left(f(X) - f(\widehat{X}^N)\right) \mathbb{1}_{\{X \in C_i(\Gamma_N)\}}\right] \\
&= \sum_{i=1}^N \mathbb{E}\left[\left(f(X) - f(x_i^N)\right) \mathbb{1}_{\{X \in (x_{i-1/2}^N, x_{i+1/2}^N]\}}\right].
\end{aligned}$$

We use the integral representation of the convex function f, of the Proposition 4.2.2, with $x := X$ and $c := x_i$ and with the stationarity conditional property given by Corollary 4.1.7, the first term cancels out, for every i,

$$\mathbb{E}\left[(X - x_i^N) f'_+(x_i^N)\, \mathbb{1}_{\{X \in C_i(\Gamma_N)\}}\right] = 0.$$

Hence, we obtain

$$\begin{aligned}
&\mathbb{E}\Big[\left(f(X) - f(x_i^N)\right) \mathbb{1}_{\{X \in (x_{i-1/2}^N, x_{i+1/2}^N]\}}\Big] \\
&= \mathbb{E}\left[\left(\int_{[a, x_i^N] \cap I} (u - X)_+ \nu(du) + \int_{(x_i^N, b] \cap I} (X - u)_+ \nu(du)\right) \mathbb{1}_{\{X \in (x_{i-1/2}^N, x_{i+1/2}^N]\}}\right] \\
&= \mathbb{E}\left[\int_{(x_{i-1/2}^N, x_i^N]} (u - X)_+ \nu(du)\, \mathbb{1}_{\{X \in (x_{i-1/2}^N, x_i^N]\}}\right] \qquad (4.6) \\
&\qquad + \mathbb{E}\left[\int_{(x_i^N, x_{i+1/2}^N)} (X - u)_+ \nu(du)\, \mathbb{1}_{\{X \in [x_i^N, x_{i+1/2}^N]\}}\right].
\end{aligned}$$

The interval $(x_{i-1/2}^N, x_i^N]$ in the integral is left-open because when $u = x_{i-1/2}^N$, as $X \in (x_{i-1/2}^N, x_i^N]$, $(u - X)_+ = 0$. The same remark can be made concerning the right open-bound of the interval

$(x_i^N, x_{i+1/2}^N)$ in the integral. Now, using a crude upper-bound for (4.6), we get

$$\begin{aligned}\mathbb{E}\left[\left(f(X)-f(x_i^N)\right)\mathbb{1}_{\{X\in(x_{i-1/2}^N,x_{i+1/2}^N]\}}\right] &\leqslant \mathbb{E}\left[(x_i^N-X)\nu\big((x_{i-1/2}^N,x_i^N]\big)\,\mathbb{1}_{\{X\in(x_{i-1/2}^N,x_i^N]\}}\right]\\ &\quad+\mathbb{E}\left[(X-x_i^N)\nu\big((x_i^N,x_{i+1/2}^N)\big)\,\mathbb{1}_{\{X\in[x_i^N,x_{i+1/2}^N]\}}\right]\\ &\leqslant \mathbb{E}\left[|x_i^N-X|\,\mathbb{1}_{\{X\in C_i(\Gamma_N)\}}\right]\nu\big(C_i(\Gamma_N)\big)\end{aligned}$$

as $\nu\big((x_{i-1/2}^N, x_{i+1/2}^N)\big) \leqslant \nu\big(C_i(\Gamma_N)\big)$. Hence

$$\begin{aligned}0\leqslant \mathbb{E}\left[f(X)-f(\widehat{X}^N)\right] &\leqslant \sum_{i=1}^N \mathbb{E}\left[|x_i^N-X|\,\mathbb{1}_{\{X\in C_i(\Gamma_N)\}}\right]\nu\big(C_i(\Gamma_N)\big)\\ &\leqslant \sum_{i=1}^N \mathbb{E}\left[|x_i^N-X|\,\mathbb{1}_{\{X\in C_i(\Gamma_N)\}}\right]\mathbb{1}_{\{x_i^N\in J_\nu\}}\,\nu\big(C_i(\Gamma_N)\big)\end{aligned}$$

with $J_\nu := [\inf_N x_{i_a-1/2}^N, \sup_N x_{i_b+1/2}^N]$ where $x_{i_a}^N$ and $x_{i_b}^N$ are the centroids of the optimal quantizer of size N that contains, respectively, the infimum and the supremum of the support of ν, denoted by a and b, respectively. Hence, $x_{i_a-1/2}^N$ is the lower bound of the Voronoï cell $C_{i_a}(\Gamma_N)$ associated to the centroid $x_{i_a}^N$ and $x_{i_b+1/2}^N$ is the upper bound of the Voronoï cell $C_{i_b}(\Gamma_N)$ associated to the centroid $x_{i_b}^N$. If a is not contained in $I_{\mathbb{P}_X}$, then the lower bound of J_ν is set to a, and the same hold for b: if it is not contained in $I_{\mathbb{P}_X}$, the upper bound of J_ν is set to b. Then,

$$\begin{aligned}N^2\,\mathbb{E}\left[f(X)-f(\widehat{X}^N)\right] &\leqslant N^2\sum_{i=1}^N \mathbb{E}\left[|x_i^N-X|\,\mathbb{1}_{\{X\in C_i(\Gamma_N)\}}\right]\mathbb{1}_{\{x_i^N\in J_\nu\}}\,\nu(C_i(\Gamma_N))\\ &\leqslant N^2\sup_{i:x_i^N\in I_{\mathbb{P}_X}\cap J_\nu}\mathbb{E}\left[|\widehat{X}^N-X|\,\mathbb{1}_{\{X\in C_i(\Gamma_N)\}}\right]\sum_{i=1}^N\nu(C_i(\Gamma_N))\\ &\leqslant \nu(I_{\mathbb{P}_X})N^2\sup_{i:x_i^N\in I_{\mathbb{P}_X}\cap J_\nu}\mathbb{E}\left[|\widehat{X}^N-X|\,\mathbb{1}_{\{X\in C_i(\Gamma_N)\}}\right]\end{aligned}$$

yielding the desired result with Theorem 4.1.10 if $I_{\mathbb{P}_X} \cap J_\nu$ is compact.
Under the hypothesis $I_{\mathbb{P}_X} \cap \ \mathrm{supp}(\nu)$ compact, then by Lemma 4.1.9,

$$\bigcup_N \bigcup_{x_i^N\in I_{\mathbb{P}_X}\cap\ \mathrm{supp}(\nu)} C_i(\Gamma_N) \subset \bigcup_N \bigcup_{C_i(\Gamma_N)\cap I_{\mathbb{P}_X}\cap\ \mathrm{supp}(\nu)\neq\varnothing} C_i(\Gamma_N) \subset K_0,$$

with $K_0 := I_{\mathbb{P}_X} \cap J_\nu$ compact, which is what we were looking for. □

Proposition 4.2.4. *Assume that the distribution $\mathbb{P}_X = \varphi\cdot\lambda$ of X satisfies the conditions of Theorem 4.1.10 not only on compact sets but uniformly. Let I be any non-trivial interval then for every function $f : I \to \mathbb{R}$ Lipschitz convex with second derivative ν defined as in Proposition*

4.2.2, there exists a real constant $C_{f,X} > 0$ such that

$$\limsup_N N^2 \big| \mathbb{E}\big[f(X)\big] - \mathbb{E}\big[f(\widehat{X}^N)\big] \big| \leqslant C_{f,X} < +\infty.$$

Proof. This proof is exactly the same as above the Proposition. □

Remark. It has not be shown yet that Gaussian or Exponential random variables satisfy the conditions of Theorem 4.1.10 uniformly but empirical tests tend to confirm that they exhibit the error bound property for Lipschitz convex functions. More details are given in the numerical part.

4.2.3 Differentiable functions

In the following proposition, we deal with functions that are piecewise-defined and where their piecewise-defined derivatives are supposed to be locally-Lipschitz continuous or locally α-Hölder continuous on the non-bounded parts of the interval. We define below what we mean by locally-Lipschitz and locally α-Hölder.

Definition 4.2.5.

- A function $f : I \to \mathbb{R}$ is supposed to be locally-Lipschitz continuous, if

$$\forall x, y \in I \quad |f(x) - f(y)| \leqslant [f]_{Lip,loc} |x - y| \big(g(x) + g(y)\big)$$

  where $[f]_{Lip,loc}$ is a real constant and $g : \mathbb{R} \to \mathbb{R}_+$.

- A function $f : I \to \mathbb{R}$ is supposed to be locally α-Hölder continuous, if

$$\forall x, y \in I \quad |f(x) - f(y)| \leqslant [f]_{\alpha,loc} |x - y|^\alpha \big(g(x) + g(y)\big)$$

  where $[f]_{\alpha,loc}$ is a real constant and $g : \mathbb{R} \to \mathbb{R}_+$.

Proposition 4.2.6. *Assume that the distribution $\mathbb{P}_X$ of X satisfies the conditions of the L^r-L^s-distortion mismatch Theorem 4.1.13 and Theorem 4.1.10 concerning the local behaviours of optimal quantizers. If $f : \mathbb{R} \to \mathbb{R}$ is a piecewise-defined continuous function with finitely many breaks of affinity $\{a_1, \ldots, a_K\}$, where $-\infty = a_0 < a_1 < \cdots < a_K < a_{K+1} = +\infty$, such that the piecewise-defined derivatives denoted $(f'_k)_{k=0,\ldots,d}$ are either*

(a) locally-Lipschitz continuous on (a_k, a_{k+1}) where $\exists\, q_k > 3$ such that the q_k-th power of $g_k : (a_k, a_{k+1}) \to \mathbb{R}_+$ defined in Definition 4.2.5 are convex and $\big(\|g_k(X)\|_{q_k}\big)_{k=1,\ldots,K} < +\infty$. Then there exists a real constant $C_{f,X} > 0$ such that

$$\limsup_N N^2 \big| \mathbb{E}\big[f(X)\big] - \mathbb{E}\big[f(\widehat{X}^N)\big] \big| \leqslant C_{f,X} < +\infty.$$

(b) or locally α-Hölder continuous on (a_k, a_{k+1}), $\alpha \in (0,1]$, where $\exists\, q_k > \frac{3}{2-\alpha}$ such that the q_k-th power of $g_k : (a_k, a_{k+1}) \to \mathbb{R}_+$ defined in Definition 4.2.5 are convex and $\big(\|g_k(X)\|_{q_k}\big)_{k=1,\dots,K} < +\infty$. Then there exists a real constant $C_{f,X} > 0$ such that

$$\limsup_N N^{1+\alpha} \big|\, \mathbb{E}\,[f(X)] - \mathbb{E}\,[f(\widehat{X}^N)]\big| \leqslant C_{f,X} < +\infty.$$

Proof. (a) Let $\Gamma_N = \{x_1^N, \dots, x_N^N\}$ be a L^2- optimal quantizer at level $N \geqslant 1$. In the first place, we define the set of all the indexes of the Voronoï cells that contains a break of affinity

$$I_{reg}^N = \big\{i = 1, \dots, N : C_i(\Gamma_N) \cap [a_1, a_K] \neq \varnothing\big\}.$$

Hence,

$$\mathbb{E}\,[f(\widehat{X}^N)] - \mathbb{E}\,[f(X)] = \underbrace{\sum_{i \in I_{reg}^N} \int_{C_i(\Gamma_N)} \big(f(x_i^N) - f(\xi)\big)\, \mathbb{P}_X(d\xi)}_{(A)} + \underbrace{\sum_{i \notin I_{reg}^N} \int_{C_i(\Gamma_N)} \big(f(x_i^N) - f(\xi)\big)\, \mathbb{P}_X(d\xi)}_{(B)}.$$

First, we deal with the (B) term. As, $i \notin I_{reg}^N$, f is differentiable in $C_i(\Gamma_N)$ and admits a first-order Taylor expansion at the point x_i^N, moreover by Corollary 4.1.7, $\int_{C_i(\Gamma_N)} f'(x_i^N)(\xi - x_i^N)\, \mathbb{P}_X(d\xi) = 0$, hence

$$\int_{C_i(\Gamma_N)} \big(f(x_i^N) - f(\xi)\big)\, \mathbb{P}_X(d\xi) = \int_{C_i(\Gamma_N)} \int_0^1 \big(f'(x_i^N) - f'(tx_i^N + (1-t)\xi)\big)(x_i^N - \xi) dt\, \mathbb{P}_X(d\xi).$$

Now, we take the absolute value and we use the locally Lipschitz property of the derivative, yielding

$$\begin{aligned}
&\left| \int_{C_i(\Gamma_N)} \big(f(x_i^N) - f(\xi)\big)\, \mathbb{P}_X(d\xi) \right| \\
&\qquad \leqslant \int_{C_i(\Gamma_N)} \int_0^1 |f'(x_i^N) - f'(tx_i^N + (1-t)\xi)||x_i^N - \xi| dt\, \mathbb{P}_X(d\xi) \\
&\qquad \leqslant [f']_{k,Lip,loc} \int_{C_i(\Gamma_N)} \int_0^1 (1-t)|x_i^N - \xi|^2 \big(g_{k_i}(x_i^N) + g_{k_i}(tx_i^N + (1-t)\xi)\big) dt\, \mathbb{P}_X(d\xi),
\end{aligned} \tag{4.7}$$

with $k_i := \{k = 0, \dots, d : x_i \in (a_k, a_{k+1})\}$. Under the convex hypothesis of $g_{k_i}^{q_{k_i}}$, we have that

$$g_{k_i}(tx_i^N + (1-t)\xi) \leqslant \max\big(g_{k_i}(x_i^N), g_{k_i}(\xi)\big) \leqslant g_{k_i}(x_i^N) + g_{k_i}(\xi),$$

thus

$$\int_{C_i(\Gamma_N)}\int_0^1 (1-t)|x_i^N-\xi|^2\big(g_k(x_i^N)+g_k(tx_i^N+(1-t)\xi)\big)dt\,\mathbb{P}_X(d\xi)$$
$$\leqslant \frac{1}{2}\int_{C_i(\Gamma_N)}|x_i^N-\xi|^2\big(2g_k(x_i^N)+g_k(\xi)\big)\,\mathbb{P}_X(d\xi).$$

Now, taking the sum over all $i\notin I_{reg}^N$ and denoting $[f']_{Lip,loc}:=\max_k[f']_{k,Lip,loc}$

$$\begin{aligned}
\big|(B)\big| &\leqslant \frac{1}{2}[f']_{Lip,loc}\sum_{i\notin I_{reg}^N}\int_{C_i(\Gamma_N)}|x_i^N-\xi|^2\big(2g_{k_i}(x_i^N)+g_{k_i}(\xi)\big)\,\mathbb{P}_X(d\xi)\\
&\leqslant \frac{K}{2}[f']_{Lip,loc}\max_k \mathbb{E}\Big[|\widehat{X}^N-X|^2\big(2g_k(\widehat{X}^N)+g_k(X)\big)\Big]\\
&\leqslant \frac{K}{2}[f']_{Lip,loc}\max_k \|\widehat{X}^N-X\|_{2p_k}^2\big(2\|g_k(\widehat{X}^N)\|_{q_k}+\|g_k(X)\|_{q_k}\big)\\
&\leqslant \frac{K}{2}[f']_{Lip,loc}\|\widehat{X}^N-X\|_{2p}^2\max_k\big(2\|g_k(\widehat{X}^N)\|_{q_k}+\|g_k(X)\|_{q_k}\big)\\
&\leqslant \frac{3K}{2}[f']_{Lip,loc}\|\widehat{X}^N-X\|_{2p}^2\max_k\|g_k(X)\|_{q_k}
\end{aligned}\tag{4.8}$$

using Hölder inequality, such that $\frac{1}{p_k}+\frac{1}{q_k}<1$ and the convexity of g^{q_k}. Under the hypothesis $q_k>3$, p_k has to be in contained in the interval $(1,3/2)$, hence p is defined as $p:=\max_k p_k$ and using the non-decreasing property of the L^p norm, we obtain the fourth inequality in (4.8). Now, if we use the L^r-L^s-distortion mismatch Theorem 4.1.13 with $r=2$ and $s=2p<3$ under the condition $X\in L^{\frac{2p}{3-2p}+\delta}(\mathbb{P})$, we have

$$\begin{aligned}
N^2\big|(B)\big| &\leqslant N^2\frac{3K}{2}[f']_{Lip,loc}\|\widehat{X}^N-X\|_{2p}^2\max_k\|g_k(X)\|_{q_k}\\
&\xrightarrow{N\to+\infty} C_2<+\infty.
\end{aligned}\tag{4.9}$$

Secondly, we take care of the (A) term. Using Lemma 4.1.9 stating that the union over all N of all the cells where their intersection with the interval $[a_1,a_K]$ is non empty lies in a compact K_0, namely

$$\bigcup_N\bigcup_{C_i(\Gamma_N)\cap[a_1,a_K]\neq\varnothing} C_i(\Gamma_N)\subset K_0$$

and using that f' is bounded on K_0 by $[f']_{Lip,K_0}$, we can use the following integral representation of f

$$f(x)=\int_0^x f'(u)du+f(0)$$

and the stationarity property of the optimal quantizer on $C_i(\Gamma_N)$, yielding

$$\left|\int_{C_i(\Gamma_N)} \big(f(x_i^N) - f(\xi)\big)\,\mathbb{P}_X(d\xi)\right| = \left|\int_{C_i(\Gamma_N)} \int_{\xi}^{x_i^N} f'(u)du\,\mathbb{P}_X(d\xi)\right|$$
$$\leqslant [f']_{Lip,K_0} \int_{C_i(\Gamma_N)} |\xi - x_i^N|\,\mathbb{P}_X(d\xi).$$

Now, we sum among all $i \in I_{reg}^N$

$$|(A)| \leqslant [f']_{Lip,K_0} \sum_{i\in I_{reg}^N} \int_{C_i(\Gamma_N)} |\xi - x_i^N|\,\mathbb{P}_X(d\xi).$$

Hence, using the result concerning the local behaviour of optimal quantizers Corollary 4.1.11 as $[a_1, a_K]$ is compact, we have

$$\begin{aligned} N^2|(A)| &\leqslant N^2[f']_{Lip,K_0} \sum_{i\in I_{reg}^N} \int_{C_i(\Gamma_N)} |\xi - x_i^N|\,\mathbb{P}_X(d\xi) \\ &\leqslant N^2K[f']_{Lip,K_0} \sup_{i:x_i^N\in K_0} \int_{C_i(\Gamma_N)} |\xi - x_i^N|\,\mathbb{P}_X(d\xi) \\ &\xrightarrow{N\to+\infty} C_1 < +\infty. \end{aligned} \tag{4.10}$$

Finally, using (4.10) and (4.9), we have the desired result

$$N^2\big|\,\mathbb{E}\,[f(X)] - \mathbb{E}\,[f(\widehat{X}^N)]\big| \leqslant N^2\Big(|(A)| + |(B)|\Big) \xrightarrow{N\to+\infty} C_1 + C_2 < +\infty.$$

(b) When the piecewise-defined derivatives are locally α-Hölder continuous on $(-\infty, a_1]$ and $[a_K, +\infty)$, $\alpha \in (0,1)$, the proof is very close to the locally Lipschitz case. Indeed, the first difference is in (4.7), where the $|x_i^N - \xi|^2$ is replaced by $|x_i^N - \xi|^{1+\alpha}$ and the constant is the one of the locally α-Hölder hypothesis. This implies that (4.8) is replaced by

$$|(B)| \leqslant \frac{3K[f']_{Hol,loc}}{2} \|\widehat{X}^N - X\|_{(1+\alpha)p}^{1+\alpha} \max_k \|g_k(X)\|_{q_k}.$$

Finally, using the L^r-L^s-distortion mismatch Theorem 4.1.13 with $r = 2$ and $s = (1+\alpha)p < 3$ under the condition $X \in L^{\frac{(1+\alpha)p}{3-(1+\alpha)p}+\delta}(\mathbb{P})$, we have

$$\begin{aligned} N^{1+\alpha}|(B)| &\leqslant N^{1+\alpha}\frac{3K[f']_{Hol,loc}}{2} \|\widehat{X}^N - X\|_{(1+\alpha)p}^{1+\alpha} \max_k \|g_k(X)\|_{q_k} \\ &\xrightarrow{N\to+\infty} C_3 < +\infty. \end{aligned}$$

The other parts of the proof are identical, yielding the desired result.

□

Remark. If one strengthens the hypothesis concerning the piecewise locally Lipschitz continuous derivative and considers in place that the derivative is piecewise Lipschitz continuous, then the hypothesis that X should satisfy the conditions of Theorem 4.1.13 can be relaxed. Indeed, the term $\frac{3K}{2}[f']_{Lip,loc}\|\widehat{X}^N - X\|_{2p}^2 \max_k \|g_k(X)\|_{q_k}$ in (4.8) would become $\frac{1}{2}[f']_{Lip}\|\widehat{X}^N - X\|_2^2$ and we would conclude using Zador's Theorem 4.1.8.

4.3 Weak Error and Richardson-Romberg Extrapolation

One can improve the previous speeds of convergence using Richardson-Romberg extrapolation method. The Richardson extrapolation is a method that was originally introduced in numerical analysis by Richardson in 1911 (see [Ric10]) and developed later by Romberg in 1955 (see [Rom55]) whose aim was to speed-up the rate of convergence of a sequence, to accelerate the research of a solution of an ODE's or to approximate more precisely integrals.

[TT90] and [Pag07; Pag18] used this concept for the computation of the expectation $\mathbb{E}\left[f(X_T)\right]$ of a diffusion $(X_t)_{t\in[0,T]}$ that cannot be simulated exactly at a given time T but can be approximated by a simulable process $\widetilde{X}_T^{(h)}$ using a Euler scheme with time step $h = T/n$ and n the number of time step. The main idea is to use the weak error expansion of the approximation in order to highlight the term we would *kill*. For example, using the following weak time discretization error of order 1

$$\mathbb{E}\left[f(X_T)\right] = \mathbb{E}\left[f(\widetilde{X}_T^{(h)})\right] + \frac{c_1}{n} + O(n^{-2}),$$

one reduces the error of the approximation using a linear combination of the approximating process $\widetilde{X}_T^{(h)}$ and a refiner process $\widetilde{X}_T^{(h/2)}$, namely

$$\mathbb{E}\left[f(X_T)\right] = \mathbb{E}\left[2f(\widetilde{X}_T^{(h/2)}) - f(\widetilde{X}_T^{(h)})\right] - \frac{1}{2}\frac{c_2}{n^2} + O(n^{-2}).$$

Our goal within the optimal quantization framework is to improve the speed of convergence of the cubature formula using the same ideas. Let us consider a random variable $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ and a quadratic-optimal quantizer $\widehat{X}^N$ of X. In our case we show that, if we are in dimension one there exists, for some functions f, a *weak error expansion* of the form:

$$\mathbb{E}\left[f(X)\right] = \mathbb{E}\left[f(\widehat{X}^N)\right] + \frac{c_2}{N^2} + O(N^{-(2+\beta)})$$

with $\beta \in (0, 1)$. We present in Section 4.3.2 a similar result in higher dimension.

4.3.1 In dimension one

This first result is focused on function $f : \mathbb{R} \to \mathbb{R}$ with Lipschitz continuous second derivative. In that case, we have a *weak error quantization* of order two. The first term of the expansion is equal to zero, thanks to the stationarity of the quadratic optimal quantizer.

Proposition 4.3.1. *Let $f : \mathbb{R} \to \mathbb{R}$ be a twice differentiable function with Lipschitz continuous second derivative. Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ be a random variable and the distribution of $\mathbb{P}_X$ of X has a non-zero absolutely continuous density φ and, for every $N \geqslant 1$, let Γ_N be an optimal quantizer at level $N \geqslant 1$ for X. Then, $\forall\, \beta \in (0,1)$, we have the following expansion*

$$\mathbb{E}\big[f(X)\big] = \mathbb{E}\big[f(\widehat{X}^N)\big] + \frac{c_2}{N^2} + O(N^{-(2+\beta)}).$$

Moreover, if $\varphi : [a,b] \to \mathbb{R}_+$ is a Lipschitz continuous probability density function, bounded away from 0 on $[a,b]$ then we can choose $\beta = 1$, yielding

$$\mathbb{E}\big[f(X)\big] = \mathbb{E}\big[f(\widehat{X}^N)\big] + \frac{c_2}{N^2} + O(N^{-3}).$$

Proof. If f is twice differentiable with Lipschitz continuous second derivatives, we have the following expansion

$$f(x) = f(y) + f'(y)(x-y) + \frac{1}{2}f''(y)(x-y)^2 + \int_0^1 (1-t)\big(f''(tx+(1-t)y) - f''(y)\big)(x-y)^2 dt$$

hence replacing x and y by X and $\widehat{X}^N$ respectively and taking the expectation yields

$$\mathbb{E}\big[f(X)\big] = \mathbb{E}\big[f(\widehat{X}^N)\big] + \frac{1}{2}\,\mathbb{E}\big[f''(\widehat{X}^N)|X - \widehat{X}^N|^2\big] + R(X, \widehat{X}^N)$$

where $R(X,\widehat{X}) = \int_0^1 (1-t)\,\mathbb{E}\big[\big(f''(tX+(1-t)\widehat{X}) - f''(\widehat{X})\big)|X-\widehat{X}|^2\big]dt$.

First, using Theorem 4.1.12 with f'', we have the following limit

$$\lim_{N\to+\infty} N^2\,\mathbb{E}\big[f''(\widehat{X}^N)|X - \widehat{X}^N|^2\big] = \mathcal{Q}_2(\mathbb{P}_X)\int f''(\xi)\,\mathbb{P}_X(d\xi),$$

hence

$$\mathbb{E}\big[f(X)\big] = \mathbb{E}\big[f(\widehat{X}^N)\big] + \frac{c_2}{N^2} + R(X, \widehat{X}^N).$$

Now, we look closely at asymptotic behaviour of $R(X, \widehat{X}^N)$. One notices that, if we consider a Lipschitz continuous function $g : \mathbb{R} \to \mathbb{R}$, for any fixed $\alpha \in (0,1)$,

$$\forall\, x,\, y \in \mathbb{R}, \quad |g(x) - g(y)| \leqslant 2\|g\|_\infty^\alpha [g]_{Lip}^{1-\alpha} |x-y|^{1-\alpha}.$$

In our case, taking $g \equiv f''$, we have

$$\begin{aligned}\mathbb{E}\Big[\big(f''(tX+(1-t)\widehat{X}^N)-f''(\widehat{X}^N)\big)|X-\widehat{X}^N|^2\Big] \\ &\leqslant \mathbb{E}\Big[2\|f''\|_\infty^\alpha [f'']_{Lip}^{1-\alpha} t^{1-\alpha}|X-\widehat{X}^N|^{1-\alpha}|X-\widehat{X}^N|^2\Big] \\ &\leqslant C_{\beta,f''} t^\beta \,\mathbb{E}\big[|X-\widehat{X}^N|^{2+\beta}\big]\end{aligned}$$

with $0<\beta<1$ where $\beta = 1-\alpha$, hence

$$R(X,\widehat{X}^N) \leqslant \widetilde{C}_{\beta,f''}\,\mathbb{E}\big[|X-\widehat{X}^N|^{2+\beta}\big],$$

with $\widetilde{C}_{\beta,f''} = C_{\beta,f''}\frac{1}{(2+\beta)(1+\beta)}$. Using now Theorem 4.1.13 with $r=2$ and $s=2+\beta$, we have the desired result: $\mathbb{E}\big[|X-\widehat{X}^N|^{2+\beta}\big] = O(N^{-(2+\beta)})$ and finally

$$\mathbb{E}\big[f(X)\big] = \mathbb{E}\big[f(\widehat{X}^N)\big] + \frac{c_2}{N^2} + O(N^{-(2+\beta)}),$$

for every $\beta \in (0,1)$. If moreover, the density φ of X is Lipschitz continuous, bounded away from 0 on $[a,b]$ then we can take $\beta = 1$.

□

Now, following the Richardson-Romberg idea, we could combine approximations with optimal quantizers $\widehat{X}^N$ of size N and $\widehat{X}^{\widetilde{N}}$ of size $\widetilde{N}$, with $\widetilde{N} > N$ in order to *kill* the residual term, leading

$$\mathbb{E}\big[f(X)\big] = \mathbb{E}\left[\frac{\widetilde{N}^2 f(\widehat{X}^{\widetilde{N}}) - N^2 f(\widehat{X}^N)}{\widetilde{N}^2 - N^2}\right] + O(N^{-(2+\beta)}). \tag{4.11}$$

Remark. For the choice of $\widetilde{N}$, we consider $\widetilde{N} := k \times N$. A natural choice for k could be $k=2$ or $k=\sqrt{2}$ but note that the complexity is proportional to $(k+1)N$. In practice it is therefore preferable to take a small k that does not increase complexity too much. For the numerical example, we choose $\widetilde{N} := k \times N$ with $k = 1.2$, this is arbitrary and probably not optimal, however even with this k, we attain a weak error of order 3.

4.3.2 A first extension in higher dimension

In this part, we give a first result on higher dimension concerning the weak error expansion of $\mathbb{E}\big[f(X)\big]$ when approximated by $\mathbb{E}\big[f(\widehat{X}^N)\big]$. In the next part, we use the following matrix norm: let $M \in \mathbb{R}^{d\times d}$, then $|||M||| := \sup_{u:|u|=1} |u^T M u|$.

Proposition 4.3.2. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a twice differentiable function with a bounded and Lipschtiz Hessian H, namely $\forall x,y \in \mathbb{R}^d$, $|||H(x)-H(y)||| \leqslant [H]_{Lip}|x-y|$. Let $X : (\Omega,\mathcal{A},\mathbb{P}) \to \mathbb{R}^d$ be a random vector with independent components $(X_k)_{k=1,\dots,d}$. For every $(N_k)_{k=1,\dots,d} \geqslant 1$, let $(\widehat{X}_d^{N_d})_{k=1,\dots,d}$ be quadratic optimal quantizers of $(X_k)_{k=1,\dots,d}$ taking values in the grids*

$(\Gamma_{N_k})_{k=1,\dots,d}$ *respectively and we define* $\widehat{X}^N$ *as the product quantizer* X *taking values in the finite grid* $\Gamma_N := \bigotimes_{k=1,\dots,d} \Gamma_{N_d}$ *of size* $N := N_1 \times \cdots \times N_d$. *Then, we have the following expansion*

$$\mathbb{E}\big[f(X)\big] = \mathbb{E}\big[f(\widehat{X}^N)\big] + \sum_{k=1}^{d} \frac{c_k}{N_k^2} + O\bigg(\Big(\min_{k=1:d} N_k\Big)^{-(2+\beta)}\bigg).$$

Proof. If f is twice differentiable, hence we have the following Taylor's expansion

$$\begin{aligned} f(x) = f(a) + \nabla f(a)(x-a) + \frac{1}{2} H(a) \cdot (x-a)^{\otimes 2} \\ + \int_0^1 (1-t)\big(H(tx + (1-t)a) - H(a)\big) \cdot (x-a)^{\otimes 2} dt \end{aligned}$$

where the notation $f(x,a) \cdot (x-a)^{\otimes 2}$ stands for $(x-a)^T f(x,a)(x-a)$. Replacing x and a by X and $\widehat{X}^N$ respectively and taking the expectation

$$\begin{aligned} \mathbb{E}\big[f(X)\big] = \mathbb{E}\big[f(\widehat{X}^N)\big] + \mathbb{E}\big[\nabla f(\widehat{X}^N)(X - \widehat{X}^N)\big] + \frac{1}{2}\mathbb{E}\big[H(\widehat{X}^N) \cdot (X - \widehat{X}^N)^{\otimes 2}\big] \\ + \int_0^1 (1-t)\,\mathbb{E}\Big[\big(H(tX + (1-t)\widehat{X}^N) - H(\widehat{X}^N)\big) \cdot (X - \widehat{X}^N)^{\otimes 2}\Big] dt. \end{aligned}$$

Noticing that, by Corollary 4.1.7,

$$\begin{aligned} \mathbb{E}\big[\nabla f(\widehat{X}^N)(X - \widehat{X}^N)\big] &= \sum_{k=1}^{d} \mathbb{E}\left[\frac{\partial f}{\partial x_k}(\widehat{X}^N)(X_k - \widehat{X}_k^{N_k})\right] \\ &= \sum_{k=1}^{d} \mathbb{E}\left[\mathbb{E}\left[\frac{\partial f}{\partial x_k}(\widehat{X}^N)(X_k - \widehat{X}_k^{N_k}) \mid \widehat{X}_{-k}\right]\right] \\ &= 0. \end{aligned}$$

where $\widehat{X}_{-k}$ denotes $(\widehat{X}_1^{N_1}, \dots, \widehat{X}_{k-1}^{N_{k-1}}, \widehat{X}_{k+1}^{N_{k+1}}, \dots, \widehat{X}_d^{N_d})$. Hence

$$\begin{aligned} \mathbb{E}\big[f(X)\big] = \mathbb{E}\big[f(\widehat{X}^N)\big] + \frac{1}{2}\mathbb{E}\big[H(\widehat{X}^N) \cdot (X - \widehat{X}^N)^{\otimes 2}\big] \\ + \int_0^1 (1-t)\,\mathbb{E}\Big[\big(H(tX + (1-t)\widehat{X}^N) - H(\widehat{X}^N)\big) \cdot (X - \widehat{X}^N)^{\otimes 2}\Big] dt \end{aligned} \tag{4.12}$$

and looking at the second term in (4.12)

$$\begin{aligned}
&\mathbb{E}\big[H(\widehat{X}^N)\cdot(X-\widehat{X}^N)^{\otimes 2}\big] \\
&\quad=\sum_{k=1}^{d}\mathbb{E}\left[\frac{\partial^2 f}{\partial x_k^2}(\widehat{X}^N)|X_k-\widehat{X}_k^{N_k}|^2\right]+2\sum_{k\neq l}\mathbb{E}\left[\frac{\partial^2 f}{\partial x_k\partial x_l}(\widehat{X}^N)(X_k-\widehat{X}_k^{N_k})(X_l-\widehat{X}_l^{N_l})\right] \\
&\quad=\sum_{k=1}^{d}\mathbb{E}\left[\mathbb{E}\left[\frac{\partial^2 f}{\partial x_k^2}(\widehat{X}^N)|X_k-\widehat{X}_k^{N_k}|^2\mid\widehat{X}_{-k}\right]\right] \\
&\qquad\qquad+2\sum_{k\neq l}\mathbb{E}\left[\underbrace{\mathbb{E}\left[\frac{\partial^2 f}{\partial x_k\partial x_l}(\widehat{X}^N)(X_k-\widehat{X}_k^{N_k})\mid X_l\right]}_{=0}(X_l-\widehat{X}_l^{N_l})\right] \\
&\quad=\sum_{k=1}^{d}\mathbb{E}\left[\mathbb{E}\left[\frac{\partial^2 f}{\partial x_k^2}(\widehat{X}^N)|X_k-\widehat{X}_k^{N_k}|^2\mid\widehat{X}_{-k}\right]\right] \\
&\quad=\sum_{k=1}^{d}\mathbb{E}\left[\mathbb{E}\left[\frac{\partial^2 f}{\partial x_k^2}(x_1,\dots,x_{k-1},\widehat{X}_k^{N_k},x_{k+1},\dots,x_d)|X_k-\widehat{X}_k^{N_k}|^2\right]\Big|_{\widehat{X}_{-k}=x_{-k}}\right] \\
&\quad=\sum_{k=1}^{d}\mathbb{E}\left[\mathbb{E}\left[g_{k,x_{-k}}(\widehat{X}_k^{N_k})|X_k-\widehat{X}_k^{N_k}|^2\right]\big|_{\widehat{X}_{-k}=x_{-k}}\right].
\end{aligned}$$

Now, using Theorem 4.1.12, we have the following limits, for each k

$$\lim_{N_k\to+\infty}N_k^2\,\mathbb{E}\left[g_{k,x_{-k}}(\widehat{X}_k^{N_k})|X_k-\widehat{X}_k^{N_k}|^2\right]=\mathcal{Q}_2(\mathbb{P}_{X_k})\int g_{k,x_{-k}}(\xi)\,\mathbb{P}_X(d\xi).$$

Giving us the first part of the desired result

$$\mathbb{E}\left[f(X)\right]=\mathbb{E}\left[f(\widehat{X}^N)\right]+\sum_{k=1}^{d}\frac{c_k}{N_k^2}+\int_0^1(1-t)\,\mathbb{E}\left[\big(H(tX+(1-t)\widehat{X}^N)-H(\widehat{X}^N)\big)\cdot(X-\widehat{X}^N)^{\otimes 2}\right]dt$$

with $c_k:=\frac{1}{2}\mathcal{Q}_2(\mathbb{P}_{X_k})\int\int g_{k,x_{-k}}(x)\,\mathbb{P}_{X_k}(dx)\,\mathbb{P}_{X_{-k}}(dy)$. Now, we take care of the integral part, we proceed using the same methodology as in the one dimensional case, using the hypothesis on the Hessian

$$\mathbb{E}\left[\left|\big(H(tX+(1-t)\widehat{X}^N)-H(\widehat{X}^N)\big)\cdot(X-\widehat{X}^N)^{\otimes 2}\right|\right]\leqslant 2t^{\beta}[H]_{Lip}^{\beta}|||H|||_{\infty}^{1-\beta}\,\mathbb{E}\left[|X-\widehat{X}^N|^{2+\beta}\right]$$

with $\beta\in(0,1)$ and $|||H|||_{\infty}:=\sup_{x\in\mathbb{R}^d}|||H(x)|||$. Hence

$$\begin{aligned}
&\int_0^1(1-t)\,\mathbb{E}\left[\big(H(tX+(1-t)\widehat{X}^N)-H(\widehat{X}^N)\big)\cdot(X-\widehat{X}^N)^{\otimes 2}\right]dt \\
&\qquad\leqslant\frac{1}{(2+\beta)(1+\beta)}C_{H,X}\,\mathbb{E}\left[|X-\widehat{X}^N|^{2+\beta}\right].
\end{aligned}$$

Using now Theorem 4.1.13, let $s = 2 + \beta$, we have the desired result: $\mathbb{E}\left[|X_k - \widehat{X}_k^{N_k}|^{2+\beta}\right] = O(N_k^{-(2+\beta)})$ and finally

$$\mathbb{E}\left[f(X)\right] = \mathbb{E}\left[f(\widehat{X}^N)\right] + \sum_{k=1}^{d} \frac{c_k}{N_k^2} + O\left(\left(\min_{k=1:d} N_k\right)^{-(2+\beta)}\right),$$

for every $\beta \in (0,1)$. If moreover, the densities φ_k of X_k, for all $k = 1, \ldots, k$, are Lipschitz continuous, bounded away from 0 on $[a,b]$ then we can take $\beta = 1$.

□

Remark. Even-though, we could be interested by considering non-independent components $(X_k)_{k=1,\ldots,d}$, the independence hypothesis on the components is necessary in the proof because we proceed component by component. For example the first order term of the expansion would not be null by stationarity if the components are not independent.

4.4 Applications

4.4.1 Quantized Control Variates in Monte Carlo simulations

Let $Z \in L^2(\mathbb{P})$ be a random vector with components $(Z_k)_{k=1,\ldots,d}$, we assume that we have a closed-form for $\mathbb{E}[Z_k]$, $k = 1, \ldots, d$, and $f : \mathbb{R}^d \to \mathbb{R}$ our function of interest. We are interested in the quantity

$$I := \mathbb{E}\left[f(Z)\right]. \tag{4.13}$$

The standard method for approximating (4.13) if we are able to simulate independent copies of Z is to devise a Monte Carlo estimator. In this part, we present a reduction variance method based on quantized control variates. Let Ξ_N our d dimensional control variate

$$\Xi^N := (\Xi_k^N)_{k=1,\ldots,d}$$

where each component Ξ_k^N is defined by

$$\Xi_k^N := f_k(Z_k) - \mathbb{E}\left[f_k(\widehat{Z}_k^N)\right],$$

with $f_k(z) := f(\mathbb{E}[Z_1], \ldots, \mathbb{E}[Z_{k-1}], z, \mathbb{E}[Z_{k+1}], \ldots, \mathbb{E}[Z_d])$ and $\widehat{Z}_k^N$ is an optimal quantizer of cardinality N of the component Z_k. One notices that the complexity for the evaluation of f_k is the same as the one of f. Now, defining $X^\lambda := f(Z) - \langle \lambda, \Xi^N \rangle$ where $\lambda \in \mathbb{R}^d$, we can introduce

$I^{\lambda,N}$ as an approximation for (4.13)

$$\begin{aligned} I^{\lambda,N} &:= \mathbb{E}\left[X^{\lambda}\right] \\ &= \mathbb{E}\left[f(Z) - \langle \lambda, \Xi^N \rangle\right] \\ &= \mathbb{E}\left[f(Z) - \sum_{k=1}^{d} \lambda_k f_k(Z_k)\right] + \sum_{k=1}^{d} \lambda_k \, \mathbb{E}\left[f_k(\widehat{Z}_k^N)\right]. \end{aligned} \tag{4.14}$$

The terms $\mathbb{E}\left[f_k(\widehat{Z}_k^N)\right]$ in (4.14) can be computed easily using the quantization-based cubature formula if we known the grids of the quantizers $(\widehat{Z}_k^N)_{k=1,\dots,d}$ and their associated weights.

Remark. We look for the $\lambda_{\min}$ minimizing the variance of X^{λ}

$$\mathbb{V}\mathrm{ar}(X^{\lambda_{\min}}) = \min\left\{ \mathbb{V}\mathrm{ar}\left(f(Z) - \langle \lambda, \Xi^N \rangle\right), \lambda \in \mathbb{R}^d \right\}.$$

The solution of the above optimization problem is the solution of following system

$$D(Z) \cdot \lambda = B$$

where $D(Z)$, the covariance-variance matrix of $\left(f_k(Z_k)\right)_{k=1,\dots,d}$, and B are given by

$$D(Z) = \begin{pmatrix} \mathbb{V}\mathrm{ar}\left(f_1(Z_1)\right) & \cdots & \mathbb{C}\mathrm{ov}\left(f_1(Z_1), f_d(Z_d)\right) \\ \vdots & \ddots & \vdots \\ \mathbb{C}\mathrm{ov}\left(f_d(Z_d), f_1(Z_1)\right) & \cdots & \mathbb{V}\mathrm{ar}\left(f_d(Z_d)\right) \end{pmatrix}, \quad B = \begin{pmatrix} \mathbb{C}\mathrm{ov}\left(f(Z), f_1(Z_1)\right) \\ \vdots \\ \mathbb{C}\mathrm{ov}\left(f(Z), f_d(Z_d)\right) \end{pmatrix}.$$

The solution to this optimization problem can easily be solved numerically using any library of linear algebra able to solve linear systems thanks to QR or LU decompositions.

Remark. If the Z_k's are independent hence λ can be determined easily. Indeed, in that case the matrix $D(Z)$ is diagonal. Then, the λ_k's are given by

$$\lambda_k = \frac{\mathbb{C}\mathrm{ov}\left(f_k(Z_k), f(Z)\right)}{\mathbb{V}\mathrm{ar}\left(f_k(Z_k)\right)}.$$

Now, we can define $\widehat{I}_M^{\lambda,N}$ the associated Monte Carlo estimator of $I^{\lambda,N}$

$$\widehat{I}_M^{\lambda,N} = \frac{1}{M} \sum_{m=1}^{M} \left(f(Z^m) - \sum_{k=1}^{d} \lambda_k f_k(Z_k^m) \right) + \sum_{k=1}^{d} \lambda_k \, \mathbb{E}\left[f_k(\widehat{Z}_k^N)\right].$$

One notices that $\mathbb{E}\left[I - I^{\lambda,N}\right] \neq 0$, with bias equal to $\sum_{k=1}^{d} \lambda_k \left(\mathbb{E}\left[f_k(\widehat{Z}_k^N)\right] - \mathbb{E}\left[f_k(Z_k)\right] \right)$. However the quantity we are really interested by is not the bias but the *MSE* (Mean Squared

Error), yielding a *bias-variance decomposition*

$$\mathrm{MSE}(\widehat{I}_M^{\lambda,N}) = \underbrace{\left(\sum_{k=1}^{d} \lambda_k \Big(\mathbb{E}\left[f_k(\widehat{Z}_k^N)\right] - \mathbb{E}\left[f_k(Z_k)\right]\Big)\right)^2}_{bias^2} + \frac{1}{M} \underbrace{\mathbb{V}\mathrm{ar}\left(f(Z) - \sum_{k=1}^{d} \lambda_k f_k(Z_k)\right)}_{Monte\ Carlo\ variance}.$$

Our aim is to minimize the cost of the Monte Carlo simulation for a given *MSE* or upper-bound of the *MSE*. Consequently, for a given Monte Carlo estimator $\widehat{I}_M^{\lambda,N}$ our minimization problem reads

$$\inf_{MSE(\widehat{I}_M^{\lambda,N}) \leqslant \epsilon^2} \mathrm{Cost}(\widehat{I}_M^{\lambda,N}). \tag{4.15}$$

Let $\kappa = \mathrm{Cost}(f(z))$ for a given $z \in \mathbb{R}^d$, the cost of a standard Monte Carlo estimator $\widehat{I}_M$ of size M is $\mathrm{Cost}(\widehat{I}_M) = \kappa M$. In our controlled case, if we neglect the cost for building an optimal quantizer, the global complexity associated to the Monte-Carlo estimator $\widehat{I}_M^{\lambda,N}$ is given by

$$\mathrm{Cost}(\widehat{I}_M^{\lambda,N}) = \kappa\big((d+1)M + dN\big)$$

where the cost of the computation of $f(z) - \lambda \sum_{k=1}^{d} f_k(z)$ is upper-bounded by $(d+1)\kappa$ whereas κdN is the cost of the quantized part. Indeed, there is d expectations of functions of N-quantizers to compute, inducing a cost of order κdN. Some optimizations can be implemented when computing $f_k(z)$, in that case $\mathrm{Cost}(f_k(z)) < \kappa$. So, (4.15) becomes

$$\inf_{MSE(\widehat{I}_M^{\lambda,N}) \leqslant \epsilon^2} \kappa\big((d+1)M + dN\big).$$

Moreover, using the results in the first part of the paper concerning the weak error, we could define an upper-bound for the $MSE(\widehat{I}_M^{\lambda,N})$, indeed if each f_k is in a class of function where the weak error of order two is attained when using a quantization-based cubature formula then

$$MSE(\widehat{I}_M^{\lambda,N}) = \left(\sum_{k=1}^{d} \lambda_k \Big(\mathbb{E}\left[f_k(\widehat{Z}_k^N)\right] - \mathbb{E}\left[f_k(Z_k)\right]\Big)\right)^2 + \frac{\sigma_\lambda^2}{M} \leqslant \frac{C}{N^4} + \frac{\sigma_\lambda^2}{M}$$

with $\sigma_\lambda^2 := \mathbb{V}\mathrm{ar}\left(f(Z) - \sum_{k=1}^{d} \lambda_k f_k(Z_k)\right)$. Now, our minimization problem becomes

$$\inf_{\frac{C}{N^4} + \frac{\sigma_\lambda^2}{M} \leqslant \epsilon^2} \kappa\big((d+1)M + dN\big).$$

$\frac{C}{N^4}$ corresponds to the squared empirical bias and $\frac{\sigma_\lambda^2}{M}$ to the empirical variance, hence a standard approach when dealing with this kind of problem, is to equally divide ϵ^2 between the bias and

the variance: $\frac{C}{N^4} = \frac{\epsilon^2}{2}$ and $\frac{\sigma_\lambda^2}{M} = \frac{\epsilon^2}{2}$ yielding

$$N = O(\epsilon^{-\frac{1}{2}}) \quad \text{and} \quad M = O(\epsilon^{-2}),$$

hence the cost would be of order $O(\epsilon^{-2})$. However, as the cost is additive and in the case where σ_λ^2 is close to $\mathbb{V}\mathrm{ar}\big(f(Z)\big)$, meaning that the control variate does not really reduce the variance, we want to reduce the bias as much as we can. So another idea could be to choose both terms M and N of order $O(\epsilon^{-2})$, because the impact on the cost of the Monte Carlo is at least of this order. Then, we search $\theta \in (0,1)$ defined by

$$\theta\epsilon^2 = \frac{C}{N^4} \quad \text{and } (1-\theta)\epsilon^2 = \frac{\sigma_\lambda^2}{M},$$

such that the impact on the cost of the Monte Carlo part and the quantization part are of same order: $O(\epsilon^{-2})$. In that case, θ is given by

$$\begin{cases} \theta\epsilon^2 = \frac{C}{N^4} \\ \kappa dN = O(\epsilon^{-2}) \end{cases} \qquad \Longrightarrow \qquad \theta = O(\epsilon^6).$$

In practice, we do take not that high value for N. Indeed, the bias converges to 0 as N^{-4}, so taking optimal quantizers of size 200 or 500 is enough for considering that the bias is negligible compared to the residual variance of the Monte Carlo estimator.

Remark. Now, if we consider that we have no closed-form for $\mathbb{E}[Z_k]$, $k = 1, \dots, d$, then we need to approximate them by m_k (this would impact the total cost of the method, as one would need to use a numerical method for computing the m_k's but this can be done once and for all before estimating $\widehat{I}_M^{\lambda,N}$). These approximations yield different control variates: the functions $\widetilde{f}_k(z) := f(m_1, \dots, m_{k-1}, z, m_{k+1}, \dots, m_d)$, inducing a different *MSE*

$$MSE(\widehat{I}_M^{\widetilde{\lambda},N}) = \left(\sum_{k=1}^{d} \widetilde{\lambda}_k \Big(\mathbb{E}\left[\widetilde{f}_k(\widehat{Z}_k^N)\right] - \mathbb{E}\left[\widetilde{f}_k(Z_k)\right] \Big) \right)^2 + \frac{\widetilde{\sigma}_{\widetilde{\lambda}}^2}{M}$$

with $\widetilde{\sigma}_{\widetilde{\lambda}}^2 := \mathbb{V}\mathrm{ar}\left(f(Z) - \sum_{k=1}^d \widetilde{\lambda}_k \widetilde{f}_k(Z_k)\right)$ and $\widetilde{\lambda}_k$, $k = 1, \dots, d$. Finally, we can conclude in the same way as before if the $\widetilde{f}_k$'s are in a class of function where the weak error of order two is attained when using a quantization-based cubature formula.

4.4.2 Numerical results

Let $(S_t)_{t\in[0,T]}$ be a geometric Brownian motion representing the dynamic of a *Black-Scholes* asset between time $t = 0$ and time $t = T$ defined by

$$S_t = S_0\, \mathrm{e}^{(r-\sigma^2/2)t + \sigma W_t}$$

with $(W_t)_{t\in[0,T]}$ a standard Brownian motion defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, r the interest rate and σ the volatility. When considering to use optimal quantization with a Black-Scholes asset, we have two possibilities: either we take an optimal quantizer of a normal distribution as $W_T \sim \mathcal{N}(0,T)$ or we build an optimal quantizer of a log-normal distribution as $\log(\mathrm{e}^{(r-\sigma^2/2)T+\sigma W_T}) \sim \mathcal{N}\left((r-\sigma^2/2)T, \sigma^2 T\right)$. In this part we consider both approaches since each one has its benefits and drawbacks.

Optimal Quantizers of log-normal random variables need to be computed each time we consider different parameters for the Black-Scholes asset. Indeed, the only operations preserving the optimality of the quantizers are translations and scaling. However, this transformations are not enough if one wishes to build an optimal quantizer of a Log-Normal random variables with parameters μ and σ from an optimal quantizer of a standardized Log-Normal random variable. However, if one looses time by computing for each set of parameters an optimal quantizer for the log-normal random variable, it gains in precision.

Now, if we consider the case of optimal quantizers of normal random variables, we loose in precision because we do not quantize directly our asset but the optimal quantizers of normal random variables can be computed once and for all and stored on a file. Indeed, we can build every normal random variable from a standard normal random variable using translations and scaling. Moreover, high precision grids of the $\mathcal{N}(0,1)$-distribution are in free access for download at the website: www.quantize.maths-fi.com.

Substantial details concerning the optimization problem and the numerical methods for building quadratic optimal quantizers can be found in [Pag18; PP03; PPP04b; McW+18]. In our case, we chose to build all the optimal quantizers with the Newton-Raphson algorithm (see [PP03] for more details on the gradient and Hessian formulas for the $\mathcal{N}(0,1)$-distribution and [McW+18] for other distributions) modified with the Levenberg-Marquardt procedure which improves the robustness of the method.

4.4.2.1 Vanilla Call

The payoff of a Call expiring at time T is

$$(S_T - K)_+$$

with K the strike and T the maturity of the option. Its price, in the special case of *Black-Scholes* model, is given by the following closed formula

$$I_0 := \mathbb{E}\left[\mathrm{e}^{-rT}(S_T-K)_+\right] = Call_{BS}(S_0, K, r, \sigma, T) = S_0\,\mathcal{N}(d_1) - K\,\mathrm{e}^{-rT}\,\mathcal{N}(d_2) \tag{4.16}$$

where $\mathcal{N}(x)$ is the cumulative distribution function of the standard normal distribution, $d_1 := \frac{\log(S_0/K)+(r+\sigma^2/2)T}{\sigma\sqrt{T}}$ and $d_2 := d_1 - \sigma\sqrt{T}$. Although the price of a Call in the Black-Scholes model can be expressed in a closed form, it is a good exercise to test new numerical methods

against this benchmark. We compare the use of optimal quantizers of normal distribution, when one quantizes the law of the Brownian motion at time T and log-normal distribution when one quantizes directly the law of the asset S_T at time T.

In the first case, we can rewrite I_0 as a function of a random variable Z with a $\mathcal{N}(0,1)$-distribution, namely a normal distributed random variable,

$$\mathbb{E}\left[\mathrm{e}^{-rT}(S_T - K)_+\right] = \mathbb{E}\left[f(Z)\right]$$

where $f(x) := \mathrm{e}^{-rT}(s_0\,\mathrm{e}^{(r-\sigma^2/2)T+\sigma\sqrt{T}x} - K)_+$ is continuous with a piecewise-defined locally-Lipschitz derivative, with respect to the function $g(x) = \mathrm{e}^{\sigma\sqrt{T}|x|}$.

In the second case, we have

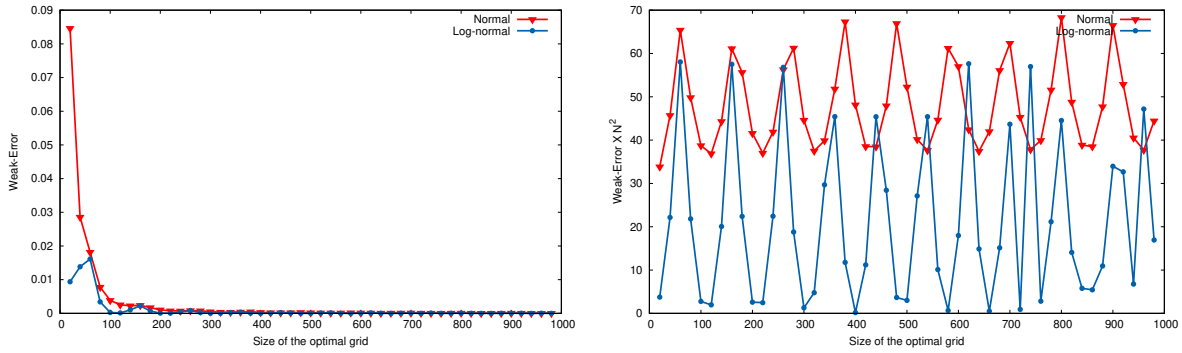
$$\mathbb{E}\left[\mathrm{e}^{-rT}(S_T - K)_+\right] = \mathbb{E}\left[\varphi(S_T)\right]$$

where $\varphi(x) := \mathrm{e}^{-rT}(x - K)_+$ is piecewise affine with one break of affinity.

The Black-Scholes parameters considered are

$$s_0 = 100, \qquad r = 0.1, \qquad \sigma = 0.5,$$

whereas those of the Call option are $T = 1$ and $K = 80$. The reference value is 34.15007. The first graphic in the Figure 4.1 represents the weak error between the benchmark and the quantization-based approximations in function of the size of the grid: $N \longmapsto \left|I_0 - \mathbb{E}\left[f(\widehat{Z}^N)\right]\right|$ and $N \longmapsto \left|I_0 - \mathbb{E}\left[\varphi(\widehat{X}^N)\right]\right|$, the second represents the weak error multiplied by N^2 in function of N: $N \longmapsto N^2 \times \left|I_0 - \mathbb{E}\left[f(\widehat{Z}^N)\right]\right|$ and $N \longmapsto N^2 \times \left|I_0 - \mathbb{E}\left[\varphi(\widehat{X}^N)\right]\right|$.



(a) $N \longmapsto \left|I_0 - \mathbb{E}\left[f(\widehat{Z}^N)\right]\right|$ (▼) *and* $N \longmapsto \left|I_0 - \mathbb{E}\left[\varphi(\widehat{X}^N)\right]\right|$ (●)

(b) $N \longmapsto N^2 \times \left|I_0 - \mathbb{E}\left[f(\widehat{Z}^N)\right]\right|$ (▼) *and* $N \longmapsto N^2 \times \left|I_0 - \mathbb{E}\left[\varphi(\widehat{X}^N)\right]\right|$ (●)

Fig. 4.1 *Call option in a Black-Scholes model.*

First, we notice that both methods yield a weak-error of order 2, as desired. Second, if we look closely at the results the log-normal grids give a more precise price. However we need to build a specific grid each time we have a new set of parameters for the asset, whereas such is

not the case when we choose to quantize the normal random variable, we can directly read precomputed grids with their associated weights in files.

4.4.2.2 Compound Option

The second product we consider is a Compound Option: a Put-on-Call. The payoff of a Put-on-Call expiring at time T_1 is the following

$$\Big(K_1 - \mathbb{E}\big[\,\mathrm{e}^{-r(T_2-T_1)}(S_{T_2}-K_2)_+ \mid S_{T_1}\big]\Big)_+$$

with price

$$I_0 := \mathbb{E}\left[\,\mathrm{e}^{-rT_1}\,\Big(K_1 - \mathbb{E}\big[\,\mathrm{e}^{-r(T_2-T_1)}(S_{T_2}-K_2)_+ \mid S_{T_1}\big]\Big)_+\right]. \tag{4.17}$$

The inner expectation can be computed, using the fact that S_{T_2} is a *Black-Scholes* asset and we know the conditional law of S_{T_2} given S_{T_1}. Using (4.16), the value of the inner expectation is

$$\mathbb{E}\big[\,\mathrm{e}^{-r(T_2-T_1)}(S_{T_2}-K_2)_+ \mid S_{T_1}\big] = Call_{BS}(S_{T_1}, K_2, r, \sigma, T_2 - T_1).$$

Hence, the price of the Put-On-Call option in (4.17) can be rewritten as

$$I_0 = \mathbb{E}\left[\,\mathrm{e}^{-rT_1}\,\big(K_1 - Call_{BS}(S_{T_1}, K_2, r, \sigma, T_2 - T_1)\big)_+\right].$$

The Black-Scholes parameters considered are

$$s_0 = 100, \qquad r = 0.03, \qquad \sigma = 0.2,$$

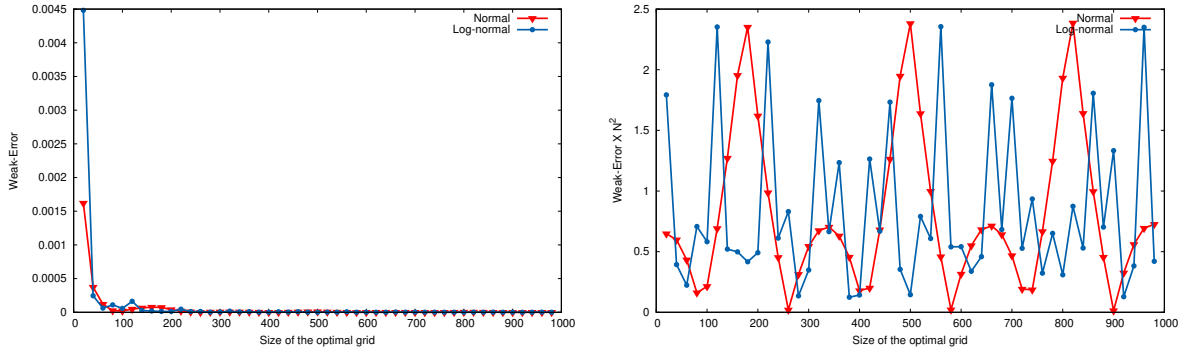
whereas those of the Put-On-Call option are $T_1 = 1/12$, $T_2 = 1/2$, $K_1 = 6.5$ and $K_2 = 100$. The reference value, obtained using an optimal quantizer of size 10000 of the $\mathcal{N}(0,1)$-distribution, is 1.3945704. As in the vanilla case, we compare the use of optimal quantizers of normal distribution and log-normal distribution. In the first case, we have

$$I_0 = \mathbb{E}\big[f(Z)\big]$$

where $Z \sim \mathcal{N}(0,1)$ and $f(z) = \mathrm{e}^{-rT_1}\,\big(K_1 - Call_{BS}(s_0\,\mathrm{e}^{(r-\sigma^2/2)T_1+\sigma\sqrt{T_1}z}, K_2, r, \sigma, T_2 - T_1)\big)_+$, and in the second case

$$I_0 = \mathbb{E}\big[\varphi(X)\big]$$

where $\log(X) \sim \mathcal{N}((r-\sigma^2/2)T, \sigma\sqrt{T})$ and $\varphi(x) = \mathrm{e}^{-rT_1}\,\big(K_1 - Call_{BS}(s_0 x, K_2, r, \sigma, T_2 - T_1)\big)_+$. The first graphic in Figure 4.2 represents the weak error between the benchmark and the quantization-based approximations in function of the size of the grid: $N \longmapsto \big|I_0 - \mathbb{E}\big[f(\widehat{Z}^N)\big]\big|$ and $N \longmapsto \big|I_0 - \mathbb{E}\big[\varphi(\widehat{X}^N)\big]\big|$, the second allows us to observe if the rate of convergence is indeed of order 2.



(a) $N \longmapsto \left|I_0 - \mathbb{E}\left[f(\widehat{Z}^N)\right]\right|$ (▼) *and* $N \longmapsto \left|I_0 - \mathbb{E}\left[\varphi(\widehat{X}^N)\right]\right|$ (●)

(b) $N \longmapsto N^2 \times \left|I_0 - \mathbb{E}\left[f(\widehat{Z}^N)\right]\right|$ (▼) *and* $N \longmapsto N^2 \times \left|I_0 - \mathbb{E}\left[\varphi(\widehat{X}^N)\right]\right|$ (●)

Fig. 4.2 *option in a Black-Scholes model.*

We notice that both methods yield a weak-error of order 2 as desired, however it is not clear that one should use the log-normal representation of (4.17) in place of the Gaussian representation. Indeed, both constants in the rate of convergence are of the desired order and getting Gaussian optimal quantizers is much cheaper than building optimal quantizers of log-normal random variables. Hence, one should choose the Gausian representation as it is as precise as the log-normal one and is much cheaper.

4.4.2.3 Exchange spread Option

In this part, we consider a higher dimensional problem. Let two Black-Scholes assets $(S_T^i)_{i=1,2}$ at time T related to two Brownian motions $(W_T^i)_{i=1,2}$, with correlation $\rho \in [-1,1]$. We are interested by an exchange spread option with strike K with payoff

$$(S_T^1 - S_T^2 - K)_+$$

whose price is

$$I_0 := \mathbb{E}\left[\,\mathrm{e}^{-rT}(S_T^1 - S_T^2 - K)_+\right]. \tag{4.18}$$

Decomposing the two Brownian motions into two independent parts, we have $(W_T^1, W_T^2) = \sqrt{T}(\sqrt{1-\rho^2}Z_1 + \rho Z_2, Z_2)$, where Z_1 and Z_2 are two independent $\mathcal{N}(0,1)$-distributed Gaussian random variables. Now, pre-conditioning on Z_2 in (4.18) and using (4.16), we have

$$I_0 = \mathbb{E}\left[\varphi(Z_2)\right]$$

where

$$\varphi(z) = Call_{BS}(s_0^1\,\mathrm{e}^{-\rho^2\sigma_1^2 T/2 + \sigma_1\rho\sqrt{T}z}, s_0^2\,\mathrm{e}^{(r-\sigma_2^2/2)T+\sigma_2\sqrt{T}z} + K, r, \sigma_1\sqrt{1-\rho^2}, T).$$

The numerical specifications of the function φ are as follows:

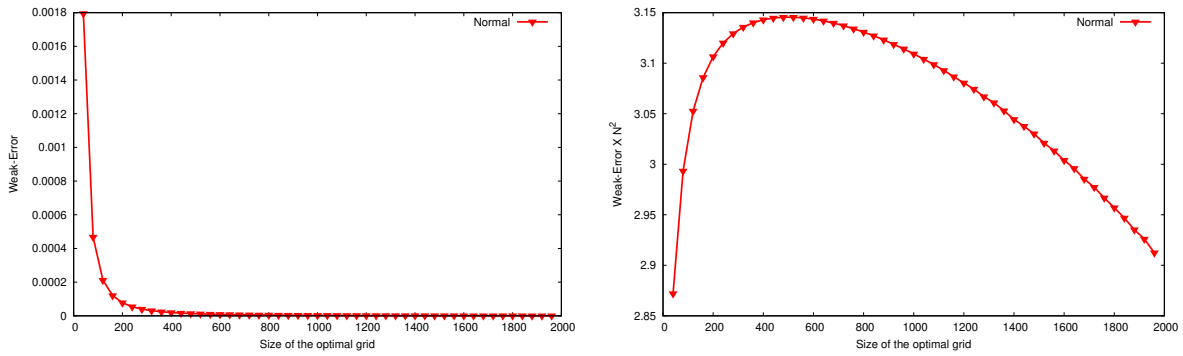
$$s_0^i = 100, \quad r = 0.02, \quad \sigma_i = 0.5, \quad \rho = 0.5, \quad T = 10, \quad K = 10.$$

In that case, the reference value is 53.552678.

First, we look at the weak error induced by the quantization-based cubature formula when approximating (4.18). We use optimal quantizers of the normal random variable Z_2. The quantization-based approximation is denoted $\widehat{I}_N$,

$$\widehat{I}_N := \mathbb{E}\left[\varphi(\widehat{Z}^N)\right].$$

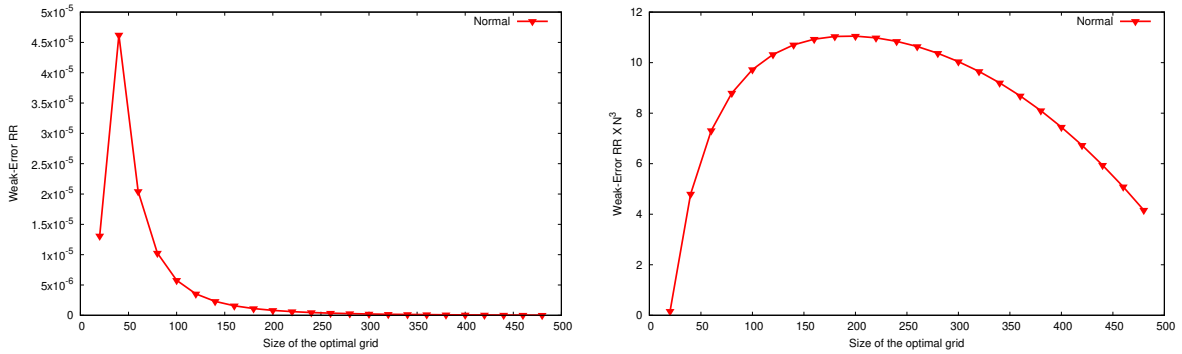
The first graphic in Figure 4.3 represents the weak error between the benchmark and the quantization-based approximation in function of the size of the grid: $N \longmapsto \left|I_0 - \mathbb{E}\left[\varphi(\widehat{Z}^N)\right]\right|$, the second plots $N \longmapsto N^2 \times \left|I_0 - \mathbb{E}\left[\varphi(\widehat{Z}^N)\right]\right|$ and allows us to observe that the rate of convergence is indeed of order 2.



(a) $N \longmapsto \left|I_0 - \mathbb{E}\left[\varphi(\widehat{Z}^N)\right]\right|$ *(▼)* (b) $N \longmapsto N^2 \times \left|I_0 - \mathbb{E}\left[\varphi(\widehat{Z}^N)\right]\right|$ *(▼)*

Fig. 4.3 *Exchange spread option pricing in a Black-Scholes model.*

Now, noticing that φ is a twice differentiable function with a bounded second derivative, we show that we can attain a weak error of order 3 when using a Richardson-Romberg extrapolation denoted $\widehat{I}^{RR}_{\widetilde{N},N}$ and defined in (4.11).



(a) $N \longmapsto |I_0 - \widehat{I}^{RR}_{\widetilde{N},N}|$ (▼) (b) $N \longmapsto N^3 \times |I_0 - \widehat{I}^{RR}_{\widetilde{N},N}|$ (▼)

Fig. 4.4 *Richardson-Romberg extrapolation, with* $\widetilde{N} = 1.2 \times N$*, for Exchange spread option pricing in a Black-Scholes model.*

4.4.2.4 Basket Option

A typical financial product that allows to diversify the market risk and to invest in options is a basket option. The simplest one is an option on a weighted average of stocks. For example, if we consider an option on the FTSE index, this is a basket option where the assets are the companies defined in the description of the index and the weights are the market capitalization of each company at the time we built the index normalized by the sum on all market capitalizations.

In this part, we consider d correlated assets $(S^k_T)_{k=1,\dots,d}$ following a Black-Scholes model and the payoff we consider is

$$f(S^1_t, \dots, S^d_T) := \left(\sum_{k=1}^{d} \alpha_k S^k_T - K \right)_+ \tag{4.19}$$

whose price is

$$I_0 := \mathrm{e}^{-rT}\, \mathbb{E}\left[\left(\sum_{k=1}^{d} \alpha_k S^k_T - K \right)_+ \right].$$

I_0 cannot be computed directly, hence we use a Monte Carlo estimator in order to approximate the expectation. The standard estimator, denoted $\widehat{I}_M$, is the crude Monte Carlo estimator and is given by

$$\widehat{I}_M := \mathrm{e}^{-rT}\, \frac{1}{M} \sum_{m=1}^{M} \left(\sum_{k=1}^{d} \alpha_k S^{k,(m)}_T - K \right)_+$$

where $(S^{k,(m)}_T)_{m=1,\dots,M}$ are i.i.d. copies of S^k_T. We compare the crude estimator to our novel approach based on a d-dimensional quantized control variates Ξ^N. In that case, I_0 is

approximated by I^N defined by

$$I^N := \mathrm{e}^{-rT}\,\mathbb{E}\left[\left(\sum_{k=1}^{d}\alpha_k S_T^k - K\right)_+ - \langle\lambda,\Xi^N\rangle\right]$$

where Ξ^N is defined later, yielding the following Monte Carlo estimator

$$\widehat{I}_M^{\lambda,N} := \mathrm{e}^{-rT}\,\frac{1}{M}\sum_{m=1}^{M}\left(\sum_{k=1}^{d}\alpha_k S_T^{k,(m)} - K\right)_+ - \langle\lambda,\Xi^{N,(m)}\rangle.$$

We propose two different control variates Ξ^N based on optimal quantizers either of log-normal random variables or of Gaussian random variables.

1. The control variate, denoted $\overline{\Xi}^N$, is defined by, $\forall k = 1,\dots,d$

$$\overline{\Xi}_k^N := f(\mathbb{E}[S_T^1],\dots,S_T^k,\dots,\mathbb{E}[S_T^d]) - \mathbb{E}\left[f(\mathbb{E}[S_T^1],\dots,\widehat{S}_T^{k,N},\dots,\mathbb{E}[S_T^d])\right]$$

   where $(\widehat{S}_T^{k,N})_{k=1,\dots,d}$ are optimal quantizers of cardinality N of S_T^k. In that case, the Monte Carlo estimator is denoted $\widehat{\overline{I}}_M^{\lambda,N}$.

2. The control variate, denoted $\widetilde{\Xi}^N$, is using another representation of the payoff (4.19), using d Gaussian random variables i.i.d in place of the assets S_T^k because the d underlying correlated Brownian Motions can be expressed from d rescaled independent Gaussian random variables, thus we define φ our new representation for the payoff as

$$\varphi(Z^1,\dots,Z^d) := f(S_T^1,\dots,S_T^d)$$

   where $(Z^k)_{k=1,\dots,d}$ are i.i.d Gaussian random variables. Now, defining our control variates with the function φ, $\forall k = 1,\dots,d$

$$\widetilde{\Xi}_k^N := \varphi(0,\dots,Z^k,\dots,0) - \mathbb{E}\left[\varphi(0,\dots,\widehat{Z}^N,\dots,0)\right]$$

   where $(\widehat{Z}^N)_{k=1,\dots,d}$ is an optimal quantizer of $Z\sim\mathcal{N}(0,1)$. In that case, the Monte Carlo estimator is denoted $\widehat{\widetilde{I}}_M^{\lambda,N}$.

The Black-Scholes parameters considered are

$$s_0^i = 100, \qquad r = 2\%, \qquad \sigma_i = \frac{i}{d+1}, \quad \rho = 0.5,$$

and the specifications of the product are

$$K = 100, \quad \alpha_i = \frac{2i}{d(d+1)}, \quad T = 1$$

such that $\sum \alpha_i = 1$. The benchmarks used for the computation of the *MSE* has been computed using a Monte Carlo estimator with control variate without quantization where the term $\sum_{k=1}^{d} \mathbb{E}[X_k]$ is computed using Black-Scholes Call pricing closed formulas. The *Mean Squared Error* of an estimator I is computed using the formula

$$MSE(I) = \frac{1}{n} \sum_{i=1}^{n} (I^{(i)} - I_0)^2$$

where $(I^{(i)})_{i=1,\dots,n}$ are n independent copies of I.

Table 4.1 compares three different types of Monte Carlo estimators: the standard (Crude) Monte Carlo estimator $\widehat{I}_M$, our novel Monte Carlo estimator with control variate based on optimal quantizers of Gaussian random variables $\widehat{\widehat{I}}_M^{\lambda,N}$ and another one with optimal quantizers of log-normal random variables $\widehat{\widehat{I}}_M^{\lambda,N}$. The notation n corresponds to the number of Monte Carlo used for computing the *MSE*, M is the size of each Monte Carlo and N is the size of the optimal quantizers. The prices of reference for each d are

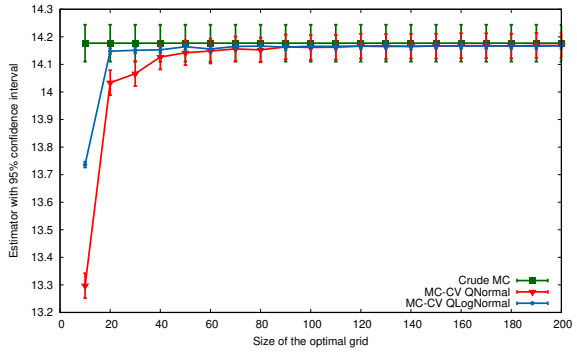
- for $d = 2$: 14.2589 ($\pm$0.0010),
- for $d = 3$: 14.1618 ($\pm$0.0015),
- for $d = 5$: 13.9005 ($\pm$0.0022),
- for $d = 10$: 13.4979 ($\pm$0.0034).

		$N = 20$		$N = 200$	
d	MC Estimator	Mean ($\pm$1.96×std)	MSE	Mean ($\pm$1.96×std)	MSE
$d = 2$	Crude	14.2695 ($\pm$0.0662)	0.1450	14.2695 ($\pm$0.0662)	0.1450
	CV Gaussian	14.1017 ($\pm$0.0399)	0.0774	14.2773 ($\pm$0.0399)	0.0530
	CV Log-Normal	14.2351 ($\pm$0.0078)	0.0026	14.2614 ($\pm$0.0078)	0.0020
$d = 3$	Crude MC	14.1770 ($\pm$0.0671)	0.1492	14.1770 ($\pm$0.0671)	0.1492
	CV Gaussian	14.0336 ($\pm$0.0451)	0.0837	14.1685 ($\pm$0.0451)	0.0673
	CV Log-Normal	14.1479 ($\pm$0.0104)	0.0038	14.1674 ($\pm$0.0104)	0.0036
$d = 5$	Crude MC	13.8803 ($\pm$0.0720)	0.1717	13.8803 ($\pm$0.0720)	0.1717
	CV Gaussian	13.6686 ($\pm$0.0562)	0.1580	13.8883 ($\pm$0.0562)	0.1044
	CV Log-Normal	13.8797 ($\pm$0.0151)	0.0080	13.9008 ($\pm$0.0151)	0.0076
$d = 10$	Crude MC	13.5046 ($\pm$0.0599)	0.1186	13.5046 ($\pm$0.0599)	0.1186
	CV Gaussian	13.2429 ($\pm$0.0515)	0.1527	13.5113 ($\pm$0.0515)	0.0878
	CV Log-Normal	13.4221 ($\pm$0.0194)	0.0181	13.4983 ($\pm$0.0194)	0.0124

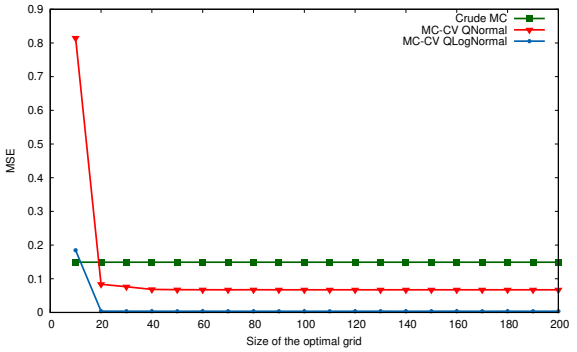
Table 4.1 $n = 128$, $M = 1e4$

One remarks in Table 4.1 the efficiency of the optimal quantization-based variance reduction method. The variance, in the best cases, can be divided by almost 100 when using the optimal quantizers of Log-Normal random variables. Figure 4.5 shows the effect of N (for $d = 3$), the

size the optimal quantizers, on the bias. The same seeds are used for all the Monte Carlo estimator, the only thing varying is N.



(a) $N \longmapsto |I_0 - \widehat{\overline{I}}_M^{\lambda,N}|$ (▼), $N \longmapsto |I_0 - \widehat{\widetilde{I}}_M^{\lambda,N}|$ (●) *and the Crude Monte Carlo estimator* (■) *with their associated confidence interval at* 95%.



(b) $N \longmapsto MSE(\widehat{I}_M)$ (■), $N \longmapsto MSE(\widehat{\overline{I}}_M^{\lambda,N})$ (▼) *and* $N \longmapsto MSE(\widehat{\widetilde{I}}_M^{\lambda,N})$ (●).

Fig. 4.5 $n = 128$, $M = 1e4$, $d = 3$.

Chapter 5

Stationary Heston model: Calibration and Pricing of exotics using Product Recursive Quantization

This chapter corresponds to the preprint "Stationary Heston model: Calibration and Pricing of exotics using Product Recursive Quantization" accessible in arXiv or HAL (see [LMP20]). This article is a joint work with Vincent Lemaire and Gilles Pagès.

Abstract A major drawback of the Standard Heston model is that its implied volatility surface does not produce a steep enough smile when looking at short maturities. For that reason, we introduce the Stationary Heston model where we replace the deterministic initial condition of the volatility by its invariant measure and show, based on calibrated parameters, that this model produce a steeper smile for short maturities than the Standard Heston model. We also present numerical solution based on Product Recursive Quantization for the evaluation of exotic options (Bermudan and Barrrier options).

Introduction

Originally introduced by Heston in [Hes93], the Heston model is a stochastic volatility model used in Quantitative Finance to model the joint dynamics of a stock and its volatility, denoted $(S_t^{(x)})_{t\geqslant 0}$ and $(v_t^x)_{t\geqslant 0}$, respectively, where $v_0^x = x$ is the initial condition of the volatility. Historically, the initial condition of the volatility x is considered as deterministic and is calibrated in the market like the other parameters of the model. This model received an important attention among practitioners for two reasons: first, it is a stochastic volatility model,

hence it introduces smile in the implied volatility surface as observed in the market, which is not the case of models with constant volatility and second, in its original form, we have access to a semi closed-form formula for the characteristic function which allows us to price European options (Call & Put) almost instantaneously using the Fast Fourier approach (Carr & Madan in [CM99]). Yet, a complaint often heard about the Heston model is that it fails to fit the implied volatility surface for short maturities because the model cannot produce a steep-enough smile for those maturities (see [Gat11]).

Noticing that the volatility process is ergodic with a unique invariant distribution $\nu = \Gamma(\alpha, \beta)$ where the parameters α and β depend on the volatility diffusion parameters, it has been first proposed by Pagès & Panloup in [PP09] to directly consider that the process evolves under its stationary regime in place of starting it at time 0 from a deterministic value. We denote by $(S_t^{(\nu)})_{t\geqslant 0}$ and $(v_t^{\nu})_{t\geqslant 0}$ the couple asset-volatility in the Stationary Heston model. Replacing the initial condition of the volatility by the stationary measure does not modify the long-term behavior of the implied volatility surface but does inject more randomness into the volatility for short maturities. This tends to produce a steeper smile for short maturities, which is the kind of behavior we are looking for. Later, the short-time and long-time behavior of the implied volatility generated by such model has been studied by Jacquier & Shi in [JS17].

In the beginning of the paper, we briefly recall the well-known methodology used for the pricing of European option in the Standard Heston model. Based on that, we express the price I_0 of a European option on the asset $S_T^{(\nu)}$ as

$$I_0 = \mathbb{E}\left[\mathrm{e}^{-rT}\,\varphi(S_T^{(\nu)})\right] = \mathbb{E}\left[f(v_0^{\nu})\right] \tag{5.1}$$

where $f(v)$ is the price of the European option in the Standard Heston model for a given set of parameters. The last expectation can be computed efficiently using quadrature formulas either based on optimal quantization of the Gamma distribution or on Laguerre polynomials.

Once we are able to price European options, we can think of calibrating our model to market data. Indeed the parameters of the model are calibrated using the implied volatility surface observed in the market. However, the calibration of the Standard Heston model is highly depending on the initial guess we choose in the minimization problem. This is due to an over-parametrization of the model (see [GR09]). Hence, when we consider the Heston model in its stationary regime, there is one parameter less to calibrate as the initial value of the volatility is no longer deterministic. The stationary model tends to be more robust when it comes to calibration.

In the second part of paper, we deal with the pricing of Exotic options such as Bermudan and Barrier options. We propose a method based on hybrid product recursive quantization. The "hybrid" term comes from the fact that we use two different types of schemes for the discretization of the volatility and the asset (Milstein and Euler-Maruyama). The recursive quantization (also called Markovian quantization) was first introduced in [PPP04b] and then

studied extensively by Pagès & Sagna in [PS15] for one dimensional diffusions discretized by an Euler-Maruyama scheme. They proposed a fast algorithm based on deterministic methods for building the quantization tree. Then, the fast recursive quantization was extended to one-dimensional higher-order schemes by [McW+18] and to higher dimensions using product quantization (see [FSP18; Rud+17; CFG18; CFG17]). Then, once the quantization tree is built, we proceed by a backward induction using the Backward Dynamic Programming Principle for the price of Bermudan options and using the methodology detailed in [Sag10; Pag18] based on the conditional law of the Brownian Bridge for the price of Barrier options.

The paper is organized as follows. First, in Section 5.1, we recall the definition of the Standard Heston model and the interesting features of the volatility diffusion which bring us to define the Stationary Heston model. In Section 5.2, we give a fast solution for the pricing of European options in the Stationary Heston model when there exists methods for the Standard model. Finally, once we are able to price European options, we can define the optimization problem of calibration on implied volatility surface. We perform the calibration of both models and compare their induced smile for short maturities options. Once this model has been calibrated, in Section 5.3, we propose a numerical method based on hybrid product recursive quantization for the pricing of exotic financial products: Bermudan and Barrier options. For this method, we give an estimate of the L^2-error introduced by the approximation.

5.1 The Heston Model

The Standard Heston model is a two-dimensional diffusion process $(S_t^{(x)}, v_t^x)$ solution to the Stochastic Differential Equation

$$\begin{cases} \dfrac{dS_t^{(x)}}{S_t^{(x)}} = (r-q)dt + \sqrt{v_t^x}\left(\rho d\widetilde{W}_t + \sqrt{1-\rho^2}dW_t\right) \\ dv_t^x = \kappa(\theta - v_t^x)dt + \xi\sqrt{v_t^x}d\widetilde{W}_t \end{cases} \tag{5.2}$$

where

- $S_t^{(x)}$ is the dynamic of the risky asset,
- v_t^x is the dynamic of the volatility process,
- $S_0^{(x)} = s_0 \geqslant 0$ is the initial value of the process,
- $r \in \mathbb{R}$ denotes the interest rate,
- $q \in \mathbb{R}$ is the dividend rate,
- $\rho \in [-1, 1]$ is the correlation between the asset and the volatility,

- $(W, \widetilde{W})$ is a two-dimensional standard Brownian motion,
- $\theta \geqslant 0$ the long run average price variance,
- $\kappa \geqslant 0$ the rate at which v_t^x reverts to θ,
- $\xi \geqslant 0$ is the volatility of the volatility,
- $v_0^x = x \geqslant 0$ is the deterministic initial condition of the volatility.

This model is widely used by practitioner for various reasons. One is that it leads to semi-closed forms for vanilla options based on a fast Fourier transform. The other is that it represents well the observed mid and long-term market behavior of the implied volatility surface observed on the market. However, it fails producing or even fitting to the smile observed for short-term maturities.

Remark (The volatility). One can notice that the volatility process is autonomous thence we are facing a one dimensional problem. Moreover, the volatility process is following a Cox-Ingersoll-Ross (CIR) diffusion also known as the square root diffusion. Existence and uniqueness of a strong solution to this stochastic differential equation has been first shown in [IW81], if $x \geqslant 0$. Moreover, it has been shown, see [LL11], that if the Feller condition holds, namely $\xi^2 \leqslant 2\kappa\theta$, for every $x > 0$, then the unique solution $(v_t^x)_{t\geqslant 0}$ satisfies

$$\forall t \geqslant 0, \quad \mathbb{P}(\tau_0^x = +\infty) = 1 \tag{5.3}$$

where τ_0^x is the first hitting time defined by

$$\tau_0^x = \inf\{t \geqslant 0 \mid v_t^x = 0\} \quad \text{where } \inf \varnothing = +\infty. \tag{5.4}$$

Moreover, the CIR diffusion admits, as a Markov process, a unique stationary regime, characterized by its invariant distribution

$$\nu = \Gamma(\alpha, \beta) \tag{5.5}$$

where

$$\alpha = \theta\beta \quad \text{and} \quad \beta = 2\kappa/\xi^2. \tag{5.6}$$

Based on the above remarks, the idea is to precisely consider the volatility process under its stationary regime, i.e., replacing the deterministic initial condition from the Standard Heston model by a ν-distributed random variable independent of $(W, \widetilde{W})$. We will refer to this model as the Stationary Heston model. Our first aim is to inject more randomness for short maturities (t small) into the volatility but also to reduce the number of free parameters to stabilize and robustify the calibration of the Heston model which is commonly known to be overparametrized (see e.g. [GR09]).

This model was first introduced by [PP09] (see also [IW81], p. 221). More recently, [JS17] studied its small-time and large-time behaviors of the implied volatility. The dynamic of the asset price $(S_t^{(\nu)})_{t\geqslant 0}$ and its stochastic volatility $(v_t^\nu)_{t\geqslant 0}$ in the Stationary Heston model are given by

$$\begin{cases} \dfrac{dS_t^{(\nu)}}{S_t^{(\nu)}} = (r-q)dt + \sqrt{v_t^\nu}\big(\rho d\widetilde{W}_t + \sqrt{1-\rho^2}dW_t\big) \\ dv_t^\nu = \kappa(\theta - v_t^\nu)dt + \xi\sqrt{v_t^\nu}d\widetilde{W}_t \end{cases} \tag{5.7}$$

where $v_0^\nu \sim \mathcal{L}(\nu) \sim \Gamma(\alpha,\beta)$ with $\beta = 2\kappa/\xi^2$, $\alpha = \theta\beta$. $S_0^{(\nu)}$, r and q are the same parameters as those defined in (5.2) and the parameters ρ, θ, κ, θ and ξ can be described as in the Standard Heston model.

5.2 Pricing of European Options and Calibration

In this section, we first calibrate both Stationary and Standard Heston models and then compare their short-term behaviors of their resulting implied volatility surfaces. For that purpose we relied on a dataset of options price on the Euro Stoxx 50 observed the 26th of September 2019 (see Figure 5.1). This is why, as a preliminary step we briefly recall the well-known methodology for the evaluation of European Call and Put in the Standard Heston model. Based on that, we outline how to price these options in the Stationary Heston model. Then, we describe the methodology employed for the calibration of both models: the Stationary Heston model (5.7) and the Standard Heston model (5.2) and then we discuss the obtained parameters and compare their short-term behaviors.

5.2.1 European options

The price of the European option with payoff φ on the asset $S_T^{(\nu)}$, under the Stationary Heston model, exercisable at time T is given by

$$I_0 = \mathbb{E}\big[\,\mathrm{e}^{-rT}\,\varphi(S_T^{(\nu)})\big]. \tag{5.8}$$

After preconditioning by v_0^ν, we have

$$I_0 = \mathbb{E}\Big[\,\mathbb{E}\big[\,\mathrm{e}^{-rT}\,\varphi(S_T^{(\nu)}) \mid \sigma(v_0^\nu)\big]\Big] = \mathbb{E}\big[f(v_0^\nu)\big] \tag{5.9}$$

where $f(v)$ is the price of the European option in the Standard Heston model with deterministic initial conditions for the set of parameters $\lambda(v) = (s_0, r, q, \theta, \kappa, \xi, \rho, v)$.

Example 5.2.1 (Call)**.** If φ is the payoff of a Call option then f is simply the price given by Fourier transform in the Standard Heston model of the European Call Option. The price at time 0, for a spot price s_0, of an European Call $C(\lambda(v), K, T)$ with expiry T and strike K

under the Standard Heston model with parameters $\lambda(v) = (s_0, r, q, \theta, \kappa, \xi, \rho, v)$ is

$$\begin{aligned} C(\lambda(v), K, T) &= \mathbb{E}\left[\mathrm{e}^{-rT}(S_T^{(v)} - K)_+\right] \\ &= \mathrm{e}^{-rT}\left(\mathbb{E}\left[S_T^{(v)}\,\mathbb{1}_{S_T^{(v)} \geqslant K}\right] - K\,\mathbb{E}\left[\mathbb{1}_{S_T^{(v)} \geqslant K}\right]\right) \\ &= s_0\,\mathrm{e}^{-qT}\,P_1\big(\lambda(v), K, T\big) - K\,\mathrm{e}^{-rT}\,P_2\big(\lambda(v), K, T\big) \end{aligned} \tag{5.10}$$

with $P_1\big(\lambda(v), K, T\big)$ and $P_2\big(\lambda(v), K, T\big)$ given by

$$\begin{aligned} P_1\big(\lambda(v), K, T\big) &= \frac{1}{2} + \frac{1}{\pi}\int_0^{+\infty} \mathrm{Re}\left(\frac{\mathrm{e}^{-\mathbf{i}u\log(K)}}{iu}\,\frac{\psi\big(\lambda(v), u - \mathbf{i}, T\big)}{s_0\,\mathrm{e}^{(r-q)T}}\right)du \\ P_2\big(\lambda(v), K, T\big) &= \frac{1}{2} + \frac{1}{\pi}\int_0^{+\infty} \mathrm{Re}\left(\frac{\mathrm{e}^{-\mathbf{i}u\log(K)}}{\mathbf{i}u}\,\psi\big(\lambda(v), u, T\big)\right)du \end{aligned} \tag{5.11}$$

where $\mathbf{i}$ is the imaginary unit s.t. $\mathbf{i}^2 = -1$, $\psi\big(\lambda(v), u, T\big)$ is the characteristic function of the logarithm of the stock price process at time T. Several representations of the characteristic function exist, we choose to use the one proposed by [SST04; Gat11; Alb+07], which is numerically more stable. It reads

$$\begin{aligned} \psi\big(\lambda(v), u, T\big) &= \mathbb{E}\left[\mathrm{e}^{\mathbf{i}u\log(S_T^{(v)})} \mid S_0^{(v)}, x\right] \\ &= \mathrm{e}^{\mathbf{i}u(\log(s_0)+(r-q)T)} \\ &\quad \times \mathrm{e}^{\theta\kappa\xi^{-2}\left((\kappa-\rho\xi u\mathbf{i}-d)T - 2\log((1-g\,\mathrm{e}^{-dt})/(1-g))\right)} \\ &\quad \times \mathrm{e}^{v^2\xi^{-2}(\kappa-\rho\xi u\mathbf{i}-d)(1-\mathrm{e}^{-dt})/(1-g\,\mathrm{e}^{-dt})} \end{aligned} \tag{5.12}$$

with

$$d = \sqrt{(\rho\xi u\mathbf{i} - \kappa)^2 - \xi^2(-u\mathbf{i} - u^2)} \quad \text{and} \quad g = (\kappa - \rho\xi u\mathbf{i} - d)/(\kappa - \rho\xi u\mathbf{i} + d). \tag{5.13}$$

Hence, in (5.9), $f(v)$ can be replaced by $C\big(\lambda(v), K, T\big)$, which yields

$$I_0 = \mathbb{E}\left[\mathrm{e}^{-rT}(S_T^{(\nu)} - K)_+\right] = \mathbb{E}\Big[C\big(\lambda(v_0^\nu), K, T\big)\Big]. \tag{5.14}$$

Now, we come to the pricing of European options in the Stationary Heston model, using the expression of the density of $v_0^\nu \sim \Gamma(\alpha, \beta)$, (5.9) reads

$$I_0 = \mathbb{E}\big[f(v_0^\nu)\big] = \int_0^{+\infty} f(v)\frac{\beta^\alpha}{\Gamma(\alpha)}v^{\alpha-1}\,\mathrm{e}^{-\beta v}\,dv. \tag{5.15}$$

Now, several approaches exists in order to approximate this integral on the positive real line.

- *Quantization based quadrature formulas.* One could use a quantization-based cubature formula with an optimal quantizer of v_0^ν with the methodology detailed in Appendix 5.D. Given

that optimal quantizer of size N, $\widehat{v}_0^N$, we approximate I_0 by $\widehat{I}_0^N$

$$\widehat{I}_0^N = \mathbb{E}\left[f(\widehat{v}_0^N)\right] = \sum_{i=1}^{N} f(v_{0,i}^N)\,\mathbb{P}\left(\widehat{v}_0^N = v_{0,i}^N\right). \tag{5.16}$$

Remarks. In one dimension, the minimization problem, that consists in building an optimal quantizer, is invariant by linear transformation. Hence applying a linear transformation to an optimal quantizer preserves its optimality. For example, if we consider an optimal quantization $\widehat{X}^N$ of a standard normal distribution $\mathcal{N}(0,1)$ then $\mu + \sigma\widehat{X}^N$ is an optimal quantizer of a normal distribution $\mathcal{N}(\mu,\sigma^2)$ and the associated probabilities of each Voronoï centroid stay the same.

In our case, noticing that if we consider a Gamma random variable $X \sim \Gamma(\alpha,1)$ then the rescaling of X by $1/\beta$ yields $X/\beta \sim \Gamma(\alpha,\beta)$. Hence, for building the optimal quantizer $\widehat{v}_0^N$ of v_0^ν, we can build an optimal quantizer of $X \sim \Gamma(\alpha,1)$ and then rescale it by $1/\beta$, yielding $\widehat{v}_0^N = \widehat{X}^N/\beta$. Our numerical tests showed that it is numerically more stable to use this approach.

In order to build the optimal quantizer, we use Lloyd's method detailed in Appendix 5.D to $X \sim \Gamma(\alpha,1)$ with the cumulative distribution function $F_X(x) = \mathbb{P}(X \leqslant x)$ and the partial first moment $K_X(x) = \mathbb{E}[X\,\mathbb{1}_{X\leqslant x}]$ given by

$$\begin{aligned} \forall x>0, \qquad & F_X(x) = \frac{1}{\Gamma(\alpha)}\gamma(\alpha,x), \qquad && K_X(x) = \alpha F_X(x) - \frac{x^\alpha\,\mathrm{e}^{-x}}{\Gamma(\alpha)}, \\ \text{otherwise}, \qquad & F_X(x) = 0, && K_X(x) = 0, \end{aligned} \tag{5.17}$$

where $\gamma(\alpha,x) = \int_0^x t^{\alpha-1}\,\mathrm{e}^{-t}\,dt$ is the lower gamma function. And the associated probabilities of the optimal quantizer $\widehat{v}_0^N$ are given by (5.113)

$$\mathbb{P}\left(\widehat{v}_0^N = v_{0,i}^N\right) = \mathbb{P}\left(\widehat{X}^N = x_i^N\right) = F_X\left(x_{i+1/2}^N\right) - F_X\left(x_{i-1/2}^N\right) \tag{5.18}$$

where $\forall i \in [\![2,N]\!], x_{i-1/2}^N = \frac{x_{i-1}^N + x_i^N}{2}$ and $x_{1/2}^N = 0$ and $x_{N+1/2}^N = +\infty$.

- *Quadrature formula from Laguerre polynomials.* One could also use an algorithm based on fixed point quadratures for the numerical integration. Indeed, noticing that the density we are integrating against is a gamma density which is exactly the Laguerre weighting function (up to a rescaling). Then, I_0 rewrites

$$I_0 = \int_0^{+\infty} f(v)\frac{\beta^\alpha}{\Gamma(\alpha)}v^{\alpha-1}\,\mathrm{e}^{-\beta v}\,dv = \frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^{+\infty} f(v)\omega(v)dv \tag{5.19}$$

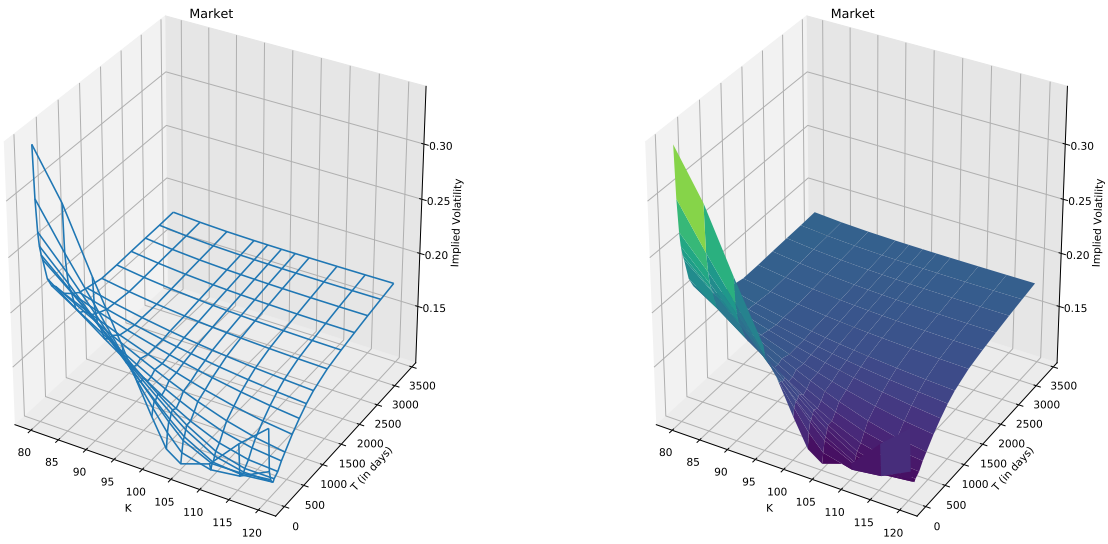


Fig. 5.1 *Implied volatility surface of the* Euro Stoxx *50 as of the 26th of September 2019. ($S_0 = 3541$, $r = -0.0032$ and $q = 0.00225$) The expiries T are given in days and the strikes K in percentage of the spot.*

where $\omega(v) = v^{\alpha-1}\,\mathrm{e}^{-\beta v}$ is the Laguerre weighting function. Then, for a fixed integer $n \geqslant 1$[1], I_0 is approximated by

$$\widetilde{I}_0^n = \frac{\beta^\alpha}{\Gamma(\alpha)} \sum_{i=1}^{n} \omega_i f(v_i) \tag{5.20}$$

where the ω_i's are the Laguerre weights and the v_i's are the associated Laguerre nodes.

5.2.2 Calibration

Now that we are able to compute the price of European options, we define the problem of minimization we wish to optimize in order to calibrate our models parameters. Let $\mathcal{P}_{SH}$ be the set of parameters of the Stationary Heston model that needs to be calibrated, defined by

$$\mathcal{P}_{SH} = \big\{(\theta, \kappa, \xi, \rho) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+ \times [-1, 1]\big\} \tag{5.21}$$

and let $\mathcal{P}_H$ be the set of parameters of the Standard Heston model that needs to be calibrated, defined by

$$\mathcal{P}_H = \big\{(x, \theta, \kappa, \xi, \rho) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+ \times [-1, 1]\big\}. \tag{5.22}$$

The others parameters are directly inferred from the market: we get $S_0 = 3541$, $r = -0.0032$ and $q = 0.00225$. In our case, we calibrate to option prices all having the same maturity.

[1] In practice, we choose $n = 20$. This number of points allows us to reach a high precision while keeping the computation time under control.

The problem can be formulated as follows: we search for the set of parameters $\phi^\star \in \mathcal{P}$ that minimizes the relative error between the implied volatility observed on the market and the implied volatility produced by the model for the given set of parameters, such that $\mathcal{P} = \mathcal{P}_{SH}$ for the Stationary Heston model and $\mathcal{P} = \mathcal{P}_H$ for the Standard Heston model. There is no need to calibrate the parameters s_0, r and q since they are directly observable in the market.

Being interested in the short-term behaviors of the models, it is natural to calibrate both models based on options prices at a small expiry. Once the optimization procedures have been performed, we compare their performances for small expiries. For that, we calibrate using only the data on the volatility surface in Figure 5.1 with expiry 50 days ($T = 50/365$) and then we compare both models to the market implied volatility at expiry 22 days which is the smallest available in the data set.

Remark. The calibration is performed in C++ on a laptop with a 2,4 GHz 8-Core Intel Core i9 CPU using the randomized version of the simplex algorithm of [NM65] proposed in the C++ library GSL. This algorithm is a derivative-free optimization method. It uses only the value of the function at each evaluation point. The computation time for calibrating the Standard Heston model is around 20s and a bit more than a minute for the Stationary model. However, these computation times need to be considered carefully because the calibration time highly depends on the initial condition we choose for the minimizer and on the implementation of the Call pricer in the Standard Heston model.

5.2.2.1 Optimization without penalization

We want to find the set of parameter $\phi^\star$ that minimizes the relative error between the volatilities observed in the market and the ones generated by the model, hence leading to the following minimization problem

$$\min_{\phi \in \mathcal{P}} \sum_K \left(\frac{\sigma_{\text{IV}}^{Market}(K,T) - \sigma_{\text{IV}}^{Model}(\phi,K,T)}{\sigma_{\text{IV}}^{Market}(K,T)} \right)^2 \tag{5.23}$$

where T is the expiry of the chosen options chosen a priori and K are their strikes. $\sigma_{\text{IV}}^{Market}(K,T)$ is the Mark-to-Market implied volatility taken from the observed implied volatility surface and the implied volatility $\sigma_{\text{IV}}^{Model}(\phi,K,T)$ is the Black-Scholes volatility σ that matches the European Call price in this model to the price given by the Standard or Stationary Heston model with the set of parameters ϕ.

In all the following figures, the strike K is given in percentage of the spot S_0.

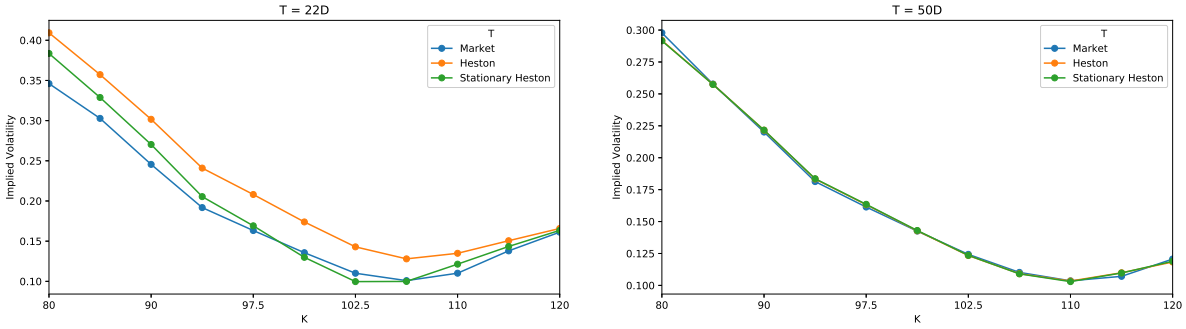


Fig. 5.2 *Implied volatilities for* 22 *(left) and* 50 *(right) days expiry options after calibration at* 50 *days without penalization.*

It is clear in Figure 5.2 (right) that both models fit really well to the market data and more precisely, the Stationary model succeeds to calibrate with the same precision as the Standard one with one less parameter. Moreover, one notices that even for 22 days maturity options, the Standard Heston model tends to over-estimate the implied volatility and fails to produce the right smile whereas the Stationary Heston model is closer to the market observations.

Now, we extrapolate the implied volatility surfaces, given by the two models, for even smaller maturities (7 and 14 days) in order to analyze the behavior of each model for short-term expiries.

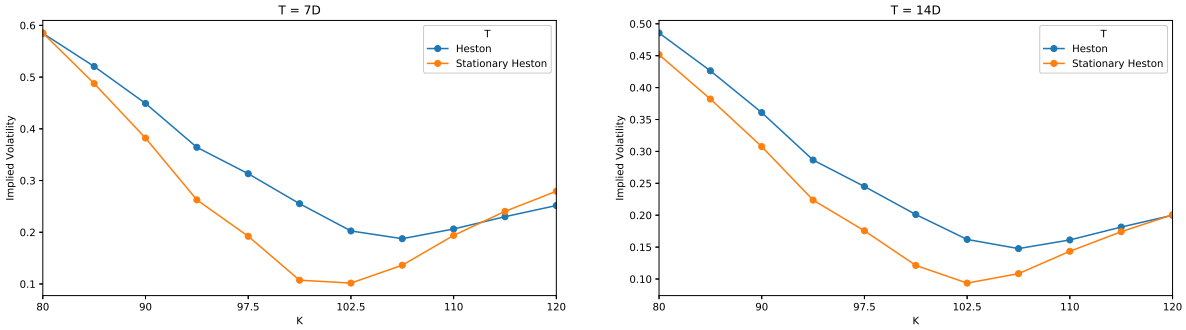


Fig. 5.3 *Implied volatilities for* 7 *(left) and* 14 *(right) days expiry options after calibration at* 50 *days without penalization.*

It is clear in Figure 5.3 that the Standard Heston model fails at producing the desired smile for very small maturities when the Stationary model meets no difficulty to generate it. The next graphics, Figure 5.4 reproduces the term-structure of the implied volatility in function of T both models.

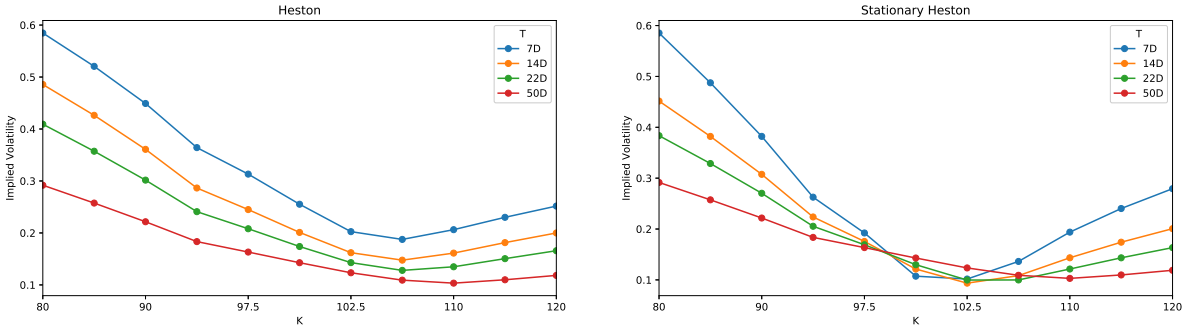


Fig. 5.4 *Term-structure of the volatility in function of T and K of both models (left: Standard Heston and right: Stationary Heston) after calibration at* 50 *days without penalization.*

Now, we investigate how these models behave for longer maturities. Do they succeed in preserving the general shape of the market volatility surface or are they only correctly fitting the maturity on which we calibrated them?

Figure 5.5 represents the relative error between the implied volatility given by the market and the one given by the models calibrated models at 50 days. Clearly, one notices that the Standard Heston model only fits at this expiry. Indeed, when looking at the expiry 22 days or for long-term maturities, the relative error explodes. The term-structure of the implied volatility surface of the market is not preserved when using the Standard Heston model. However, the Stationary Heston model does fit well at both short and long term expiries. The Stationary model produces a steep smile for very short maturities and flattens correctly to the appropriate mean for long expiries.

$\phi^\star$	ρ	v_0	θ	κ	ξ
Standard Heston	-0.74	0.152584	0.01487	80.05	5.22
Stationary Heston	-0.75		0.02744	593.46	36.80

Table 5.1 *Parameters obtained for both models after calibration without penalization for options with maturity* 50 *days ($S_0 = 3541$, $r = -0.0032$ and $q = 0.00225$).*

However, looking closely at the parameters obtained after calibration (which are summarized in Table 5.1), one notices that both sets of calibrated parameters are far from satisfying the Feller condition. And we have to keep in mind that the calibration procedure is performed in order to price path-dependent or American style derivatives using Monte-Carlo simulation or alternative numerical methods, as developed in the next Section. Hence, the Feller condition has to be satisfied, this is the reason why we add a constraint to the minimization problem in order to penalize the sets of parameters not satisfying the condition.

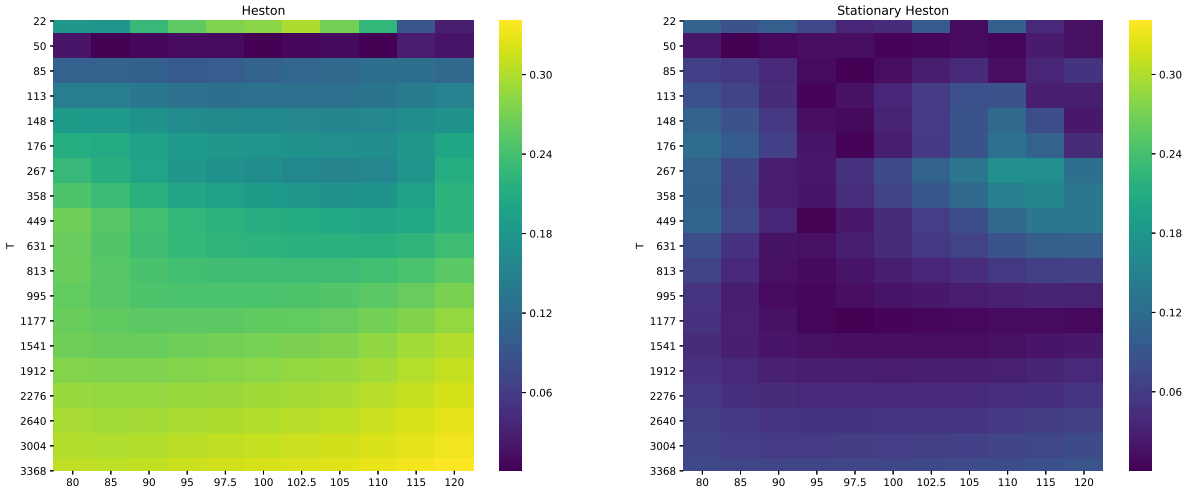


Fig. 5.5 $(K,T) \longrightarrow \frac{|\sigma_{\text{IV}}^{Market}(K,T)-\sigma_{\text{IV}}^{Model}(\phi^\star,K,T)|}{\sigma_{\text{IV}}^{Market}(K,T)}$ *for both models after calibration at* 50 *days without penalization. The expiries* T *are given in days and the strikes* K *are in percentage of the spot. (left: Standard Heston and right: Stationary Heston).*

5.2.2.2 Optimization with penalization using the Feller condition

The minimization problem becomes

$$\min_{\phi\in\mathcal{P}} \sum_K \left(\frac{\sigma_{\text{IV}}^{Market}(K,T) - \sigma_{\text{IV}}^{Model}(\phi,K,T)}{\sigma_{\text{IV}}^{Market}(K,T)} \right)^2 + \lambda \max(\xi^2 - 2\kappa\theta, 0) \tag{5.24}$$

where λ is the penalization factor to be adjusted during the procedure. The obtained parameters after calibration are summarized in Table 5.2. The Feller condition is still not fulfilled for both models but it is not far from being satisfied. We choose $\lambda = 0.01$ which seems to be right the compromise in order to avoid underfitting the model because of the constraint.

$\phi^\star$	ρ	v_0	θ	κ	ξ
Standard Heston	-0.83	0.0045	0.17023	2.19	1.04
Stationary Heston	-0.99		0.02691	19.28	1.15

Table 5.2 *Parameters obtained for both models after calibration with penalization* ($\lambda = 0.01$) *for options with maturity* 50 *days* ($S_0 = 3541$, $r = -0.0032$ *and* $q = 0.00225$).

Figure 5.6 displays the resulting implied volatility curves at 50 days and 22 days for both calibrated models and observed in the market with calibration at 50 days. Adding a penalization term deteriorates the calibration results compared to the non-penalized case (see Figure 5.2 (right)) but the results are still acceptable.

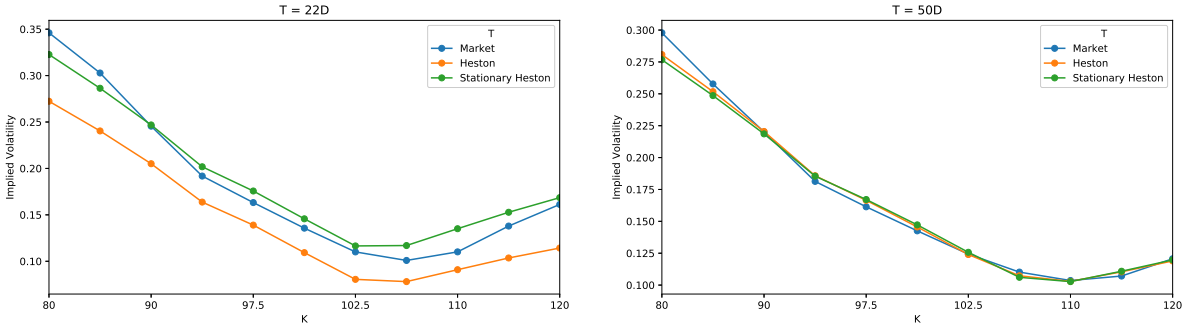


Fig. 5.6 *Implied volatilities for* 22 *(left) and* 50 *(right) days expiry options after calibration at* 50 *days with penalization.*

Now, again, we extrapolate the implied volatility of both models for very short term maturities in Figure 5.7. The Stationary Heston model produces the desired smile, however the Standard Heston model fails to produce prices sensibly different than 0 for strikes higher than 105 with this set of parameters, this is why there is no values in implied volatility curves.

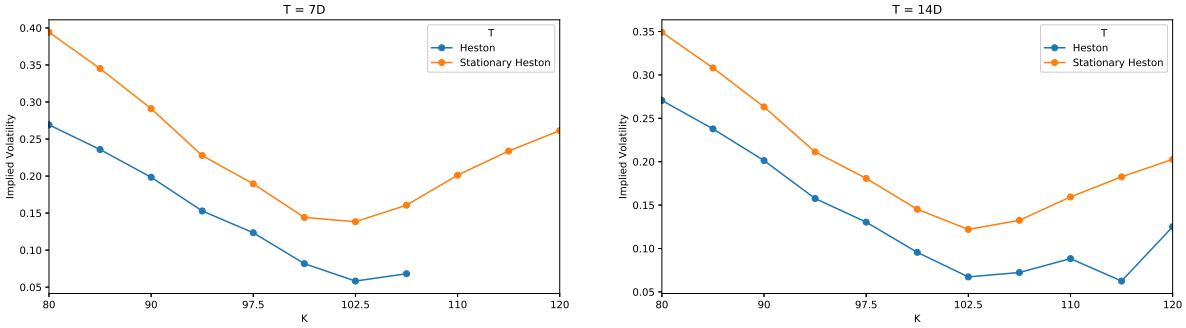


Fig. 5.7 *Implied volatilities for* 7 *(left) and* 14 *(right) days expiry options after calibration at* 50 *days with penalization.*

Figure 5.8 represents, as in the non-penalized case, the relative error between the implied volatility given by the market and the one given by the models calibrated models at 50 days using a penalization. The Standard Heston model completely fails to preserve the term-structure while being calibrated at 50 days. In comparison, the Stationary Heston behaves much better and the relative error does not explodes for long-term expiries, meaning that the long run average price variance is well caught.

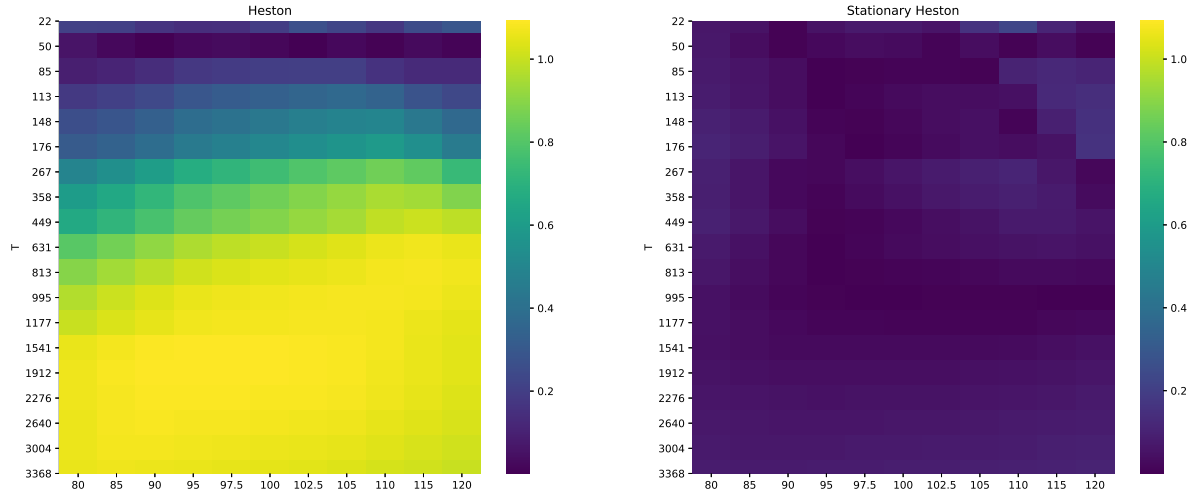


Fig. 5.8 $(K,T) \longrightarrow \frac{|\sigma_{\text{IV}}^{Market}(K,T) - \sigma_{\text{IV}}^{Model}(\phi^\star,K,T)|}{\sigma_{\text{IV}}^{Market}(K,T)}$ *for both models after calibration at* 50 *days with penalization. The expiries* T *are given in days and the strikes* K *are in percentage of the spot. (left: Standard Heston and right: Stationary Heston).*

5.3 Toward the pricing of Exotic Options

In this Section, we evaluate first Bermudan options and then Barrier options under the Stationary Heston model. For both products, the pricing rely on a *Backward Dynamic Programming Principle*. The numerical solution we propose is based on a two-dimensional product recursive quantization scheme. We extend the methodology previously developed by [FSP18; CFG18; CFG17], where they considered an Euler-Maruyama scheme for both components. In this paper, we consider a hybrid scheme made up with an Euler-Maruyama scheme for the log-stock price dynamics and a Milstein scheme for the (boosted) volatility process. Finally, we apply the backward algorithm that corresponds to the financial product we are dealing with (the *Quantized Backward Dynamic Programming Principle* for Bermudan Options, see [BP03; BPP05; Pag18] and the algorithm by [Sag10; Pag18] for Barrier Options based on the conditional law of the Brownian motion).

5.3.1 Discretization scheme of a stochastic volatility model

We first present the time discretization schemes we use for the asset-volatility couple $(S_t^{(\nu)}, v_t^\nu)_{t\in[0,T]}$. For the volatility, we choose a Milstein on a *boosted* version of the process in order to preserve the positivity of the volatility and we select an Euler-Maruyama scheme for the log of the asset.

The boosted volatility. Based on the discussion in Appendix 5.A, we will work with the following *boosted* volatility process: $Y_t = \mathrm{e}^{\kappa t} v_t^\nu, t \in [0,T]$ for some $\kappa > 0$, whose diffusion is

given by

$$dY_t = \mathrm{e}^{\kappa t}\,\kappa\theta dt + \xi\,\mathrm{e}^{\kappa t/2}\sqrt{Y_t}d\widetilde{W}_t. \tag{5.25}$$

The Milstein discretization scheme of Y_t is given by

$$\bar{Y}_{t_{k+1}} = \mathcal{M}_{\widetilde{b},\widetilde{\sigma}}\big(t_k, \bar{Y}_{t_k}, Z_{k+1}^2\big) \tag{5.26}$$

with $t_k = \frac{Tk}{n}$ and $\widetilde{b}$ and $\widetilde{\sigma}$ are given by

$$\widetilde{b}(t,x) = \mathrm{e}^{\kappa t}\,\kappa\theta, \qquad \widetilde{\sigma}(t,x) = \xi\sqrt{x}\,\mathrm{e}^{\kappa t/2} \qquad \text{and} \qquad \widetilde{\sigma}'_x(t,x) = \frac{\xi\,\mathrm{e}^{\kappa t/2}}{2\sqrt{x}} \tag{5.27}$$

and $\mathcal{M}_{\widetilde{b},\widetilde{\sigma}}(t,x,z)$ defined by

$$\mathcal{M}_{\widetilde{b},\widetilde{\sigma}}(t,x,z) = x - \frac{\widetilde{\sigma}(t,x)}{2\widetilde{\sigma}'_x(t,x)} + h\left(\widetilde{b}(t,x) - \frac{(\widetilde{\sigma}\widetilde{\sigma}'_x)(t,x)}{2}\right) + \frac{(\widetilde{\sigma}\widetilde{\sigma}'_x)(t,x)h}{2}\left(z + \frac{1}{\sqrt{h}\widetilde{\sigma}'_x(t,x)}\right)^2. \tag{5.28}$$

We made this choice of scheme because, under the Feller condition, the positivity of $\mathcal{M}_{\widetilde{b},\widetilde{\sigma}}$ is ensured, since

$$\mathcal{M}_{\widetilde{b},\widetilde{\sigma}}(t,x,z) = h\,\mathrm{e}^{\kappa t}\left(\kappa\theta - \frac{\xi^2}{4}\right) + h\frac{\xi^2\,\mathrm{e}^{\kappa t}}{4}\left(z + \frac{2\sqrt{x}}{\sqrt{h}\xi\,\mathrm{e}^{\kappa t/2}}\right)^2 \tag{5.29}$$

and

$$\xi^2 \leqslant 2\kappa\theta \leqslant 4\kappa\theta.$$

Other schemes could have been used, see [Alf05] for an extensive review of the existing schemes for the discretization of the CIR model, but in our case we needed one allowing us to use the fast recursive quantization, i.e., where we can express explicitly and easily the cumulative distribution function and the first partial moment of the scheme, which is the case of the Milstein scheme (we give more details in SubSection 5.3.2).

Hence, as our time-discretized scheme is well defined because its positivity is ensured if the Feller condition is satisfied, we can start to think of the time-discretization of our process $(S_{t_k}^{(\nu)})_{k\in[\![0,n]\!]}$.

The log-asset. For the asset, the standard approach is to consider the process which is the logarithm of the asset $X_t = \log(S_t)$. Applying Itô's formula, the dynamics of X_t is given by

$$dX_t = \left(r - q - \frac{v_t}{2}\right)dt + \sqrt{v_t}dW_t. \tag{5.30}$$

Now, using a standard Euler-Maruyama scheme for the discretization of X_t, we have

$$\begin{cases} \bar{X}_{t_{k+1}} = \mathcal{E}_{b,\sigma}\big(t_k, \bar{X}_{t_k}, \bar{Y}_{t_k}, Z^1_{k+1}\big) \\ \bar{Y}_{t_{k+1}} = \mathcal{M}_{\tilde{b},\tilde{\sigma}}\big(t_k, \bar{Y}_{t_k}, Z^2_{k+1}\big) \end{cases} \tag{5.31}$$

where $Z^1_{k+1} \sim \mathcal{N}(0,1)$, $Z^2_{k+1} \sim \mathcal{N}(0,1)$, $\mathbb{C}\mathrm{orr}(Z^1_{k+1}, Z^2_{k+1}) = \rho$ and

$$\mathcal{E}_{b,\sigma}(t,x,y,z) = x + b(t,x,y)h + \sigma(t,x,y)\sqrt{h}\,z \tag{5.32}$$

with

$$b(t,x,y) = r - q - \frac{\mathrm{e}^{-\kappa t}\,y}{2} \qquad \text{and} \qquad \sigma(t,x,y) = \mathrm{e}^{-\kappa t/2}\,\sqrt{y}. \tag{5.33}$$

5.3.2 Hybrid Product Recursive Quantization

In this part, we describe the methodology used for the construction of the product recursive quantization tree of the couple log asset- boosted volatility in the Heston model.

In Figure 5.9, as an example, we synthesise the main idea behind the recursive quantization of a diffusion v_t which has been time-discretized with $F_0(t,x,z)$. We start at time $t_0 = 0$ with a quantizer $\widehat{v}_0$ taking values in the grid $\Gamma_{t_0} = \{v^0_1, \ldots, v^0_{10}\}$ of size 10, where each point is represented by a black bullet ($\bullet$) with probability $p^0_i = \mathbb{P}(\widehat{v}_0 = v^0_i)$ is represented by a bar. In the Stationary Heston model, $\widehat{v}_0$ is an optimal quantization of the Gamma distribution given by (5.5) and (5.6). Then, starting from this grid, we simulate the process from time t_0 to time $t_1 = 5$ days with our chosen time-discretization scheme $F_0(t,x,z)$, yielding $\widetilde{v}_1 = F_0(t_0, \widehat{v}_0, Z_1)$, where Z_1 is a standardized Gaussian random variable. Each trajectory starts from point v^0_i with probability p^0_i. And finally we project the obtained distribution at time t_1 onto a grid $\Gamma_{t_1} = \{v^1_1, \ldots, v^1_{10}\}$ of cardinality 10, represented by black triangles ($\blacktriangle$) such that $\widehat{v}_1$ is an optimal quantizer of the discretized and simulated process starting from quantizer $\widehat{v}_0$ at time $t_0 = 0$.

Remark. In practice, for low dimensions, we do not simulate trajectories. We use the information on the law of $\widetilde{v}_1$ conditionally of starting from $\widehat{v}_0$. The knowledge of the distribution allows us to use deterministic algorithms during the construction of the optimal quantizer of $\widetilde{v}_1$ that are a lot faster than algorithms based on simulation.

In our case, we consider the following stochastic volatility system

$$\begin{cases} dX_t = b(t, X_t, Y_t)dt + \sigma(t, X_t, Y_t)dW_t \\ dY_t = \widetilde{b}(t, Y_t)dt + \widetilde{\sigma}(t, Y_t)d\widetilde{W}_t \end{cases} \tag{5.34}$$

where W_t and $\widetilde{W}_t$ are two correlated Brownian motions with correlation $\rho \in [-1,1]$, b and σ are defined in (5.33) and $\widetilde{b}$ and $\widetilde{\sigma}$ are defined in (5.27). Our aim is to build a quantization tree of the couple (X_t, Y_t) at given dates t_k, $k = 0, \ldots, n$ based on a recursive product quantization

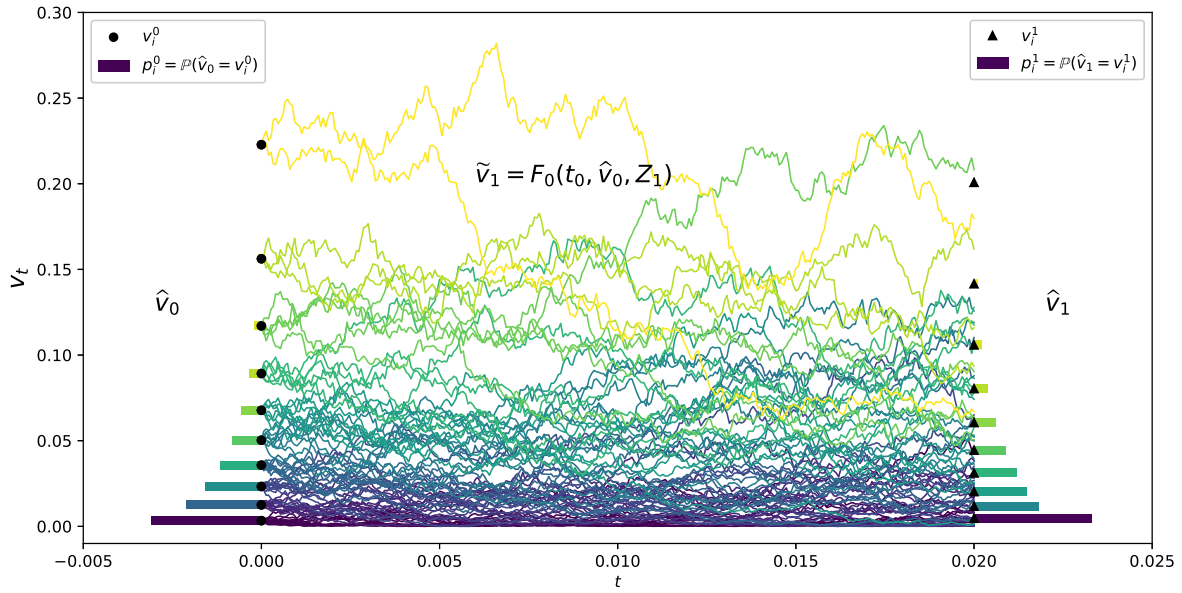


Fig. 5.9 *Example of recursive quantization of the volatility process in the Heston model for one time-step.*

scheme. The product recursive quantization of such diffusion system has already been studied by [CFG17] and [Rud+17] in the case case where both processes are discretized using an Euler-Maruyama scheme.

One can notice that building the quantization tree $(\widehat{Y}_k)_{k\in[\![0,n]\!]}$ approximating $(Y_t)_{t\in[0,T]}$ is a one dimensional problem as the diffusion of Y_t is autonomous. Hence, based on our choice of discretization scheme, we will apply the fast recursive quantization (detailed above in Figure 5.9) that was introduced in [PS15] for one dimensional diffusion discretized by an Euler-Maruyama discretization scheme and then extended to higher order schemes, still in one dimension, by [McW+18]. The minor difference with existing literature is that, in our problem, the initial condition y_0 is not deterministic.

Then, using the quantization tree of $(\widehat{Y}_k)_{k\in[\![0,n]\!]}$ we will be able to build the tree $(\widehat{X}_k)_{k\in[\![0,n]\!]}$ following ideas developed in [FSP18; Rud+17; CFG18; CFG17]. Indeed, once the quantization tree of the volatility is built, we are in a one-dimensional setting and we are able to use fast deterministic algorithms.

5.3.2.1 Quantizing the volatility (a one-dimensional case)

Let $(Y_t)_{t\in[0,T]}$ be a stochastic process in $\mathbb{R}$ and solution to the stochastic differential equation

$$dY_t = \widetilde{b}(t, Y_t)dt + \widetilde{\sigma}(t, Y_t)d\widetilde{W}_t \tag{5.35}$$

where Y_0 has the same law than the stationary measure ν: $\mathcal{L}(Y_0) = \nu$. In order to approximate our diffusion process, we choose a Milstein scheme for the time discretization, as defined in 5.28 and we build recursively the Markovian quantization tree $(\widehat{Y}_{t_k})_{k\in[\![0,n]\!]}$ where $\widehat{Y}_{t_{k+1}}$ is the Voronoï quantization of $\widetilde{Y}_{t_{k+1}}$ defined by

$$\widetilde{Y}_{t_{k+1}} = \mathcal{M}_{\widetilde{b},\widetilde{\sigma}}\big(t_k, \widehat{Y}_{t_k}, Z^2_{k+1}\big), \qquad \widehat{Y}_{t_{k+1}} = \mathrm{Proj}_{\Gamma^Y_{N_{2,k+1}}}\big(\widetilde{Y}_{t_{k+1}}\big) \tag{5.36}$$

and the projection operator $\mathrm{Proj}_{\Gamma^Y_{N_{2,k+1}}}(\cdot)$ is defined in (5.105), $\Gamma^Y_{N_{2,k+1}} = \big\{y^{k+1}_1, \dots, y^k_{N_{2,k+1}}\big\}$ is the grid of the optimal quantizer of $\widetilde{Y}_{t_{k+1}}$ and $Z^2_{k+1} \sim \mathcal{N}(0,1)$. In order alleviate the notations, we will denote $\widetilde{Y}_k$ and $\widehat{Y}_k$ in place of $\widetilde{Y}_{t_k}$ and $\widehat{Y}_{t_k}$.

The first step consists in building $\widehat{Y}_0$, an optimal quantizer of size $N_{2,0}$ of Y_0. Noticing that $Y_0 = v_0^\nu$, we use the optimal quantizer we built for the pricing of European options. Then, we build recursively $(\widehat{Y}_k)_{k=1,\dots,n}$, where the $N_{2,k}$-tuple are defined by $y^k_{1:N_{2,k}} = \big(y^k_1, \dots, y^k_{N_{2,k}}\big)$, by solving iteratively the minimization problem defined in the Appendix 5.D in (5.109), with the help of Lloyd's method I. Replacing X by $\widetilde{Y}_{k+1}$ in (5.109) yields

$$\begin{aligned} y^{k+1}_j &= \frac{\mathbb{E}\left[\widetilde{Y}_{k+1}\,\mathbb{1}_{Y_{k+1}\in C_j\left(\Gamma^Y_{N_{2,k+1}}\right)}\right]}{\mathbb{P}\left(\widetilde{Y}_{k+1}\in C_j\big(\Gamma^Y_{N_{2,k+1}}\big)\right)} \\ &= \frac{\mathbb{E}\left[\mathcal{M}_{\widetilde{b},\widetilde{\sigma}}\big(t_k,\widehat{Y}_k,Z^2_{k+1}\big)\,\mathbb{1}_{\mathcal{M}_{\widetilde{b},\widetilde{\sigma}}\left(t_k,\widehat{Y}_k,Z^2_{k+1}\right)\in C_j\left(\Gamma^Y_{N_{2,k+1}}\right)}\right]}{\mathbb{P}\left(\mathcal{M}_{\widetilde{b},\widetilde{\sigma}}\big(t_k,\widehat{Y}_k,Z^2_{k+1}\big)\in C_j\big(\Gamma^Y_{N_{2,k+1}}\big)\right)}. \end{aligned} \tag{5.37}$$

Now, preconditioning by $\widehat{Y}_k$ in the numerator and the denominator and using $p^k_i = \mathbb{P}\left(\widehat{Y}_k = y^k_i\right)$, we have

$$\begin{aligned} y^{k+1}_j &= \frac{\mathbb{E}\left[\mathbb{E}\left[\mathcal{M}_{\widetilde{b},\widetilde{\sigma}}\big(t_k,\widehat{Y}_k,Z^2_{k+1}\big)\,\mathbb{1}_{\mathcal{M}_{\widetilde{b},\widetilde{\sigma}}\left(t_k,\widehat{Y}_k,Z^2_{k+1}\right)\in C_j\left(\Gamma^Y_{N_{2,k+1}}\right)}\mid\widehat{Y}_k\right]\right]}{\mathbb{E}\left[\mathbb{P}\left(\mathcal{M}_{\widetilde{b},\widetilde{\sigma}}\big(t_k,\widehat{Y}_k,Z^2_{k+1}\big)\in C_j\big(\Gamma^Y_{N_{2,k+1}}\big)\mid\widehat{Y}_k\right)\right]} \\ &= \frac{\sum_{i=1}^{N_{2,k}}\mathbb{E}\left[\mathcal{M}_{\widetilde{b},\widetilde{\sigma}}\big(t_k,y^k_i,Z^2_{k+1}\big)\,\mathbb{1}_{\mathcal{M}_{\widetilde{b},\widetilde{\sigma}}\left(t_k,y^k_i,Z^2_{k+1}\right)\in C_j\left(\Gamma^Y_{N_{2,k+1}}\right)}\right]p^k_i}{\sum_{i=1}^{N_{2,k}}\mathbb{P}\left(\mathcal{M}_{\widetilde{b},\widetilde{\sigma}}\big(t_k,y^k_i,Z^2_{k+1}\big)\in C_j\big(\Gamma^Y_{N_{2,k+1}}\big)\right)p^k_i} \\ &= \frac{\sum_{i=1}^{N_{2,k}}\left(K^k_i\big(y^{k+1}_{j+1/2}\big)-K^k_i\big(y^{k+1}_{j-1/2}\big)\right)p^k_i}{\sum_{i=1}^{N_{2,k}}\left(F^k_i\big(y^{k+1}_{j+1/2}\big)-F^k_i\big(y^{k+1}_{j-1/2}\big)\right)p^k_i} \end{aligned} \tag{5.38}$$

where $C_j\big(\Gamma^Y_{N_{2,k+1}}\big) = \big(y^{k+1}_{j-1/2}, y^{k+1}_{j+1/2}\big]$ is defined in (5.104). F^k_i and K^k_i are the cumulative distribution function and the first partial moment function of $U^k_i \sim \mu^k_i + \kappa^k_i(Z^1_{k+1} + \lambda^k_i)^2$ respectively with

$$\begin{aligned} \kappa^k_j &= \frac{(\widetilde{\sigma}\widetilde{\sigma}'_x)(t_k, y^k_j)h}{2}, \qquad & \lambda^k_j &= \frac{1}{\sqrt{h}\widetilde{\sigma}'_x(t_k, y^k_j)}, \\ \text{and} \qquad \mu^k_j &= y^k_j - \frac{\sigma(t_k, y^k_j)}{2\widetilde{\sigma}'_x(t_k, y^k_j)} + h\left(\widetilde{b}(t_k, y^k_j) - \frac{(\widetilde{\sigma}\widetilde{\sigma}'_x)(t_k, y^k_j)}{2}\right). \end{aligned} \tag{5.39}$$

The functions F^k_i and K^k_i can explicitly be determined in terms of the density and the cumulative distribution function of the normal distribution.

Lemma 5.3.1. *Let $U = \mu + \kappa(Z + \lambda)^2$, with $\mu, \kappa, \lambda \in \mathbb{R}$, $\lambda \geqslant 0$, $\kappa > 0$ and $Z \sim \mathcal{N}(0,1)$ then the cumulative distribution function F_X and the first partial moment K_U of U are given by*

$$\begin{aligned} F_U(x) &= \big(F_Z(x_+) - F_Z(x_-)\big)\,\mathbb{1}_{x>\mu} \\ K_U(x) &= \left(F_U(x)\big(\mu + \kappa(\lambda^2+1)\big) + \frac{\kappa}{\sqrt{2\pi}}\left(x_-\,\mathrm{e}^{-\frac{x_+^2}{2}} - x_+\,\mathrm{e}^{-\frac{x_-^2}{2}}\right)\right)\mathbb{1}_{x>\mu} \end{aligned} \tag{5.40}$$

where $x_+ = \sqrt{\frac{x-\mu}{\kappa}} - \lambda$, $x_- = -\sqrt{\frac{x-\mu}{\kappa}} - \lambda$ and F_Z is the cumulative distribution function of Z.

Finally, we can apply the Lloyd algorithm defined in Appendix 5.112 with F_X and K_X defined by

$$F_X(x) = \sum_{i=1}^{N_{2,k}} p^k_i\, F^k_i(x) \qquad \text{and} \qquad K_X(x) = \sum_{i=1}^{N_{2,k}} p^k_i\, K^k_i(x). \tag{5.41}$$

In order to be able to build recursively the tree quantization $(\widehat{Y}_k)_{k=0,\dots,n}$, we need to have access to the weights $p^k_i = \mathbb{P}\left(\widehat{Y}_k = y^k_i\right)$, which can be themselves computed recursively, as well as the conditional probabilities $p^k_{ij} = \mathbb{P}\left(\widehat{Y}_{k+1} = y^{k+1}_j \mid \widehat{Y}_k = y^k_i\right)$.

Lemma 5.3.2. *The conditional probabilities p^k_{ij} are given by*

$$p^k_{ij} = F^k_i\big(y^{k+1}_{j+1/2}\big) - F^k_i\big(y^{k+1}_{j-1/2}\big). \tag{5.42}$$

And the probabilities p^{k+1}_j are given by

$$p^{k+1}_j = \sum_{i=1}^{N_{2,k}} p^k_i p^k_{ij}. \tag{5.43}$$

Proof. The

$$\begin{aligned} p_{ij}^k &= \mathbb{P}\left(\widehat{Y}_{k+1} = y_j^{k+1} \mid \widehat{Y}_k = y_i^k\right) \\ &= \mathbb{P}\left(\mathcal{M}_{\widetilde{b},\widetilde{\sigma}}\left(t_k, \widehat{Y}_k, Z_{k+1}^2\right) \in C_j\left(\Gamma_{N_{2,k+1}}^Y\right) \mid \widehat{Y}_k = y_i^k\right) \\ &= \mathbb{P}\left(\mathcal{M}_{\widetilde{b},\widetilde{\sigma}}\left(t_k, y_i^k, Z_{k+1}^2\right) \in C_j\left(\Gamma_{N_{2,k+1}}^Y\right)\right) \\ &= F_i^k\left(y_{j+1/2}^{k+1}\right) - F_i^k\left(y_{j-1/2}^{k+1}\right) \end{aligned}$$

and

$$\begin{aligned} p_j^{k+1} &= \mathbb{P}\left(\widehat{Y}_{k+1} = y_j^{k+1}\right) = \sum_{i=1}^{N_{2,k}} \mathbb{P}\left(\widehat{Y}_{k+1} = y_j^{k+1} \mid \widehat{Y}_k = y_i^k\right) \mathbb{P}\left(\widehat{Y}_k = y_i^k\right) \\ &= \sum_{i=1}^{N_{2,k}} p_i^k \, p_{ij}^k. \end{aligned}$$

□

As an illustration, we display in Figure 5.10 the rescaled grids obtained after recursive quantization of the boosted-volatility, where $\widehat{v}_k = \mathrm{e}^{-\kappa t_k} \widehat{Y}_k$ and $(\widehat{Y}_k)_{k=1,\dots,n}$ are the quantizers built using the fast recursive quantization approach.

5.3.2.2 Quantizing the asset (a one-dimensional case again)

Now, using the fact that $(Y_t)_t$ has already been quantized and the Euler-Maruyama scheme of $(X_t)_t$, as defined (5.32), we define the Markov quantized scheme

$$\widetilde{X}_{t_{k+1}} = \mathcal{E}_{b,\sigma}\left(t_k, \widehat{X}_{t_k}, \widehat{Y}_{t_k}, Z_{k+1}^1\right), \qquad \widehat{X}_{t_{k+1}} = \mathrm{Proj}_{\Gamma_{N_{1,k+1}}^X}\left(\widetilde{X}_{t_{k+1}}\right) \tag{5.44}$$

where the projection operator $\mathrm{Proj}_{\Gamma_{N_{1,k+1}}^X}(\cdot)$ is defined in (5.105), $\Gamma_{N_{1,k+1}}^X$ is the optimal $N_{1,k+1}$-quantizer of $\widetilde{X}_{t_{k+1}}$ and $Z_{k+1}^1 \sim \mathcal{N}(0,1)$. Again, in order to simplify the notations, $\widetilde{X}_{t_k}$ and $\widehat{X}_{t_k}$ are denoted in what follows by $\widetilde{X}_k$ and $\widehat{X}_k$.

Note that we are still in an one-dimensional case, hence we can apply the same methodology as developed in Appendix 5.D and build recursively the quantization $\left(\widehat{X}_k\right)_{k=0,\dots,n}$ as detailed above, where the $N_{1,k}$-tuple are defined by $x_{1:N_{1,k}}^k = \left(x_1^k, \dots, x_{N_{1,k}}^k\right)$. Replacing X by $\widetilde{X}_k$ in

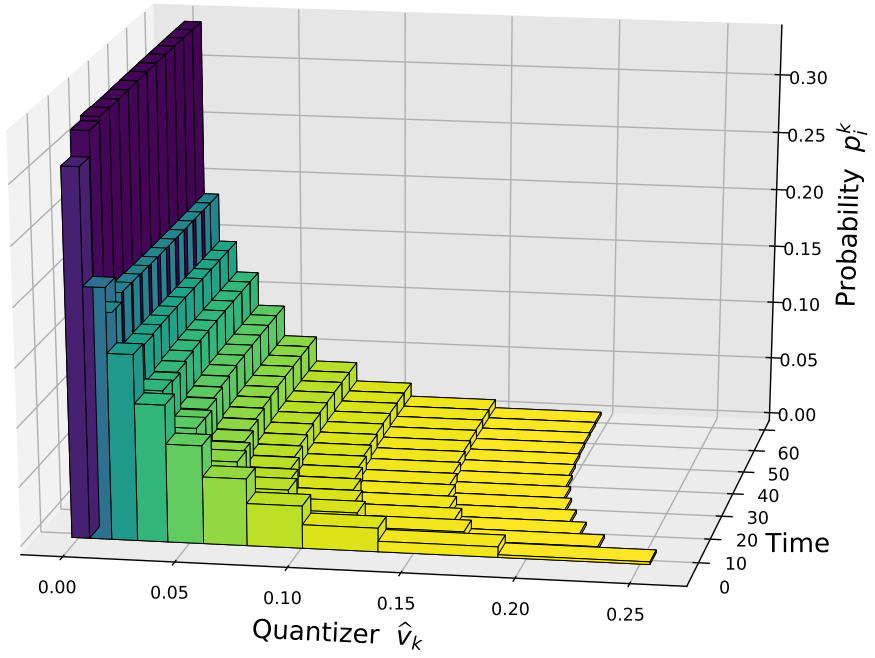


Fig. 5.10 *Rescaled Recursive quantization of the boosted-volatility process with its associated weights from* $t = 0$ *to* $t = 60$ *days with a time step of* 5 *days with grids of size* $N = 10$. *The recursive quantization methodology is applied to* $\widehat{Y}_k$ *and then we display the rescaled volatility* $\widehat{v}_k = \mathrm{e}^{-\kappa t_k}\,\widehat{Y}_k$.

(5.109) yield

$$
\begin{aligned}
x_{j_1}^{k+1} &= \frac{\mathbb{E}\left[\mathcal{E}_{b,\sigma}\left(t_k,\widehat{X}_{t_k},\widehat{Y}_{t_k},Z_{k+1}^1\right)\mathbb{1}_{\mathcal{E}_{b,\sigma}\left(t_k,\widehat{X}_{t_k},\widehat{Y}_{t_k},Z_{k+1}^1\right)\in C_{j_1}\left(\Gamma_{N_{1,k+1}}^X\right)}\right]}{\mathbb{P}\left(\mathcal{E}_{b,\sigma}\left(t_k,\widehat{X}_{t_k},\widehat{Y}_{t_k},Z_{k+1}^1\right)\in C_{j_1}\left(\Gamma_{N_{1,k+1}}^X\right)\right)} \\
&= \frac{\displaystyle\sum_{i_1=1}^{N_{1,k}}\sum_{i_2=1}^{N_{2,k}}\mathbb{E}\left[\mathcal{E}_{b,\sigma}\left(t_k,x_{i_1}^k,y_{i_2}^k,Z_{k+1}^1\right)\mathbb{1}_{\mathcal{E}_{b,\sigma}\left(t_k,x_{i_1}^k,y_{i_2}^k,Z_{k+1}^1\right)\in C_{j_1}\left(\Gamma_{N_{1,k+1}}^X\right)}\right]p_{(i_1,i_2)}^k}{\displaystyle\sum_{i_1=1}^{N_{1,k}}\sum_{i_2=1}^{N_{2,k}}\mathbb{P}\left(\mathcal{E}_{b,\sigma}\left(t_k,x_{i_1}^k,y_{i_2}^k,Z_{k+1}^1\right)\in C_{j_1}\left(\Gamma_{N_{1,k+1}}^X\right)\right)p_{(i_1,i_2)}^k} \\
&= \frac{\displaystyle\sum_{i_1=1}^{N_{1,k}}\sum_{i_2=1}^{N_{2,k}}\left(K_{(i_1,i_2)}^k\left(x_{j_1+1/2}^{k+1}\right)-K_{(i_1,i_2)}^k\left(x_{j_1-1/2}^{k+1}\right)\right)p_{(i_1,i_2)}^k}{\displaystyle\sum_{i_1=1}^{N_{1,k}}\sum_{i_2=1}^{N_{2,k}}\left(F_{(i_1,i_2)}^k\left(x_{j_1+1/2}^{k+1}\right)-F_{(i_1,i_2)}^k\left(x_{j_1-1/2}^{k+1}\right)\right)p_{(i_1,i_2)}^k}
\end{aligned}
\tag{5.45}
$$

where $p^k_{(i_1,i_2)} = \mathbb{P}\left(\widehat{X}_k = x^k_{i_1}, \widehat{Y}_k = y^k_{i_2}\right)$ and $F^k_{(i_1,i_2)}$ and $K^k_{(i_1,i_2)}$ are the cumulative distribution function and the first partial moment function of the normal distribution $\mu^k_{(i_1,i_2)} + Z^1_{k+1}\sigma^k_{(i_1,i_2)}$ and they are defined by

$$\begin{aligned} F^k_{(i_1,i_2)}(x) &= F_Z\left(\frac{x-\mu^k_{(i_1,i_2)}}{\sigma^k_{(i_1,i_2)}}\right) \\ K^k_{(i_1,i_2)}(x) &= \mu^k_{(i_1,i_2)}F_Z\left(\frac{x-\mu^k_{(i_1,i_2)}}{\sigma^k_{(i_1,i_2)}}\right) + \sigma^k_{(i_1,i_2)}K_Z\left(\frac{x-\mu^k_{(i_1,i_2)}}{\sigma^k_{(i_1,i_2)}}\right) \end{aligned} \tag{5.46}$$

with

$$\mu^k_{(i_1,i_2)} = x^k_{i_1} + b(t_k, x^k_{i_1}, y^k_{i_2})h \qquad \text{and} \qquad \sigma^k_{(i_1,i_2)} = \sigma(t_k, x^k_{i_1}, y^k_{i_2})\sqrt{h} \tag{5.47}$$

and F_Z and K_Z are the cumulative distribution function and the first partial moment of the standard normal distribution.

Finally, we apply the Lloyd method defined in Appendix (5.112) with F_X and K_X defined by

$$F_X(x) = \sum_{i_1=1}^{N_{1,k}}\sum_{i_2=1}^{N_{2,k}} p^k_{(i_1,i_2)}\, F^k_{(i_1,i_2)}(x) \qquad \text{and} \qquad K_X(x) = \sum_{i_1=1}^{N_{1,k}}\sum_{i_2=1}^{N_{2,k}} p^k_{(i_1,i_2)}\, K^k_{(i_1,i_2)}(x). \tag{5.48}$$

The sensitive part concerns the computation of the joint probabilities $p^k_{(i_1,i_2)}$. Indeed, they are needed at each step in order to be able to design recursively the quantization tree.

Lemma 5.3.3. *The joint probabilities $p^k_{(i_1,i_2)}$ are given by the following forward induction*

$$p^{k+1}_{(j_1,j_2)} = \sum_{i=1}^{N_{1,k}}\sum_{j=1}^{N_{2,k}} p^k_{(i_1,i_2)}\, \mathbb{P}\left(\widehat{X}_{k+1} = x^{k+1}_{j_1}, \widehat{Y}_{k+1} = y^{k+1}_{j_2} \mid \widehat{X}_k = x^k_{i_1}, \widehat{Y}_k = y^k_{i_2}\right) \tag{5.49}$$

where the joint conditional probabilities $\mathbb{P}\left(\widehat{X}_{k+1} = x^{k+1}_{j_1}, \widehat{Y}_{k+1} = y^{k+1}_{j_2} \mid \widehat{X}_k = x^k_{i_1}, \widehat{Y}_k = y^k_{i_2}\right)$ are given by the formulas below, depending on the correlation

- *if* $\mathbb{C}\mathrm{orr}(Z^1_{k+1}, Z^2_{k+1}) = \rho = 0$

$$\mathbb{P}\left(\widehat{X}_{k+1} = x^{k+1}_{j_1}, \widehat{Y}_{k+1} = y^{k+1}_{j_2} \mid \widehat{X}_k = x^k_{i_1}, \widehat{Y}_k = y^k_{i_2}\right) = p^k_{i_2 j_2}\left[\mathcal{N}\left(x^k_{i_1,i_2,j_1,+}\right) - \mathcal{N}\left(x^k_{i_1,i_2,j_1,-}\right)\right], \tag{5.50}$$

where $p^k_{i_2 j_2}$ is defined in (5.42) and

$$x^k_{i_1,i_2,j_1,-} = \frac{x^{k+1}_{j_1-1/2} - \mu^k_{(i_1,i_2)}}{\sigma^k_{(i_1,i_2)}}, \qquad x^k_{i_1,i_2,j_1,+} = \frac{x^{k+1}_{j_1+1/2} - \mu^k_{(i_1,i_2)}}{\sigma^k_{(i_1,i_2)}}, \tag{5.51}$$

with $\mu^k_{(i_1,i_2)}$ and $\sigma^k_{(i_1,i_2)}$ defined in (5.47).

- *if* $\mathbb{C}\mathrm{orr}(Z^1_{k+1}, Z^2_{k+1}) = \rho \neq 0$

$$\begin{aligned}
&\mathbb{P}\big(\widehat{X}_{k+1} = x^{k+1}_{j_1}, \widehat{Y}_{k+1} = y^{k+1}_{j_2} \mid \widehat{X}_k = x^k_{i_1}, \widehat{Y}_k = y^k_{i_2}\big) \\
&\quad = \mathbb{P}\left(Z^1_{k+1} \in \big(x^k_{i_1,i_2,j_1,-}, x^k_{i_1,i_2,j_1,+}\big], Z^2_{k+1} \in \left(\sqrt{y^k_{i_2,j_2,-}} - \lambda^k_{i_2}, \sqrt{y^k_{i_2,j_2,+}} - \lambda^k_{i_2}\right]\right) \\
&\qquad + \mathbb{P}\left(Z^1_{k+1} \in \big(x^k_{i_1,i_2,j_1,-}, x^k_{i_1,i_2,j_1,+}\big], Z^2_{k+1} \in \left[-\sqrt{y^k_{i_2,j_2,+}} - \lambda^k_{i_2}, -\sqrt{y^k_{i_2,j_2,-}} - \lambda^k_{i_2}\right)\right)
\end{aligned} \tag{5.52}$$

where

$$y^k_{i_2,j_2,-} = 0 \vee \frac{y^{k+1}_{j_2-1/2} - \mu^k_{i_2}}{\kappa^k_{i_2}}, \qquad y^k_{i_2,j_2,+} = 0 \vee \frac{y^{k+1}_{j_2+1/2} - \mu^k_{i_2}}{\kappa^k_{i_2}}, \tag{5.53}$$

with $\mu^k_{i_2}$, $\kappa^k_{i_2}$ *and* $\lambda^k_{i_2}$ *defined in* (5.39).

Remark. The probability in the right hand side of (5.52) can be computed using the cumulative distribution function of a correlated bivariate normal distribution[2]. Indeed, let

$$F_\rho(x_1, x_2) = \mathbb{P}(X_1 \leqslant x_1, X_2 \leqslant x_2)$$

the cumulative distribution function of the correlated centered Gaussian vector (X_1, X_2) with unit variance and correlation ρ, we have

$$\mathbb{P}\left(X_1 \in [a, b], X_2 \in [c, d]\right) = F_\rho(b, d) - F_\rho(b, c) - F_\rho(a, d) + F_\rho(a, c) \tag{5.54}$$

with $a, c \geqslant -\infty$ and $b, d \leqslant +\infty$.

Proof.

$$\begin{aligned}
p^{k+1}_{(j_1,j_2)} &= \mathbb{P}\left(\widehat{X}_{k+1} = x^{k+1}_{j_1}, \widehat{Y}_{k+1} = y^{k+1}_{j_2}\right) \\
&= \sum_{i=1}^{N_{1,k}} \sum_{j=1}^{N_{2,k}} \mathbb{P}\left(\widehat{X}_{k+1} = x^{k+1}_{j_1}, \widehat{Y}_{k+1} = y^{k+1}_{j_2} \mid \widehat{X}_k = x^k_{i_1}, \widehat{Y}_k = y^k_{i_2}\right) \mathbb{P}\left(\widehat{X}_k = x^k_{i_1}, \widehat{Y}_k = y^k_{i_2}\right) \\
&= \sum_{i=1}^{N_{1,k}} \sum_{j=1}^{N_{2,k}} p^k_{(i_1,i_2)} \, \mathbb{P}\left(\widehat{X}_{k+1} = x^{k+1}_{j_1}, \widehat{Y}_{k+1} = y^{k+1}_{j_2} \mid \widehat{X}_k = x^k_{i_1}, \widehat{Y}_k = y^k_{i_2}\right).
\end{aligned}$$

[2] C++ implementation of the upper right tail of a bivariate normal distribution can be found in John Burkardt's website https://people.sc.fsu.edu/~jburkardt/cpp_src/toms462/toms462.html.

- if $\mathbb{C}\mathrm{orr}(Z^1_{k+1}, Z^2_{k+1}) = \rho = 0$

$$\begin{aligned}
&\mathbb{P}\left(\widehat{X}_{k+1} = x^{k+1}_{j_1}, \widehat{Y}_{k+1} = y^{k+1}_{j_2} \mid \widehat{X}_k = x^k_{i_1}, \widehat{Y}_k = y^k_{i_2}\right) \\
&\qquad = p^k_{i_2 j_2}\, \mathbb{P}\left(\widehat{X}_{k+1} = x^{k+1}_{j_1} \mid \widehat{X}_k = x^k_{i_1}, \widehat{Y}_k = y^k_{i_2}\right) \\
&\qquad = p^k_{i_2 j_2}\, \mathbb{P}\left(\bar{X}_{k+1} \in \left(x^{k+1}_{j_1-1/2}, x^{k+1}_{j_1+1/2}\right] \mid \widehat{X}_k = x^k_{i_1}, \widehat{Y}_k = y^k_{i_2}\right) \\
&\qquad = p^k_{i_2 j_2}\, \mathbb{P}\left(\mathcal{E}_{b,\sigma}\left(t_k, x^k_{i_1}, y^k_{i_2}, Z^1_{k+1}\right) \in \left(x^{k+1}_{j_1-1/2}, x^{k+1}_{j_1+1/2}\right]\right) \\
&\qquad = p^k_{i_2 j_2}\left[\mathcal{N}\left(x^k_{i_1,i_2,j_1,+}\right) - \mathcal{N}\left(x^k_{i_1,i_2,j_1,-}\right)\right],
\end{aligned}$$

- if $\mathbb{C}\mathrm{orr}(Z^1_{k+1}, Z^2_{k+1}) = \rho \neq 0$

$$\begin{aligned}
&\mathbb{P}\big(\widehat{X}_{k+1} = x^{k+1}_{j_1}, \widehat{Y}_{k+1} = y^{k+1}_{j_2} \mid \widehat{X}_k = x^k_{i_1}, \widehat{Y}_k = y^k_{i_2}\big) \\
&\quad = \mathbb{P}\left(\mathcal{E}_{b,\sigma}\big(t_k, x^k_{i_1}, y^k_{i_2}, Z^1_{k+1}\big) \in \left(x^{k+1}_{j_1-1/2}, x^{k+1}_{j_1+1/2}\right], \mathcal{M}_{\tilde{b},\tilde{\sigma}}\big(t_k, y^k_{i_2}, Z^2_{k+1}\big) \in \left(y^{k+1}_{j_2-1/2}, y^{k+1}_{j_2+1/2}\right]\right) \\
&\quad = \mathbb{P}\left(\mu^k_{(i_1,i_2)} + \sigma^k_{(i_1,i_2)} Z^1_{k+1} \in \left(x^{k+1}_{j_1-1/2}, x^{k+1}_{j_1+1/2}\right], \mu^k_{i_2} + \kappa^k_{i_2}(Z^2_{k+1} + \lambda^k_{i_2})^2 \in \left(y^{k+1}_{j_2-1/2}, y^{k+1}_{j_2+1/2}\right]\right) \\
&\quad = \mathbb{P}\left(Z^1_{k+1} \in \left(x^k_{i_1,i_2,j_1,-}, x^k_{i_1,i_2,j_1,+}\right], (Z^2_{k+1} + \lambda^k_{i_2})^2 \in \left(y^k_{i_2,j_2,-}, y^k_{i_2,j_2,+}\right]\right) \\
&\quad = \mathbb{P}\left(Z^1_{k+1} \in \left(x^k_{i_1,i_2,j_1,-}, x^k_{i_1,i_2,j_1,+}\right], Z^2_{k+1} \in \left(\sqrt{y^k_{i_2,j_2,-}} - \lambda^k_{i_2}, \sqrt{y^k_{i_2,j_2,+}} - \lambda^k_{i_2}\right]\right) \\
&\qquad + \mathbb{P}\left(Z^1_{k+1} \in \left(x^k_{i_1,i_2,j_1,-}, x^k_{i_1,i_2,j_1,+}\right], Z^2_{k+1} \in \left[-\sqrt{y^k_{i_2,j_2,+}} - \lambda^k_{i_2}, -\sqrt{y^k_{i_2,j_2,-}} - \lambda^k_{i_2}\right)\right).
\end{aligned}$$

□

Remark. Another possibility for the quantization of the Stationary Heston model could be to use optimal quantizers for the volatility at each date t_k in place of using recursive quantization. Indeed, the volatility $(v_t)_t$ being stationary and the fact that we required the volatility to start at time 0 from the invariant measure, we could use the grid of the optimal quantization $\widehat{v}_0$ of size N of the stationary measure with its associated weights for every dates, hence setting $\widehat{v}_k = \widehat{v}_0$. We need as well the transitions from time t_k to t_{k+1} defined by

$$\mathbb{P}\left(\widehat{v}_{k+1} = v^{k+1}_{j_2} \mid \widehat{v}_k = v^k_{i_2}\right). \tag{5.55}$$

These probabilities can be computed using the conditional law of the CIR process described in [CIR05; And07], which is a non-central chi-square distribution. Then, we would build the recursive quantizer of the log-asset at date $\widehat{X}_{k+1}$ with the standard methodology of recursive quantization using the already built quantizers of the volatility $\widehat{v}_k$ and the log-asset $\widehat{X}_k$ at time t_k, i.e.

$$\widetilde{X}_{k+1} = \mathcal{E}_{b,\sigma}\big(t_k, \widehat{X}_k, \widehat{v}_k, Z^1_{k+1}\big) \quad \text{and} \quad \widehat{X}_{k+1} = \mathrm{Proj}_{\Gamma^X_{N_{1,k+1}}}\left(\widetilde{X}_{k+1}\right) \tag{5.56}$$

where, this time, the Euler scheme is not defined in function of the boosted-volatility but directly in function of the volatility and is given by

$$\mathcal{E}_{b,\sigma}(t,x,v,z) = x + h\Big(r - q - \frac{v}{2}\Big) + \sqrt{v}\sqrt{h}z. \tag{5.57}$$

However, the difficulties with this approach come from the computation of the couple transitions

$$\mathbb{P}\left(\widehat{X}_{k+1} = x_{j_1}^{k+1}, \widehat{v}_{k+1} = v_{j_2}^{k+1} \mid \widehat{X}_k = x_{i_1}^k, \widehat{v}_k = v_{i_2}^k\right). \tag{5.58}$$

Indeed, these probability weights would not be as straightforward to compute as the methodology we adopt in this paper, namely using time-discretization schemes for both components. Our approach allows us to express the conditional probability of the couple as the probability that a correlated bi-variate Gaussian vector lies in a rectangle domain and this can be easily be computed numerically.

5.3.2.3 About the L^2-error

In this part, we study the L^2-error induced by the product recursive quantization approximation $\widehat{U}_k = (\widehat{X}_k, \widehat{Y}_k)$ of $\bar{U}_k = (\bar{X}_k, \bar{Y}_k)$, the time-discretized processes defined in (5.26) and (5.31) by

$$\bar{U}_k = F_{k-1}(\bar{U}_{k-1}, Z_k) \tag{5.59}$$

where $Z_k = (Z_k^1, Z_k^2)$ is a standardized correlated Gaussian vector and the hybrid discretization scheme $F_k(u, Z)$ is given by

$$F_k(u,Z) = \begin{pmatrix} \mathcal{E}_{b,\sigma}(t_k, x, y, Z_{k+1}^1) \\ \mathcal{M}_{\widetilde{b},\widetilde{\sigma}}(t_k, y, Z_{k+1}^2) \end{pmatrix}. \tag{5.60}$$

We recall the definition of the product recursive quantizer $\widehat{U}_k = (\widehat{X}_k, \widehat{Y}_k)$. Its first component $\widehat{X}_k$ is the projection of $\widetilde{X}_k$ onto $\Gamma_{N_{1,k}}^X$ and the second component $\widehat{Y}_k$ is the projection of $\widetilde{Y}_k$ onto $\Gamma_{N_{2,k}}^Y$, i.e.,

$$\widehat{X}_{k+1} = \mathrm{Proj}_{\Gamma_{N_{1,k+1}}^X}\left(\widetilde{X}_{k+1}\right) \quad \text{and} \quad \widehat{Y}_{k+1} = \mathrm{Proj}_{\Gamma_{N_{2,k+1}}^Y}\left(\widetilde{Y}_{k+1}\right) \tag{5.61}$$

where $\widetilde{X}_k$ and $\widetilde{Y}_k$ are defined in (5.36) and (5.44), respectively. Moreover, if we consider the couple $\widetilde{U}_k = (\widetilde{X}_k, \widetilde{Y}_k)$, using the above notations we have

$$\widetilde{U}_k = F_{k-1}(\widehat{U}_{k-1}, Z_k). \tag{5.62}$$

It has been shown in [FSP18; PS18b] that if, for all $k = 0, \dots, n-1$, the schemes $F_k(u, z)$ are Lipschitz in u, then there exists constants $j = 1, \dots, n$, $C_j < +\infty$ such that

$$\|\widehat{U}_k - \bar{U}_k\|_2 \leqslant \sum_{j=1}^{k} C_j \big(N_{1,j} \times N_{2,j}\big)^{-1/2} \tag{5.63}$$

where $\widehat{U}_k$ and $\bar{U}_k$ are the processes defined in (5.61) and (5.62). The proof of this result is based on the extension of Pierce's lemma to the case of product quantization (see Lemma 2.3 in [PS18b]).

In our case, the diffusion of the boosted volatility in the CIR model does not have Lipschitz drift and volatility components, hence the above result from [FSP18; PS18b] does not apply in our context. Even if we can hope to obtain similar results by applying the same kind of arguments, the results we obtain have to considered carefully. Indeed, when we take the limit in $n \to +\infty$, the number of time-step, the error upper-bound term goes to infinity. However, in practice, we consider $h = kT/n$ fixed and then study the behavior of $\widehat{U}_k$ in function of $N_{1,j}$ and $N_{2,j}$ for $j \geqslant k$. The proof of the following proposition is given in Appendix 5.C.

Proposition 5.3.4. *Let b, σ, $\widetilde{b}$ and $\widetilde{\sigma}$, defined by (5.27) and (5.33), the coefficients of the* log*-asset and the boosted-volatility of the Heston model. Let, for every $k = 0, \dots, n$, $\widehat{U}_k$ the hybrid recursive product quantizer at level $N_{1,k} \times N_{2,k}$ of $\bar{U}_k$. Then, for every $k = 0, \dots, n$*

$$\|\widehat{U}_k - \bar{U}_k\|_2 \leqslant \sum_{j=0}^{k} \widetilde{A}_{j,k} \big(N_{1,j} \times N_{2,j}\big)^{-1/2} + B_k \sqrt{h} \tag{5.64}$$

where

$$\widetilde{A}_{j,k} = 2^{\frac{p-2}{2p}} C_p^2 A_{j,k} \left(2^{(\frac{p}{2}-1)j} \beta_p^j \|\widehat{U}_0\|_2^p + \alpha_p \frac{1 - 2^{(\frac{p}{2}-1)j} \beta_p^j}{1 - 2^{\frac{p}{2}-1} \beta_p} \right)^{1/p} \tag{5.65}$$

with

$$A_{j,k} = 2^{\frac{k-j}{2}} \, \mathrm{e}^{\frac{\sqrt{h}}{2}(k-j)} \quad \text{and} \quad B_k = C_T(h) \sum_{j=0}^{k-1} 2^{\frac{k-1-j}{2}} \, \mathrm{e}^{\frac{\sqrt{h}}{2}(k-1-j)} \tag{5.66}$$

where $\sum_{\varnothing} = 0$ by convention and $C_T(h) = O(1)$.

5.3.3 Backward algorithm for Bermudan and Barrier options

Bermudan Options A Bermudan option is a financial derivative product that gives the right to its owner to buy or sell (or to enter to, in the case of a swap) an underlying product with a given payoff $\psi_t(\cdot, \cdot)$ at predefined exercise dates $\{t_0, \cdots, t_n\}$. Its price, at time $t_0 = 0$, is given by

$$\sup_{\tau \in \{t_0, \cdots, t_n\}} \mathbb{E}\left[\mathrm{e}^{-r\tau} \, \psi_\tau(X_\tau, Y_\tau) \mid \mathcal{F}_{t_0} \right]$$

where X_t and Y_t are solutions to the system defined in (5.34).

In this part, we follow the numerical solution first introduced by [BPP05; BP03]. They proposed to solve discrete-time optimal stopping problems using a quantization tree of the risk factors X_t and Y_t.

Let $\mathcal{F}^{X,Y} = (\mathcal{F})_{0\leqslant k\leqslant n}$ the natural filtration of X and Y. Hence, we can define recursively the sequence of random variable L^p-integrable $(V_k)_{0\leqslant k\leqslant n}$

$$\begin{cases} V_n = \mathrm{e}^{-rt_n}\,\psi_n(X_n, Y_n), \\ V_k = \max\left(\mathrm{e}^{-rt_k}\,\psi_k(X_k, Y_k), \mathbb{E}[V_{k+1} \mid \mathcal{F}_k]\right), \qquad 0 \leqslant k \leqslant n-1 \end{cases} \tag{5.67}$$

called *Backward Dynamic Programming Principle*. Then

$$V_0 = \sup\left\{\mathbb{E}[\mathrm{e}^{-r\tau}\,\psi_\tau(X_\tau, Y_\tau) \mid \mathcal{F}_0], \tau \in \Theta_{0,n}\right\}$$

with $\Theta_{0,n}$ the set of all stopping times taking values in $\{t_0, \cdots, t_n\}$. The sequence $(V_k)_{0\leqslant k\leqslant n}$ is also known as the Snell envelope of the obstacle process $\left(\mathrm{e}^{-rt_k}\,\psi_k(X_k, Y_k)\right)_{0\leqslant k\leqslant n}$. In the end, $\mathbb{E}[V_0]$ is the quantity we are interested in. Indeed, $\mathbb{E}[V_0]$ is the price of the Bermudan option whose payoff is ψ_k and is exercisable at dates $\{t_1, \cdots, t_n\}$.

Following what was defined in (5.67), in order to compute $\mathbb{E}[V_0]$, we will need to use the previously defined quantizer of X_k and Y_k: $\widehat{X}_k$ and $\widehat{Y}_k$. Hence, for a given global budget $N = N_{1,0}N_{2,0} + \cdots + N_{1,n}N_{2,n}$, the total number of nodes of the tree by the couple $(\widehat{X}_k, \widehat{Y}_k)_{0\leqslant k\leqslant n}$, we can approximate the *Backward Dynamic Programming Principle* (5.67) by the following sequence involving the couple $(\widehat{X}_k, \widehat{Y}_k)_{0\leqslant k\leqslant n}$

$$\begin{cases} \widehat{V}_n = \mathrm{e}^{-rt_n}\,\psi_n(\widehat{X}_n, \widehat{Y}_n), \\ \widehat{V}_k = \max\left(\mathrm{e}^{-rt_k}\,\psi_k(\widehat{X}_k, \widehat{Y}_k), \mathbb{E}[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{Y}_k)]\right), \qquad k = 0, \ldots, n-1. \end{cases} \tag{5.68}$$

Remark. A direct consequence of choosing recursive Markovian Quantization to spatially discretize the problem is that the sequence $(\widehat{X}_k, \widehat{Y}_k)_{0\leqslant k\leqslant n}$ is Markovian. Hence $(\widehat{V}_k)_{0\leqslant k\leqslant n}$ defined in (5.68) obeying a *Backward Dynamic Programming Principle* is the Snell envelope of $\left(\mathrm{e}^{-rt_k}\,\psi_k(\widehat{X}_k, \widehat{Y}_k)\right)_{0\leqslant k\leqslant n}$. This is the main difference with the first approach of [BPP05; BP03], where in there case they only had a pseudo-Snell envelope of $\left(\mathrm{e}^{-rt_k}\,\psi_k(\widehat{X}_k, \widehat{Y}_k)\right)_{0\leqslant k\leqslant n}$.

Using the discrete feature of the quantizers, (5.68) can be rewritten

$$\begin{cases} \widehat{v}_n(x^n_{i_1}, y^n_{i_2}) = \mathrm{e}^{-rt_n}\,\psi_n(x^n_{i_1}, y^n_{i_2}), & \begin{matrix} i_1 = 1, \ldots, N_{1,n} \\ i_2 = 1, \ldots, N_{2,n} \end{matrix} \\ \widehat{v}_k(x^k_{i_1}, y^k_{i_2}) = \max\left(\mathrm{e}^{-rt_k}\,\psi_k(x^k_{i_1}, y^k_{i_2}), \displaystyle\sum_{j_1=1}^{N_{1,k+1}}\sum_{j_2=1}^{N_{2,k+1}} \pi^k_{(i_1,i_2),(j_1,j_2)}\,\widehat{v}_{k+1}(x^{k+1}_{j_1}, y^{k+1}_{j_2})\right), & \begin{matrix} k = 0, \ldots, n-1 \\ i_1 = 1, \ldots, N_{1,k} \\ i_2 = 1, \ldots, N_{2,k} \end{matrix} \end{cases} \tag{5.69}$$

where $\pi^k_{(i_1,i_2),(j_1,j_2)} = \mathbb{P}\left(\widehat{X}_{k+1} = x^{k+1}_{j_1}, \widehat{Y}_{k+1} = y^{k+1}_{j_2} \mid \widehat{X}_k = x^k_{i_1}, \widehat{Y}_k = y^k_{i_2}\right)$ is the conditional probability weight given in (5.52). Finally, the approximation of the price of the Bermudan option is given by

$$\mathbb{E}\left[\widehat{v}_0(x_0, \widehat{Y}_0)\right] = \sum_{i=1}^{N_{2,0}} p_i \, \widehat{v}_0(x_0, y^0_i) \tag{5.70}$$

with $p_i = \mathbb{P}\left(\widehat{Y}_0 = y^0_i\right)$ given by (5.18).

Barrier Options A Barrier option is a path-dependent financial product whose payoff at maturity date T depends on the value of the process X_T at time T and its maximum or minimum over the period $[0, T]$. More precisely, we are interested by options with the following types of payoff h

$$h = f(X_T)\,\mathbb{1}_{\{\sup_{t\in[0,T]} X_t \in I\}} \qquad or \qquad h = f(X_T)\,\mathbb{1}_{\{\inf_{t\in[0,T]} X_t \in I\}} \tag{5.71}$$

where I is an unbounded interval of $\mathbb{R}$, usually of the forme $(-\infty, L]$ or $[L, +\infty)$ (L is the barrier) and f can be any vanilla payoff function (Call, Put, Spread, Butterfly, ...).

In this part, we follow the methodology initiated in [Sag10] in the case of functional quantization. This work is based on the Brownian bridge method applied to the Euler-Maruyama scheme as described e.g. in [Pag18]. We generalize it to stochastic volatility models and product Markovian recursive quantization. X_t being discretized by an Euler-Maruyama scheme, yielding $\bar{X}_k$ with $k = 0, \ldots, n$, we can determine the law of $\max_{t\in[0,T]} \bar{X}_t$ and $\min_{t\in[0,T]} \bar{X}_t$ given the values $\bar{X}_k = x_k, \bar{Y}_k = y_k, k = 0, \ldots, n$

$$\mathcal{L}\left(\max_{t\in[0,T]} \bar{X}_t \mid \bar{X}_k = x_k, \bar{Y}_k = y_k, k = 0, \ldots, n\right) = \mathcal{L}\left(\max_{k=0,\ldots,n-1} \left(G^k_{(x_k,y_k),x_{k+1}}\right)^{-1}(U_k)\right) \tag{5.72}$$

and

$$\mathcal{L}\left(\min_{t\in[0,T]} \bar{X}_t \mid \bar{X}_k = x_k, \bar{Y}_k = y_k, k = 0, \ldots, n\right) = \mathcal{L}\left(\max_{k=0,\ldots,n-1} \left(F^k_{(x_k,y_k),x_{k+1}}\right)^{-1}(U_k)\right) \tag{5.73}$$

where $(U_k)_{k=0,\ldots,n-1}$ are i.i.d uniformly distributed random variables over the unit interval and $(G^k_{(x,y),z})^{-1}$ and $(F^k_{(x,y),z})^{-1}$ are the inverse of the conditional distribution functions $G^k_{(x,y),z}$ and $F^k_{(x,y),z}$ defined by

$$G^k_{(x,y),z}(u) = \left(1 - \mathrm{e}^{-2n\frac{(x-u)(z-u)}{T\sigma^2(t_k,x,y)}}\right)\mathbb{1}_{\{u\geqslant\max(x,z)\}} \tag{5.74}$$

and

$$F^k_{(x,y),z}(u) = 1 - \left(1 - \mathrm{e}^{-2n\frac{(x-u)(z-u)}{T\sigma^2(t_k,x,y)}}\right)\mathbb{1}_{\{u\leqslant\min(x,z)\}}\,. \tag{5.75}$$

Now, using the resulting representation formula for $\mathbb{E}\, f(\bar{X}_T, \max_{t\in[0,T]} \bar{X}_t)$ (see e.g. [Sag10; Pag18]), we have a new representation formula for the price of up-and-out options $\bar{P}_{UO}$ and

down-and-out options $\bar{P}_{DO}$

$$\bar{P}_{UO} = \mathrm{e}^{-rT}\,\mathbb{E}\left[f(\bar{X}_T)\,\mathbb{1}_{\sup_{t\in[0,T]}\bar{X}_t\leqslant L}\right] = \mathrm{e}^{-rT}\,\mathbb{E}\left[f(\bar{X}_T)\prod_{k=0}^{n-1} G^k_{(\bar{X}_k,\bar{Y}_k),\bar{X}_{k+1}}(L)\right] \tag{5.76}$$

and

$$\bar{P}_{DO} = \mathrm{e}^{-rT}\,\mathbb{E}\left[f(\bar{X}_T)\,\mathbb{1}_{\inf_{t\in[0,T]}\bar{X}_t\geqslant L}\right] = \mathrm{e}^{-rT}\,\mathbb{E}\left[f(\bar{X}_T)\prod_{k=0}^{n-1}\left(1-F^k_{(\bar{X}_k,\bar{Y}_k),\bar{X}_{k+1}}(L)\right)\right] \tag{5.77}$$

where L is the barrier.

Finally, replace $\bar{X}_k$ and $\bar{Y}_k$ by $\widehat{X}_k$ and $\widehat{Y}_k$ and apply the recursive algorithm in order to approximate $\bar{P}_{UO}$ or $\bar{P}_{DO}$ by $\mathbb{E}[\widehat{V}_0]$ or equivalently $\mathbb{E}[\widehat{v}_0(x_0,\widehat{Y}_0)]$

$$\begin{cases} \widehat{V}_n = \mathrm{e}^{-rT}\,f(\widehat{X}_n), \\ \widehat{V}_k = \mathbb{E}\left[g_k(\widehat{X}_k,\widehat{Y}_k,\widehat{X}_{k+1})\widehat{V}_{k+1}\mid(\widehat{X}_k,\widehat{Y}_k)\right], \qquad 0\leqslant k\leqslant n-1 \end{cases} \tag{5.78}$$

that can be rewritten

$$\begin{cases} \widehat{v}_n(x^n_{i_1},y^n_{i_2}) = \mathrm{e}^{-rT}\,f(x^n_i), & \begin{matrix} i=1,\ldots,N_{1,n} \\ j=1,\ldots,N_{2,n} \end{matrix} \\ \widehat{v}_k(x^k_{i_1},y^n_{i_2}) = \displaystyle\sum_{j_1=1}^{N_{1,k+1}}\sum_{j_2=1}^{N_{2,k+1}} \pi^k_{(i_1,i_2),(j_1,j_2)}\widehat{v}_{k+1}(x^{k+1}_{j_1},y^{k+1}_{j_2})g_k(x^k_{i_1},y^k_{i_2},x^{k+1}_{j_1}), & \begin{matrix} k=0,\ldots,n-1 \\ i=1,\ldots,N_{1,k} \\ j=1,\ldots,N_{2,k} \end{matrix} \end{cases} \tag{5.79}$$

with $\pi^k_{(i_1,i_2),(j_1,j_2)} = \mathbb{P}\left(\widehat{X}_{k+1}=x^{k+1}_{j_1},\widehat{Y}_{k+1}=y^{k+1}_{j_2}\mid\widehat{X}_k=x^k_{i_1},\widehat{Y}_k=y^k_{i_2}\right)$ the conditional probabilities given in (5.52) and $g_k(x,y,z)$ is either equal to $G^k_{(x,y),z}(L)$ or $1-F^k_{(x,y),z}(L)$ depending on the option type. Finally, the approximation of the price of the barrier option is given by

$$\mathbb{E}[\widehat{V}_0] = \mathbb{E}\left[\widehat{v}_0(x_0,\widehat{Y}_0)\right] = \sum_{i=1}^{N_{2,0}} p_i\,\widehat{v}_0(x_0,y^0_i) \tag{5.80}$$

with $p_i = \mathbb{P}\left(\widehat{Y}_0 = y^0_i\right)$ given by (5.18).

5.3.4 Numerical illustrations

In this part, we deal with numerical experiments in the Stationary Heston model. We will apply the methodology based on hybrid product recursive quantization to the pricing of European, Bermudan and Barrier options. For the model parameters, we consider the parameters given in Table 5.2 obtained after the penalized calibration procedure and instead of considering the market value for S_0, we take $S_0 = 100$ in order to get prices of an order we are used to. For the size of the quantization grids, we consider grids of constant size for all time-steps: for all

$k = 0, \dots, n$, we take $N_{1,k} = N_1$ and $N_{2,k} = N_2$ where n is the number of time steps. During the numerical tests, we vary the tuple values (n, N_1, N_2).

All the numerical tests have been carried out in C++ on a laptop with a 2,4 GHz 8-Core Intel Core i9 CPU. The computations of the transition probabilities are parallelized on the CPU.

European options First, we compare, in Table 5.3, the price of European options with maturity $t_n = T = 0.5$ (6 months) computed using the quantization tree to the benchmark price computed using the methodology based on the quadrature formula (the quadrature formula with Laguerre polynomials) explained in Section 5.2. In place of using the backward algorithm (5.68) (without the function max) for computing the expectation at the expiry date, we use the weights $p^k_{(i_1,i_2)}$ defined in (5.49) and built by forward induction, in order to compute

$$\mathbb{E}\left[\mathrm{e}^{-rt_n}\,\psi_n(\widehat{X}_n, \widehat{Y}_n)\right] = \mathrm{e}^{-rt_n} \sum_{i_1=1}^{N_{1,n}} \sum_{i_2=1}^{N_{2,n}} \psi_n(x^n_{i_1}, y^n_{i_2}). \tag{5.81}$$

We give, in parenthesis, the relative error induced by the quantization-based approximation. We compare the behavior of the pricers with different size of grids and numbers of discretization steps. We notice that the main part of the error is explained by the size of the time-step n.

			(N_1, N_2)			
	K	Benchmark	$(20, 5)$	$(50, 10)$	$(100, 10)$	$(150, 10)$
Call	80	20.17	19.68 (2.46%)	19.99 (0.92%)	20.04 (0.64%)	20.06 (0.57%)
	85	15.56	14.97 (3.75%)	15.35 (1.31%)	15.42 (0.89%)	15.43 (0.79%)
	90	11.24	10.60 (5.68%)	11.03 (1.84%)	11.10 (1.18%)	11.12 (1.02%)
	95	7.383	6.781 (8.14%)	7.202 (2.44%)	7.286 (1.30%)	7.306 (1.03%)
	100	4.196	3.727 (11.1%)	4.081 (2.73%)	4.173 (0.54%)	4.194 (0.04%)
Put	100	4.469	4.160 (6.90%)	4.396 (1.61%)	4.459 (0.22%)	4.472 (0.08%)
	105	7.171	7.034 (1.91%)	7.178 (0.09%)	7.244 (1.01%)	7.257 (1.19%)
	110	10.86	10.84 (0.18%)	10.91 (0.46%)	10.97 (1.02%)	10.98 (1.11%)
	115	15.38	15.43 (0.33%)	15.40 (0.12%)	15.43 (0.37%)	15.44 (0.41%)
	120	20.30	20.43 (0.60%)	20.31 (0.02%)	20.29 (0.05%)	20.29 (0.04%)
	Time		2.6s	39s	192s	480s

Table 5.3 *Comparison between European options prices, with maturity $T = 0.5$ (6 months), given by quantization and the benchmark, in function of the strike K and (N_1, N_2) where we set $n = 180$.*

			n			
	K	Benchmark	30	60	90	180
Call	80	20.17	20.00 (0.83%)	20.03 (0.70%)	20.03 (0.72%)	19.99 (0.92%)
	85	15.56	15.33 (1.47%)	15.38 (1.11%)	15.39 (1.07%)	15.35 (1.31%)
	90	11.24	10.94 (2.60%)	11.04 (1.78%)	11.05 (1.63%)	11.03 (1.84%)
	95	7.383	7.045 (4.57%)	7.170 (2.87%)	7.203 (2.43%)	7.202 (2.44%)
	100	4.196	3.879 (7.55%)	4.016 (4.29%)	4.057 (3.31%)	4.081 (2.73%)
Put	100	4.469	4.161 (6.89%)	4.306 (3.64%)	4.354 (2.56%)	4.396 (1.61%)
	105	7.171	6.972 (2.77%)	7.081 (1.25%)	7.125 (0.64%)	7.178 (0.09%)
	110	10.86	10.81 (0.44%)	10.85 (0.05%)	10.87 (0.12%)	10.91 (0.46%)
	115	15.38	15.39 (0.06%)	15.38 (0.04%)	15.39 (0.08%)	15.40 (0.12%)
	120	20.30	20.29 (0.08%)	20.29 (0.09%)	20.29 (0.06%)	20.31 (0.02%)
	Time		9s	16s	24s	42s

Table 5.4 *Comparison between European options prices, with maturity $T = 0.5$ (6 months), given by quantization and the benchmark, in function of the strike K and of the size n where we set $(N_1, N_2) = (50, 10)$.*

Bermudan options Then, in Figure 5.11, we display the prices of monthly exercisable Bermudan options with maturity $T = 0.5$ (6 months) for Call and Put of strikes $K = 100$. The prices are computed by quantization and we compare the behavior of the pricer for different choices of time-step n and sizes of the asset grids N_1 where we set $N_2 = 10$. Again, we notice that the choice of n has a high impact on the price given by quantization compared to the choice of the grid size.

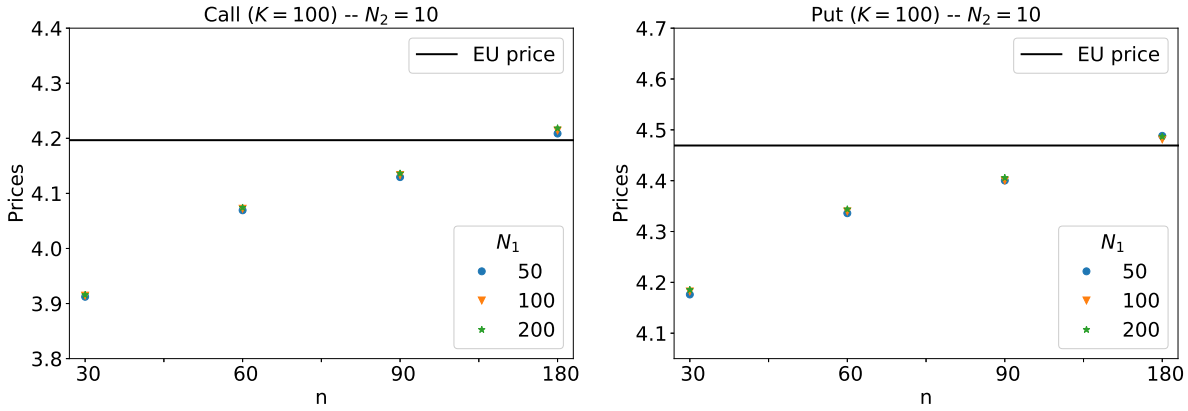


Fig. 5.11 *Prices of Bermudan options in the stationary Heston model given by product hybrid recursive quantization with fixed value $N_2 = 10$.*

Barrier options Finally, in Figure 5.12, we display the prices of an up-and-out Barrier option with strike $K = 100$, maturity $T = 0.5$ (6 months), barrier $L = 115$ and $N_2 = 10$ computed with quantization. Again, we can notice the impact of n on the approximated price.

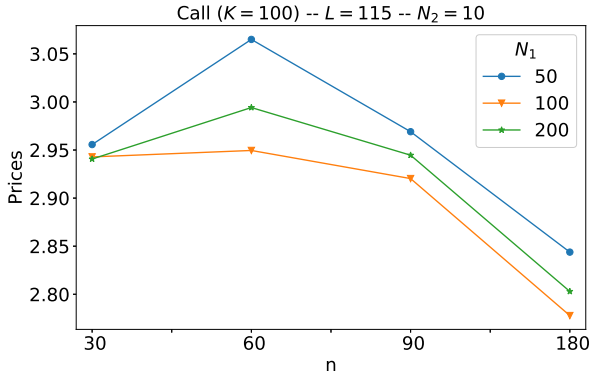


Fig. 5.12 *Prices of Barrier options with strike* $K = 100$ *in the stationary Heston model given by product hybrid recursive quantization with fixed value* $N_2 = 10$.

Appendix 5.A Discretization scheme for the volatility preserving the positivity

We recall the dynamics of the volatility

$$dv_t = \kappa(\theta - v_t)dt + \xi\sqrt{v_t}d\widetilde{W}_t$$

with $\kappa > 0$, $\theta > 0$ and $\xi > 0$. In this section, we discuss the choice of the discretization scheme under the Feller condition, which ensures the positivity of the process.

Euler-Maruyama scheme. Discretizing the volatility using an Euler-Maruyama scheme

$$\bar{v}_{t_{k+1}} = \bar{v}_{t_k} + \kappa(\theta - \bar{v}_{t_k})h + \xi\sqrt{\bar{v}_{t_k}}\sqrt{h}\, Z_{k+1}^2$$

with $t_k = kh$, $h = T/n$ and $Z_{k+1}^2 = (\widetilde{W}_{t_{k+1}} - \widetilde{W}_{t_k})/\sqrt{h}$ may look natural. However, such a scheme clearly does not preserve positivity of the process even if the Feller condition is fulfilled since

$$\mathbb{P}\left(\bar{v}_{t_1} < 0\right) = \mathbb{P}\left(Z < \frac{-v_0 - \kappa(\theta - v_0)h}{\xi\sqrt{v_0}\sqrt{h}}\right) > 0$$

with $Z \sim \mathcal{N}(0,1)$. This suggests to introduce the Milstein scheme which is quite tractable in one dimension in absence of Lévy areas.

Milstein scheme. The Milstein scheme of the stochastic volatility is given by

$$\bar{v}_{t_{k+1}} = \mathcal{M}_{b,\sigma}\big(t_k, \bar{v}_{t_{k+1}}, Z_{k+1}^2\big)$$

where (see (5.28))

$$\mathcal{M}_{b,\sigma}(t,x,z) = x - \frac{\sigma(x)}{2\sigma_x'(x)} + h\left(b(t,x) - \frac{(\sigma\sigma_x')(x)}{2}\right) + \frac{(\sigma\sigma_x')(x)h}{2}\left(z + \frac{1}{\sqrt{h}\sigma_x'(x)}\right)^2.$$

with $b(x) = \kappa(\theta - x)$, $\sigma(x) = \xi\sqrt{x}$ and $\sigma_x'(x) = \frac{\xi}{2\sqrt{x}}$. Consequently, under the Feller condition, the positivity of $\mathcal{M}_{b,\sigma}(t,x,z)$ is ensured if

$$x \geqslant \frac{\sigma(x)}{2\sigma_x'(x)} \geqslant 0, \qquad b(t,x) \geqslant \frac{(\sigma\sigma_x')(x)}{2} \geqslant 0.$$

In our case, if the first condition holds true since

$$\frac{\sigma(x)}{2\sigma_x'(x)} = \frac{\xi\sqrt{x}}{2\frac{\xi}{2\sqrt{x}}} = x$$

the second one fails. Indeed

$$\frac{(\sigma\sigma'_x)(x)}{2} = \frac{\xi\sqrt{x}\frac{\xi}{2\sqrt{x}}}{2} = \frac{\xi^2}{4}$$

can be bigger than $b(t,x)$. In order to solve this problem, we consider the following *boosted* volatility process

$$Y_t = \mathrm{e}^{\kappa t}\, v_t,\ t \in [0,T]. \tag{5.82}$$

Milstein scheme for the *boosted* volatility. Let $Y_t = \mathrm{e}^{\kappa t}\, v_t$, $t \in [0,T]$ for some $\kappa > 0$, which satisfies, owing to Itô's formula

$$dY_t = \mathrm{e}^{\kappa t}\,\kappa\theta dt + \xi\,\mathrm{e}^{\kappa t/2}\,\sqrt{Y_t}d\widetilde{W}_t.$$

Remark. The process $(Y_t)_{t\in[0,T]}$ will have a higher variance but, having in mind a quantized scheme, this has no real impact (by contrast with a Monte Carlo simulation).

Now, if we look at the Milstein discretization scheme of Y_t

$$\bar{Y}_{t_{k+1}} = \mathcal{M}_{\widetilde{b},\widetilde{\sigma}}\big(t_k, \bar{Y}_{t_k}, Z^2_{k+1}\big)$$

using the notation defined in (5.28) where drift and volatility terms of the *boosted* process, now time-dependents, are given by

$$\widetilde{b}(t,x) = \mathrm{e}^{\kappa t}\,\kappa\theta, \qquad \widetilde{\sigma}(t,x) = \xi\sqrt{x}\,\mathrm{e}^{\kappa t/2} \qquad \text{and} \qquad \widetilde{\sigma}'_x(t,x) = \frac{\xi\,\mathrm{e}^{\kappa t/2}}{2\sqrt{x}}.$$

Under the Feller condition, the positivity of the scheme is ensured, since

$$\frac{\widetilde{\sigma}(t,x)}{2\widetilde{\sigma}'_x(t,x)} = x \qquad \text{and} \qquad \frac{(\widetilde{\sigma}\widetilde{\sigma}'_x)(t,x)}{2} = \frac{\xi^2\,\mathrm{e}^{\kappa t}}{4} \leqslant \widetilde{b}(t,x) = \mathrm{e}^{\kappa t}\,\kappa\theta.$$

The last inequality is satisfied thanks to the condition $\frac{\xi^2}{2\kappa\theta} \leqslant 1$ ensuring the positivity of the scheme.

Appendix 5.B L^p-linear growth of the hybrid scheme

The aim of this section is to show the L^p-linear growth of the scheme $F_k(u,z)$ with $u = (x,y)$ defined by

$$F_k(u,Z) = \begin{pmatrix} \mathcal{E}_{b,\sigma}\big(t_k, x, y, Z^1_{k+1}\big) \\ \mathcal{M}_{\widetilde{b},\widetilde{\sigma}}\big(t_k, y, Z^2_{k+1}\big) \end{pmatrix}. \tag{5.83}$$

where the schemes $\mathcal{E}_{b,\sigma}$ and $\mathcal{M}_{\widetilde{b},\widetilde{\sigma}}$ are defined in (5.32) and (5.28), respectively.

The results on the L^p-linear growth of the schemes are essentially based on the key Lemma 2.1 proved in [PS18b] in $\mathbb{R}^d$ that we recall below.

Lemma 5.B.1. *(a) Let $u \in \mathbb{R}^d$ and $A(u)$ be a $d \times q$-matrix and let $a(u) \in \mathbb{R}^d$. Let $p \in [2,3)$. For any centered random vector $\zeta \in L^p_{\mathbb{R}^d}(\Omega, \mathcal{A}, \mathbb{P})$, one has for every $h \in (0, +\infty)$*

$$\mathbb{E}\left[|a(u)+\sqrt{h}A(u)\zeta|^p\right] \leqslant \left(1+\frac{(p-1)(p-2)}{2}h\right)|a(u)|^p + h\big(1+p+h^{\frac{p}{2}-1}\big)\|A(u)\|^p\,\mathbb{E}\left[|\zeta|^p\right] \tag{5.84}$$

where $\|A(u)\| = \big(\mathrm{Tr}(A(u)A^\star(u))\big)^{1/2}$.

(b) In particular, if $|a(u)| \leqslant |u|(1+Lh)+Lh$ and $\|A(u)\|^p \leqslant 2^{p-1}\Upsilon^p(1+|u|^p)$, then

$$\mathbb{E}\left[|a(u)+\sqrt{h}A(u)\zeta|^p\right] \leqslant \big(\mathrm{e}^{\kappa_p h}L + K_p\big)h + \big(\mathrm{e}^{\kappa_p h} + K_p h\big)|u|^p, \tag{5.85}$$

where

$$\kappa_p = \frac{(p-1)(p-2)}{2} + 2pL \quad \text{and} \quad K_p = 2^{p-1}\Upsilon^p\big(1+p+h^{\frac{p}{2}-1}\big)\,\mathbb{E}\left[|\zeta|^p\right]. \tag{5.86}$$

Now, we will apply Lemma 5.B.1 to $F_k(u,z)$ defined in (5.83) further on in order to show its L^p-linear growth. Let $a(u) \in \mathbb{R}^2$ and let $A(u)$ be a 2×3-matrix defined by

$$a(u) = \begin{pmatrix} x + h\big(r - \frac{\mathrm{e}^{-\kappa t_k}\,y}{2}\big) \\ y + \mathrm{e}^{\kappa t_k}\,\kappa\theta h \end{pmatrix}, \quad A(u) = \begin{pmatrix} \mathrm{e}^{-\kappa t_k/2}\sqrt{y} & 0 & 0 \\ 0 & \sqrt{y}\,\mathrm{e}^{\kappa t_k/2} & \sqrt{h}\frac{\xi^2\,\mathrm{e}^{\kappa t_k}}{4} \end{pmatrix}$$

$$\text{and} \quad \zeta = \begin{pmatrix} Z^1_{k+1} \\ Z^2_{k+1} \\ (Z^2_{k+1})^2 - 1 \end{pmatrix}.$$

First, we show the linear growth of $a(u)$

$$\begin{aligned}
|a(u)| &= \Big(\Big|x + h\big(r - \frac{\mathrm{e}^{-\kappa t_k}\,y}{2}\big)\Big|^2 + \big|y + \mathrm{e}^{\kappa t_k}\,\kappa\theta h\big|^2\Big)^{1/2} \\
&= \Big(|x|^2 + |y|^2 + h^2\Big(r^2 + \frac{\mathrm{e}^{-2\kappa t_k}}{4}|y|^2\Big) + \mathrm{e}^{2\kappa t_k}\,\kappa^2\theta^2h^2\Big)^{1/2} \\
&\leqslant \Big(|u|^2\Big(1 + h^2\frac{\mathrm{e}^{-2\kappa t_k}}{4}\Big) + h^2\big(r^2 + \mathrm{e}^{2\kappa t_k}\,\kappa^2\theta^2\big)\Big)^{1/2} \\
&\leqslant |u|\Big(1 + h^2\frac{\mathrm{e}^{-2\kappa t_k}}{4}\Big)^{1/2} + h\big(r^2 + \mathrm{e}^{2\kappa t_k}\,\kappa^2\theta^2\big)^{1/2} \\
&\leqslant |u|\Big(1 + h\frac{h}{2}\Big) + h\big(r^2 + \mathrm{e}^{2\kappa T}\,\kappa^2\theta^2\big)^{1/2} \\
&\leqslant |u|(1+Lh) + Lh
\end{aligned}$$

where $L = \max\left(\frac{1}{2}, \left(r^2 + \mathrm{e}^{2\kappa T}\,\kappa^2\theta^2\right)^{1/2}\right)$. Then, we study $\|A(u)\|^p$

$$\begin{aligned}
\|A(u)\|^p &= \left(\mathrm{e}^{-\kappa t_k}\,|y| + |y|\,\mathrm{e}^{\kappa t_k} + h\frac{\xi^4\,\mathrm{e}^{2\kappa t_k}}{16}\right)^{p/2}\\
&= \left(|y|(\mathrm{e}^{-\kappa t_k} + \mathrm{e}^{\kappa t_k}) + h\frac{\xi^4\,\mathrm{e}^{2\kappa t_k}}{16}\right)^{p/2}\\
&\leqslant 2^{\frac{p}{2}-1}\left(|y|^{\frac{p}{2}}(\mathrm{e}^{-\kappa t_k} + \mathrm{e}^{\kappa t_k})^{\frac{p}{2}} + h^{\frac{p}{2}}\frac{\xi^{2p}\,\mathrm{e}^{p\kappa t_k}}{4^p}\right)\\
&\leqslant 2^{\frac{p}{2}-1}\left(\frac{|y|^p+1}{2}(\mathrm{e}^{-\kappa t_k} + \mathrm{e}^{\kappa t_k})^{\frac{p}{2}} + h^{\frac{p}{2}}\frac{\xi^{2p}\,\mathrm{e}^{p\kappa t_k}}{4^p}\right)\\
&\leqslant 2^{\frac{p}{2}-1}\frac{(1+\mathrm{e}^{\kappa T})^{\frac{p}{2}}}{2}\left(|y|^p + 1 + h^{\frac{p}{2}}\frac{\xi^{2p}\,\mathrm{e}^{p\kappa T}}{2^{2p-1}}\frac{1}{(1+\mathrm{e}^{\kappa T})^{\frac{p}{2}}}\right)\\
&\leqslant 2^{p-1}\Upsilon^p\left(1 + |u|^p\right)
\end{aligned}$$

where $\Upsilon^p = \frac{(1+\mathrm{e}^{\kappa T})^{\frac{p}{2}}}{2} + h^{\frac{p}{2}}\frac{\xi^{2p}\,\mathrm{e}^{p\kappa T}}{2^{2p}}$. Hence, by Lemma 5.B.1, the discretization scheme F_k has an L^p-linear growth

$$\mathbb{E}\left[|F_k(u, Z_{k+1})|^p\right] \leqslant \alpha_p + \beta_p|u|^p$$

with

$$\alpha_p = \left(\mathrm{e}^{\kappa_p h}\,L + K_p\right)h \quad \text{and} \quad \beta_p = \mathrm{e}^{\kappa_p h} + K_p h \tag{5.87}$$

where K_p and κ_p are defined in the Lemma 5.B.1.

Appendix 5.C Proof of the L^2-error estimation of Proposition 5.3.4

We have, for every $k = 0, \ldots, n-1$

$$\begin{aligned}
\widehat{U}_{k+1} - \bar{U}_{k+1} &= \widehat{U}_{k+1} - \widetilde{U}_{k+1} + \widetilde{U}_{k+1} - \bar{U}_{k+1}\\
&= \widehat{U}_{k+1} - \widetilde{U}_{k+1} + F_k(\widehat{U}_k, Z_{k+1}) - F_k(\bar{U}_k, Z_{k+1})
\end{aligned} \tag{5.88}$$

by the very definition of $\widetilde{U}_{k+1}$ and $\bar{U}_{k+1}$. Hence,

$$\begin{aligned}
\|\widehat{U}_{k+1} - \bar{U}_{k+1}\|_2 &\leqslant \|\widehat{U}_{k+1} - \widetilde{U}_{k+1}\|_2 + \|\widetilde{U}_{k+1} - \bar{U}_{k+1}\|_2\\
&\leqslant \|\widehat{U}_{k+1} - \widetilde{U}_{k+1}\|_2 + \|F_k(\widehat{U}_k, Z_{k+1}) - F_k(\bar{U}_k, Z_{k+1})\|_2.
\end{aligned} \tag{5.89}$$

Using the definition of Milstein scheme of the *boosted*-volatility models $\mathcal{M}_{\tilde{b},\tilde{\sigma}}$ in (5.29), the $\frac{1}{2}$-Hölder property of $\sqrt{x}$, for every $y, y' \in \mathbb{R}_+$ one has

$$\begin{aligned}|\mathcal{M}_{\tilde{b},\tilde{\sigma}}(t,y,z) - \mathcal{M}_{\tilde{b},\tilde{\sigma}}(t,y',z)| &= \left|\Big(z\frac{\xi\,\mathrm{e}^{\kappa t/2}\sqrt{h}}{2} + \sqrt{y}\Big)^2 - \Big(z\frac{\xi\,\mathrm{e}^{\kappa t/2}\sqrt{h}}{2} + \sqrt{y'}\Big)^2\right| \\ &\leqslant \big|\sqrt{y} - \sqrt{y'}\big|\big(|z|\xi\,\mathrm{e}^{\kappa t/2}\sqrt{h} + \sqrt{y} + \sqrt{y'}\big) \\ &\leqslant \sqrt{|y-y'|}\sqrt{h}|z|\xi\,\mathrm{e}^{\kappa t/2} + |y-y'|\end{aligned} \tag{5.90}$$

and using the definition of the Euler-Maruyama scheme of the log-asset $\mathcal{E}_{b,\sigma}$ defined in (5.32) we have, for any $x, x', y, y' \in \mathbb{R}_+$

$$\big|\mathcal{E}_{b,\sigma}(t,x,y,z) - \mathcal{E}_{b,\sigma}(t,x',y',z)\big| \leqslant |x-x'| + \frac{\mathrm{e}^{-\kappa t}}{2}h|y-y'| + \mathrm{e}^{-\kappa t/2}\sqrt{h}|z|\sqrt{|y-y'|}. \tag{5.91}$$

Now, when we replace x, y, x', y' by $\widehat{X}_k, \widehat{Y}_k, \bar{X}_k, \bar{Y}_k$ in the last expression, we get an upper-bound for the last term of (5.89)

$$\begin{aligned}&\big\|F_k(\widehat{U}_k, Z_{k+1}) - F_k(\bar{U}_k, Z_{k+1})\big\|_2 \\ &\quad\leqslant \big\|\mathcal{E}_{b,\sigma}\big(t_k, \widehat{X}_k, \widehat{Y}_k, Z^1_{k+1}\big) - \mathcal{E}_{b,\sigma}\big(t_k, \bar{X}_k, \bar{Y}_k, Z^1_{k+1}\big)\big\|_2 \\ &\qquad\qquad + \big\|\mathcal{M}_{\tilde{b},\tilde{\sigma}}\big(t_k, \widehat{Y}_k, Z^2_{k+1}\big) - \mathcal{M}_{\tilde{b},\tilde{\sigma}}\big(t_k, \bar{Y}_k, Z^2_{k+1}\big)\big\|_2 \\ &\quad\leqslant \|\widehat{X}_k - \bar{X}_k\|_2 + \Big(1 + \frac{\mathrm{e}^{-\kappa t_k}}{2}h\Big)\|\widehat{Y}_k - \bar{Y}_k\|_2 + \Big\|\sqrt{h}\big(\xi\,\mathrm{e}^{\kappa t_k/2} + \mathrm{e}^{-\kappa t_k/2}\big)\sqrt{|\widehat{Y}_k - \bar{Y}_k|}\Big\|_2 \\ &\quad\leqslant \|\widehat{X}_k - \bar{X}_k\|_2 + \Big(1 + \frac{\mathrm{e}^{-\kappa t_k}}{2}h\Big)\|\widehat{Y}_k - \bar{Y}_k\|_2 + \Big\|\sqrt{2h(\xi^2\,\mathrm{e}^{\kappa t_k} + \mathrm{e}^{-\kappa t_k})}\sqrt{|\widehat{Y}_k - \bar{Y}_k|}\Big\|_2.\end{aligned} \tag{5.92}$$

Now, using that $\sqrt{a}\sqrt{b} \leqslant \frac{1}{2}\big(\frac{a}{\lambda} + b\lambda\big)$ with $\sqrt{a} = \sqrt{2h(\xi^2\,\mathrm{e}^{\kappa t_k} + \mathrm{e}^{-\kappa t_k})}$ and $\sqrt{b} = \sqrt{|\widehat{Y}_k - \bar{Y}_k|}$ where we considere that $\lambda = \sqrt{h}(1-\sqrt{h})$. Wo choose λ of this order because we wish to divide equally the impact of h and get $\sqrt{h}$ on each side. Hence, we have

$$\sqrt{2h(\xi^2\,\mathrm{e}^{\kappa t_k} + \mathrm{e}^{-\kappa t_k})}\sqrt{|\widehat{Y}_k - \bar{Y}_k|} \leqslant \frac{1}{2}\left(\frac{2h(\xi^2\,\mathrm{e}^{\kappa t_k} + \mathrm{e}^{-\kappa t_k})}{\lambda} + |\widehat{Y}_k - \bar{Y}_k|\lambda\right). \tag{5.93}$$

Then,

$$\begin{aligned}
&\big\|F_k(\widehat{U}_k, Z_{k+1}) - F_k(\bar{U}_k, Z_{k+1})\big\|_2 \\
&\quad\leqslant \|\widehat{X}_k - \bar{X}_k\|_2 + \Big(1 + \frac{\mathrm{e}^{-\kappa t_k}}{2}h\Big)\|\widehat{Y}_k - \bar{Y}_k\|_2 + \Big\|\frac{1}{2}\Big(\frac{2h(\xi^2\,\mathrm{e}^{\kappa t_k} + \mathrm{e}^{-\kappa t_k})}{\lambda} + |\widehat{Y}_k - \bar{Y}_k|\lambda\Big)\Big\|_2 \\
&\quad\leqslant \|\widehat{X}_k - \bar{X}_k\|_2 + \Big(1 + \frac{\mathrm{e}^{-\kappa t_k}}{2}h + \frac{\lambda}{2}\Big)\|\widehat{Y}_k - \bar{Y}_k\|_2 + (\xi^2\,\mathrm{e}^{\kappa t_k} + \mathrm{e}^{-\kappa t_k})\frac{h}{\lambda} \\
&\quad\leqslant \sqrt{2}\Big(1 + \frac{h}{2} + \frac{\lambda}{2}\Big)\|\widehat{U}_k - \bar{U}_k\|_2 + C_T\frac{h}{\lambda} \\
&\quad\leqslant \sqrt{2}\Big(1 + \frac{\sqrt{h}}{2}\Big)\|\widehat{U}_k - \bar{U}_k\|_2 + C_T(h)\sqrt{h}
\end{aligned} \tag{5.94}$$

where $C_T(h) = (1 + \xi^2\,\mathrm{e}^{\kappa T})(1 - \sqrt{h})^{-1} = O(1)$.

Finally, (5.89) is upper-bounded by

$$\begin{aligned}
\|\widehat{U}_{k+1} - \bar{U}_{k+1}\|_2 &\leqslant \|\widehat{U}_{k+1} - \widetilde{U}_{k+1}\|_2 + \sqrt{2}\Big(1 + \frac{\sqrt{h}}{2}\Big)\|\widehat{U}_k - \bar{U}_k\|_2 + C_T(h)\sqrt{h} \\
&\leqslant \sum_{j=0}^{k+1}\|\widehat{U}_j - \widetilde{U}_j\|_2 2^{\frac{k-j+1}{2}}\Big(1 + \frac{\sqrt{h}}{2}\Big)^{k-j+1} + \sqrt{h}C_T(h)\sum_{j=0}^{k}2^{\frac{k-j}{2}}\Big(1 + \frac{\sqrt{h}}{2}\Big)^{k-j} \\
&\leqslant \sum_{j=0}^{k+1}A_{j,k+1}\|\widehat{U}_j - \widetilde{U}_j\|_2 + B_{k+1}\sqrt{h}
\end{aligned} \tag{5.95}$$

where

$$A_{j,k} = 2^{\frac{k-j}{2}}\,\mathrm{e}^{\frac{\sqrt{h}}{2}(k-j)} \quad\text{and}\quad B_k = C_T(h)\sum_{j=0}^{k-1}2^{\frac{k-1-j}{2}}\,\mathrm{e}^{\frac{\sqrt{h}}{2}(k-1-j)} \tag{5.96}$$

and $\sum_{\varnothing} = 0$ by convention.

Now, we follow the lines of the proof developed in [PS18b], we apply the revisited Pierce's lemma for product quantization (Lemma 2.3 in [PS18b]) with $r = 2$ and let $p > r = 2$, which yields

$$\|\widehat{U}_{k+1} - \bar{U}_{k+1}\|_2 \leqslant 2^{\frac{p-2}{2p}}C_p\sum_{j=0}^{k+1}A_{j,k+1}\|\widetilde{U}_j\|_p\big(N_{1,j}\times N_{2,j}\big)^{-1/2} + B_{k+1}\sqrt{h} \tag{5.97}$$

where $C_p = 2C_{1,p}$ and $C_{1,p}$ is the constant appearing in Pierce lemma (see the second item in Theorem 5.D.7 and [GL00] for further details) and we used that $\|\widetilde{U}_j\|_p \geqslant \sigma_p(\widetilde{U}_j) = \inf_{a\in\mathbb{R}^2}\|\widetilde{U}_j - a\|_p$. Moreover, noting that the hybrid discretization scheme F_k has an L^p-linear growth, (see Appendix 5.B), i.e.

$$\forall k = 0,\ldots,n-1,\quad \forall u\in\mathbb{R}^2,\quad \mathbb{E}\big[|F_k(u, Z_{k+1})|^p\big] \leqslant \alpha_p + \beta_p|x|^p, \tag{5.98}$$

where the coefficients α_p and β_p are defined in (5.87). Hence, for all $j = 0, \dots, n-1$, we have

$$\|\widetilde{U}_{j+1}\|_p^p = \mathbb{E}\left[\,\mathbb{E}\left[|F_j(\widehat{U}_j, Z_{j+1})|^p \mid \widehat{U}_j\right]\right] \leqslant \alpha_p + \beta_p \|\widehat{U}_j\|_p^p. \tag{5.99}$$

Furthermore, $\mathbb{E}\left[|\widehat{U}_j|^p\right]$ can be upper-bounded using Jensen's inequality and the stationary property satisfied by $\widehat{X}_j$ and $\widehat{Y}_j$ independently. Indeed, they are one-dimensional quadratic optimal quantizers of $\widetilde{X}_j$ and $\widetilde{Y}_j$, respectively, hence they are stationary in the sense of Proposition 5.D.5.

$$\begin{aligned}
\|\widehat{U}_j\|_p^p &\leqslant 2^{\frac{p}{2}-1}\Big(\mathbb{E}\left[|\widehat{X}_j|^p\right] + \mathbb{E}\left[|\widehat{Y}_j|^p\right]\Big) \\
&\leqslant 2^{\frac{p}{2}-1}\left(\mathbb{E}\left[\left|\,\mathbb{E}\left[\widetilde{X}_j \mid \widehat{X}_j\right]\right|^p\right] + \mathbb{E}\left[\left|\,\mathbb{E}\left[\widetilde{Y}_j \mid \widehat{Y}_j\right]\right|^p\right]\right) \\
&\leqslant 2^{\frac{p}{2}-1}\Big(\mathbb{E}\left[|\widetilde{X}_j|^p\right] + \mathbb{E}\left[|\widetilde{Y}_j|^p\right]\Big) \\
&= 2^{\frac{p}{2}-1}\|\widetilde{U}_j\|_p^p \\
&\leqslant 2^{\frac{p}{2}-1}\|\widetilde{U}_j\|_2^p.
\end{aligned} \tag{5.100}$$

Now, plugging this upper-bound in (5.99) and by a standard induction argument, we have

$$\begin{aligned}
\|\widetilde{U}_j\|_p^p &\leqslant \alpha_p + \beta_p 2^{\frac{p}{2}-1}\|\widetilde{U}_{j-1}\|_2^p \\
&\leqslant 2^{(\frac{p}{2}-1)j}\beta_p^j\|\widehat{U}_0\|_2^p + \alpha_p \sum_{i=0}^{j-1}\left(2^{\frac{p}{2}-1}\beta_p\right)^i \\
&\leqslant 2^{(\frac{p}{2}-1)j}\beta_p^j\|\widehat{U}_0\|_2^p + \alpha_p \frac{1 - 2^{(\frac{p}{2}-1)j}\beta_p^j}{1 - 2^{\frac{p}{2}-1}\beta_p}.
\end{aligned} \tag{5.101}$$

Hence, using the upper-bound (5.101) in (5.97), we have

$$\begin{aligned}
&\|\widehat{U}_{k+1} - \bar{U}_{k+1}\|_2 \\
&\quad\leqslant 2^{\frac{p-2}{2p}} C_p^2 \sum_{j=0}^{k+1} A_{j,k+1}\left(2^{(\frac{p}{2}-1)j}\beta_p^j\|\widehat{U}_0\|_2^p + \alpha_p \frac{1 - 2^{(\frac{p}{2}-1)j}\beta_p^j}{1 - 2^{\frac{p}{2}-1}\beta_p}\right)^{1/p}\left(N_{1,j} \times N_{2,j}\right)^{-1/2} + B_{k+1}\sqrt{h} \\
&\quad\leqslant \sum_{j=0}^{k+1} \widetilde{A}_{j,k+1}\left(N_{1,j} \times N_{2,j}\right)^{-1/2} + B_{k+1}\sqrt{h}
\end{aligned} \tag{5.102}$$

yielding the desired result with

$$\widetilde{A}_{j,k} = 2^{\frac{p-2}{2p}} C_p^2 A_{j,k}\left(2^{(\frac{p}{2}-1)j}\beta_p^j\|\widehat{U}_0\|_2^p + \alpha_p \frac{1 - 2^{(\frac{p}{2}-1)j}\beta_p^j}{1 - 2^{\frac{p}{2}-1}\beta_p}\right)^{1/p}. \tag{5.103}$$

Appendix 5.D Quadratic Optimal Quantization: Generic Approach

Let X be a $\mathbb{R}$-valued random variable with distribution $\mathbb{P}_X$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that $X \in L^2_{\mathbb{R}}(\Omega, \mathcal{A}, \mathbb{P})$.

Definition 5.D.1. Let $\Gamma_N = \{x_1^N, \dots, x_N^N\} \subset \mathbb{R}$ be a subset of size N, called N-quantizer. A Borel partition $(C_i(\Gamma_N))_{i\in\{1,\dots,N\}}$ of $\mathbb{R}$ is a Voronoï partition of $\mathbb{R}$ induced by the N-quantizer Γ_N if, for every $i \in \{1, \dots, N\}$,

$$C_i(\Gamma_N) \subset \big\{\xi \in \mathbb{R}, |\xi - x_i^N| \leqslant \min_{j\neq i} |\xi - x_j^N|\big\}.$$

The Borel sets $C_i(\Gamma_N)$ are called Voronoï cells of the partition induced by Γ_N.

Remark. Any such N-quantizer is in correspondence with the N-tuple $x = (x_1^N, \dots, x_N^N) \in (\mathbb{R})^N$ as well as with all N-tuples obtained by a permutation of the components of x. This is why we will sometimes replace Γ_N by x.

If the quantizers are in non-decreasing order: $x_1^N < x_2^N < \cdots < x_{N-1}^N < x_N^N$, then the Voronoï cells are given by

$$C_i(\Gamma_N) = \big(x_{i-1/2}^N, x_{i+1/2}^N\big], \qquad i \in \{1, \dots, N-1\}, \qquad C_N(\Gamma_N) = \big(x_{N-1/2}^N, x_{N+1/2}^N\big) \quad (5.104)$$

where $\forall i \in\in \{2, \dots, N\}, x_{i-1/2}^N = \frac{x_{i-1}^N + x_i^N}{2}$ and $x_{1/2}^N = -\infty$ and $x_{N+1/2}^N = +\infty$.

Definition 5.D.2. The Voronoï quantization of X by Γ_N, $\widehat{X}^N$, is defined as the nearest neighbour projection of X onto Γ_N

$$\widehat{X}^N = \mathrm{Proj}_{\Gamma_N}(X) = \sum_{i=1}^N x_i^N \mathbb{1}_{X\in C_i(\Gamma_N)} \quad (5.105)$$

and its associated probabilities, also called weights, are given by

$$\mathbb{P}\left(\widehat{X}^N = x_i^N\right) = \mathbb{P}_X\left(C_i(\Gamma_N)\right) = \mathbb{P}\left(X \in \big(x_{i-1/2}^N, x_{i+1/2}^N\big]\right).$$

Definition 5.D.3. The quadratic distortion function at level N induced by an N-tuple $x = (x_1^N, \dots, x_N^N)$ is given by

$$\mathcal{Q}_{2,\mathrm{N}} : x \longmapsto \frac{1}{2}\,\mathbb{E}\Big[\min_{i\in\{1,\dots,N\}} |X - x_i^N|^2\Big] = \frac{1}{2}\,\mathbb{E}\big[\,\mathrm{dist}(X, \Gamma_N)^2\big] = \frac{1}{2}\|X - \widehat{X}^N\|_2^2.$$

Of course, the above result can be extended to the L^p case by considering the L^p-mean quantization error in place of the quadratic one.

We briefly recall some classical theoretical results, see [GL00; Pag18] for further details. The first one treats of existence of optimal quantizers.

Theorem 5.D.4. *(Existence of optimal N-quantizers) Let $X \in L^2_{\mathbb{R}}(\mathbb{P})$ and $N \in \mathbb{N}^\star$.*

(a) The quadratic distortion function $\mathcal{Q}_{2,\mathrm{N}}$ at level N attains a minimum at a N-tuple $x^\star = (x_1^N, \ldots, x_N^N)$ and $\Gamma_N^\star = \{x_i^N,\, i \in \{1, \ldots, N\}\}$ is a quadratic optimal quantizer at level N.

(b) If the support of the distribution $\mathbb{P}_X$ of X has at least N elements, then $x^\star = (x_1^N, \ldots, x_N^N)$ has pairwise distinct components, $\mathbb{P}_X\left(C_i(\Gamma_N^\star)\right) > 0$, $i \in \{1, \ldots, N\}$. Furthermore, the sequence $N \mapsto \inf_{x \in (\mathbb{R})^N} \mathcal{Q}_{2,\mathrm{N}}(x)$ converges to 0 and is decreasing as long as it is positive.

A really interesting and useful property concerning quadratic optimal quantizers is the stationary property, this property is deeply connected to the addressed problem after for the optimization of the quadratic optimal quantizers in (5.109).

Proposition 5.D.5. *(Stationarity) Assume that the support of $\mathbb{P}_X$ has at least N elements. Any L^2-optimal N-quantizer $\Gamma_N \in (\mathbb{R})^N$ is stationary in the following sense: for every Voronoï quantization $\widehat{X}^N$ of X,*

$$\mathbb{E}\left[X \mid \widehat{X}^N\right] = \widehat{X}^N.$$

Moreover $\mathbb{P}\left(X \in \bigcup_{i=1,\ldots,N} \partial C_i(\Gamma_N)\right) = 0$, so all optimal quantization induced by Γ_N a.s. coincide.

The uniqueness of an optimal N-quantizer, due to Kieffer [Kie82], was shown in dimension one under some assumptions on the density of X.

Theorem 5.D.6. *(Uniqueness of optimal N-quantizers see [Kie82]) If $\mathbb{P}_X(d\xi) = \varphi(\xi)d\xi$ with $\log \varphi$ concave, then for every $N \geqslant 1$, there is exactly one stationary N-quantizer (up to the permutations of the N-tuple). This unique stationary quantizer is a global (local) minimum of the distortion function, i.e.*

$$\forall N \geqslant 1, \qquad \arg\min_{\mathbb{R}^N} \mathcal{Q}_{2,\mathrm{N}} = \{x^\star\}.$$

In what follows, we will drop the star notation ($\star$) when speaking of optimal quantizers, $x^\star$ and $\Gamma_N^\star$ will be replaced by x and Γ_N.

The next result elucidates the asymptotic behavior of the distortion. We saw in Theorem 5.D.4 that the infimum of the quadratic distortion converges to 0 as N goes to infinity. The next theorem, known as Zador's Theorem, establishes the sharp rate of convergence of the L^p-mean quantization error.

Theorem 5.D.7. *(Zador's Theorem) Let $p \in (0, +\infty)$.*

(a) Sharp rate [Zad82; GL00]. *Let $X \in L^{p+\delta}_{\mathbb{R}}(\mathbb{P})$ for some $\delta > 0$. Let $\mathbb{P}_X(d\xi) = \varphi(\xi)\cdot\lambda(d\xi) + \nu(d\xi)$, where $\nu \perp \lambda$ i.e., is singular with respect to the Lebesgue measure λ on $\mathbb{R}$. Then, there is a constant $\widetilde{J}_{p,1} \in (0,+\infty)$ such that*

$$\lim_{N\to+\infty} N \min_{\Gamma_N\subset\mathbb{R},|\Gamma_N|\leqslant N} \|X-\widehat{X}^N\|_p = \frac{1}{2^p(p+1)}\left[\int_{\mathbb{R}} \varphi^{\frac{1}{1+p}} d\lambda\right]^{1+\frac{1}{p}}. \tag{5.106}$$

(b) Non asymptotic upper-bound [GL00; Pag18]. *Let $\delta > 0$. There exists a real constant $C_{1,p} \in (0,+\infty)$ such that, for every $\mathbb{R}$-valued random variable X,*

$$\forall N \geqslant 1, \qquad \min_{\Gamma_N\subset\mathbb{R},|\Gamma_N|\leqslant N} \|X-\widehat{X}^N\|_p \leqslant C_{1,p}\sigma_{\delta+p}(X)N^{-1} \tag{5.107}$$

where, for $r \in (0,+\infty), \sigma_r(X) = \min_{a\in\mathbb{R}} \|X-a\|_r < +\infty$ is the L^r-pseudo-standard deviation.

Now, we will be interested by the construction of such quadratic optimal quantizer. We differentiate $\mathcal{Q}_{2,\mathrm{N}}$, whose gradient is given by

$$\nabla\,\mathcal{Q}_{2,\mathrm{N}}(x) = \left(\mathbb{E}\left[(x_i^N - X)\,\mathbb{1}_{X\in\left(x^N_{i-1/2},x^N_{i+1/2}\right]}\right]\right)_{i=1,\dots,N}. \tag{5.108}$$

Moreover, if x is solution to the distortion minimization problem then it satisfies

$$\begin{aligned} \nabla\,\mathcal{Q}_{2,\mathrm{N}}(x) = 0 \quad &\Longleftrightarrow \quad x_i^N = \frac{\mathbb{E}\left[X\,\mathbb{1}_{X\in\left(x^N_{i-1/2},x^N_{i+1/2}\right]}\right]}{\mathbb{P}\left(X\in\left(x^N_{i-1/2},x^N_{i+1/2}\right]\right)}, \qquad i=1,\dots,N \\ &\Longleftrightarrow \quad x_i^N = \frac{K_X\left(x^N_{i+1/2}\right)-K_X\left(x^N_{i-1/2}\right)}{F_X\left(x^N_{i+1/2}\right)-F_X\left(x^N_{i-1/2}\right)}, \qquad i=1,\dots,N \end{aligned} \tag{5.109}$$

where $K_X(\cdot)$ and $F_X(\cdot)$ are the first partial moment and the cumulative distribution respectively, function of X, i.e.

$$K_X(x) = \mathbb{E}\left[X\,\mathbb{1}_{X\leqslant x}\right] \qquad \text{and} \qquad F_X(x) = \mathbb{P}\left(X\leqslant x\right). \tag{5.110}$$

Hence, one can notices that the optimal quantizer that cancel the gradient defined in (5.109), hence is an optimal quantizer, is a stationary quantizer in the following sense

$$\mathbb{E}\left[\widehat{x}^N \mid X\right] = \widehat{X}^N. \tag{5.111}$$

The last equality in (5.109) was the starting point to the development of the first method devoted to the numerical computation of optimal quantizers: the Lloyd's method I. This method was first devised in 1957 by S.P. Lloyd and published later [Llo82]. Starting from a

sorted N-tuple $x^{[0]}$ and with the knowledge of the first partial moment K_X and the cumulative distribution function F_X of X, the algorithm, which is essentially a deterministic fixed point method, is defined as follows

$$x_i^{N,[n+1]} = \frac{K_X\left(x_{i+1/2}^{N,[n]}\right) - K_X\left(x_{i-1/2}^{N,[n]}\right)}{F_X\left(x_{i+1/2}^{N,[n]}\right) - F_X\left(x_{i-1/2}^{N,[n]}\right)}, \qquad i = 1, \ldots, N. \tag{5.112}$$

In the seminal paper of [Kie82], it has been shown that $\left(x^{[n]}\right)_{n\geqslant 1}$ converges exponentially fast toward x, the optimal quantizer, when the density φ of X is log-concave and not piecewise affine. Numerical optimizations can be made in order to increase the rate of convergence to the optimal quantizer such as fixed point search acceleration, for example the Anderson acceleration (see [And65] for the original paper and [WN11] for details on the procedure).

Of course, other algorithms exist, such as the Newton Raphson zero search procedure or its variant the Levenberg–Marquardt algorithm which are deterministic procedures as well if the density, the first partial moment and the cumulative distribution function of X are known. Additionally, we can cite stochastic procedures such as the CLVQ procedure (Competitive Learning Vector Quantization) which is a zero search stochastic gradient and the randomized version of the Lloyd's method I. For more details, the reader can refer to [Pag18; PY16].

Once the algorithm (5.112) has been converging, we have at hand the quadratic optimal quantizer $\widehat{X}^N$ of X and its associated probabilities given by

$$\mathbb{P}\left(\widehat{X}^N = x_i^n\right) = F_X\left(x_{i+1/2}^N\right) - F_X\left(x_{i-1/2}^N\right), \qquad i = 1, \ldots, n. \tag{5.113}$$

Chapter 6

Quantization-based Bermudan option pricing in the FX world

This chapter corresponds to the article "Quantization-based Bermudan option pricing in the FX world" submitted to *Journal of Computational Finance* and accessible in arXiv or HAL (see [Fay+19]). This article is a joint work with Jean-Michel Fayolle, Vincent Lemaire and Gilles Pagès.

Abstract This paper proposes two numerical solution based on Product Optimal Quantization for the pricing of Foreign Exchange (FX) linked long term Bermudan options e.g. Bermudan Power Reverse Dual Currency options, where we take into account stochastic domestic and foreign interest rates on top of stochastic FX rate, hence we consider a 3-factor model. For these two numerical methods, we give an estimation of the L^2-error induced by such approximations and we illustrate them with market-based examples that highlight the speed of such methods.

Introduction

Persistent low levels of interest rates in Japan in the latter decades of the 20th century were one of the core sources that led to the creation of structured financial products responding to the need of investors for coupons higher than the low yen-based ones. This started with relatively simple dual currency notes in the 80s where coupons were linked to foreign (i.e. non yen-based) currencies enabling payments of coupons significantly higher. As time (and issuers' competition) went by, such structured notes were iteratively "enhanced" to reverse dual currency, power reverse dual currency (PRDC), cancellable power reverse dual currency etc., each version adding further features such as limits, early repayment options, etc. Finally, in the early 2000s, the denomination xPRD took root to describe those structured notes typically long-dated (over 30y initial term) and based on multiple currencies (see [Wys17]). The total notional invested in such notes is likely to be in the hundreds of billions of USD. The valuation

of such investments obviously requires the modeling of the main components driving the key risks, namely the interest rates of each pair of currencies involved as well as the corresponding exchange rates. In its simplest and most popular version, that means 3 sources of risk: domestic and foreign rates and the exchange rate. The 3-factor model discussed herein is an answer to that problem.

Gradually, as the note's features became more and more complex, further refinements to the modeling were needed, for instance requiring the inclusion of the volatility smile, the dependence of implied volatilities on both the expiry and the strike[1] of the option, prevalent in the FX options market. Such more complete modeling should ideally consist in successive refinements of the initial modeling enabling consistency across the various flavors of xPRDs at stake.

The model discussed herein was one of the answers popular amongst practitioners for multiple reasons: it was accounting for the main risks – interest rates in the currencies involved and exchange rates – in a relatively simple manner and the numerical implementations proposed at that time were based on simple extensions of well-known single dimensional techniques such as 3 dimensional trinomial trees, PDE based method (see [Pit05]) or on Monte Carlo simulations.

Despite the qualities of these methods, the calculation time could be rather slow (around 20 minutes with a trinomial tree for one price), especially when factoring in the cost for hedging (that is, measuring the sensitivities to all the input parameters) and even more post 2008, where the computation of risk measures and their sensitivities to market values became a central challenge for the financial markets participants. Indeed, even though these products were issued towards the end of the 20th century, they are still present in the banks's books and need to be considered when evaluating counterparty risk computations such as Credit Valuation Adjustment (CVA), Debt Valuation Adjustment (DVA), Funding Valuation Adjustment (FVA), Capital Valuation Adjustment (KVA), ..., in short xVA's (see [BMP13; CBB14; Gre15] for more details on the subject). Hence, a fast and accurate numerical method is important for being able to produce the correct values in a timely manner. The present paper aims at providing an elegant and efficient answer to that problem of numerical efficiency based on Optimal Quantization. Our novel method allows us reach a computation time of 1 or 2 seconds at the expense of a systematic error that we quantify in Section 6.3.

Let $P(t,T)$ be the value at time t of one unit of the currency delivered (that is, paid) at time T, also known as a zero coupon price or discount factor. A few iterations were needed by researchers and practitioners before the seminal family of Heath-Jarrow-Morton models came about. The general Heath-Jarrow-Morton (HJM) family of yield curve models can be expressed as follows – although originally expressed by its authors in terms of rates dynamics, the two

[1] In the case of the FX, the implied volatility is expressed in function of the delta.

are equivalent, see [HJM92] – in a n-factor setting, we have for the curve $P(t,T)$ that

$$\frac{dP(t,T)}{P(t,T)} = r_t dt + \sum_i \sigma_i\big(t,T,P(t,T)\big)dW_t^i \tag{6.1}$$

where r_t is the instantaneous rate at time t (therefore a random variable), W^i, $i = 1,\cdots,n$ are n correlated Brownian motions and $\sigma_i\big(t,T,P(t,T)\big)$ are volatility functions in the most general settings (with the obvious constraint that $\sigma_i\big(T,T,P(T,T)\big) = 0$). Indeed, the general HJM framework allows for the volatility functions $\sigma_i\big(t,T,P(t,T)\big)$ to also depend on the yield curve's (random) levels up to t – actually through forward rates – and therefore be random too. However, it has been demonstrated in [EMV92] that, to keep a tractable version (i.e. a finite number of state variables), the volatility functions must be of a specific form, namely, of the mean-reverting type (where the mean reversion can also depend on time). We use this way of expressing the model as a mean to recall that such model is essentially the usual and well-known Black Scholes model applied to all and any zero-coupon prices, with various enhancements regarding number of factors and volatility functions, to keep the calculations tractable. For further details and theory, one can refer to some of the following articles [EFG96; EMV92; HJM92; BS73]. Of course, such a framework can be applied to any yield curve. In its simplest form (i.e. flat volatility and one-factor), we have under the risk-neutral measure

$$\frac{dP(t,T)}{P(t,T)} = r_t dt + \sigma(T-t)dW_t \tag{6.2}$$

where W is a standard Brownian motion under the risk-neutral probability. In that case, σ is the flat volatility, which means the volatility of (zero-coupon) interest rates. That is often referred to as a Hull-White model without mean reversion (see [HW93]) or a continuous-time version of the Ho-Lee model. In the rest of the paper, we work with the model presented in (6.2) for the diffusion of the zero coupon although the extension to non-flat volatilities is easily feasible.

About the Foreign Exchange (FX) rate, we denote by S_t the value at time $t > 0$ of one unit of foreign currency in the domestic one. The diffusion is that of a standard Black-Scholes model with the following equation

$$\frac{dS_t}{S_t} = (r_t^d - r_t^f)dt + \sigma_S dW_t^S \tag{6.3}$$

where r_t^d is the instantaneous rate of the domestic currency at time t, r_t^f is the instantaneous rate of the foreign currency at time t, σ_S is the volatility of the FX rate and W^S is a standard Brownian motion under the risk-neutral probability.

Let us briefly recall the principle of the adopted numerical method, Optimal quantization. Optimal Quantization is a numerical method whose aim is to approximate optimally, for a

given norm, a continuous random signal by a discrete one with a given cardinality at most N. [She97] was the first to work on it for the uniform distribution on unit hypercubes. Since then, it has been extended to more general distributions with applications to Signal transmission in the 50's at the Bell Laboratory (see [GG82]). Formally, let Z be an $\mathbb{R}^d$-valued random vector with distribution $\mathbb{P}_Z$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that $Z \in L^2(\mathbb{P})$. We search for Γ_N, a finite subset of $\mathbb{R}^d$ defined by $\Gamma_N := \{z_1^N, \dots, z_N^N\} \subset \mathbb{R}^d$, solution to the following problem

$$\min_{\Gamma_N \subset \mathbb{R}^d, |\Gamma_N| \leqslant N} \|Z - \widehat{Z}^N\|_2$$

where $\widehat{Z}^N$ denotes the nearest neighbour projection of Z onto Γ_N. This problem can be extended to the L^p-optimal quantization by replacing the L^2-norm by the L^p-norm but this not in the scope of this paper. In our case, we mostly consider quadratic one-dimensional optimal quantization, i.e $d = 1$ and $p = 2$. The existence of an optimal quantizer at level N goes back to [CGM97] (see also [Pag98; GL00] for further developments). In the one-dimensional case, if the distribution of Z is absolutely continuous with a *log-concave* density, then there exists a unique optimal quantizer at level N, see [Kie83]. We scale to the higher dimension using Optimal Product Quantization which deals with multi-dimensional quantizers built by considering the cartesian product of one-dimensional optimal quantizers.

Considering again $Z = (Z^\ell)_{\ell=1:d}$, a $\mathbb{R}^d$-valued random vector. First, we look separately at each component Z^ℓ independently by building a one-dimensional optimal quantization $\widehat{Z}^\ell$ of size N^ℓ, with quantizer $\Gamma_\ell^{N_\ell} = \left\{z_{i_\ell}^\ell, i_\ell \in \{1, \cdots, N_\ell\}\right\}$ and then, by applying the cartesian product between the one-dimensional optimal quantizers, we build the product quantizer $\Gamma^N = \prod_{\ell=1}^d \Gamma_\ell^{N_\ell}$ with cardinality $N = N^1 \times \cdots \times N^d$ by

$$\Gamma^N = \left\{(z_{i_1}^1, \cdots, z_{i_\ell}^\ell, \cdots, z_{i_d}^d), \quad i_\ell \in \{1, \cdots, N_\ell\}, \quad \ell \in \{1, \cdots, d\}\right\}. \tag{6.4}$$

Then, in the 90s, [Pag98] developed quantization-based cubature formulas for numerical integration purposes and expectation approximations. Indeed, let f be a continuous function $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ such that $f(Z) \in L^1(\mathbb{P})$, we can define the following quantization-based cubature formula using the discrete property of the quantizer $\widehat{Z}^N$

$$\mathbb{E}\left[f(\widehat{Z}^N)\right] = \sum_{i=1}^N p_i f(z_i^N)$$

where $p_i = \mathbb{P}(\widehat{Z}^N = z_i^N)$. Then, one could want to approximate $\mathbb{E}\left[f(Z)\right]$ by $\mathbb{E}\left[f(\widehat{Z}^N)\right]$ when the first expression cannot be computed easily. For example, this case is exactly the problem one encounters when trying to price European options. We know the rate of convergence of the weak error induced by this cubature formula, i.e $\exists \alpha \in (0, 2]$, depending on the regularity of f

such that

$$\lim_{N\to+\infty} N^{\alpha} \left| \mathbb{E}\left[f(Z)\right] - \mathbb{E}\left[f(\widehat{Z}^N)\right] \right| \leqslant C_{f,X} < +\infty. \tag{6.5}$$

For more results on the rate of convergence, the value of α, we refer to [Pag18] for a survey in $\mathbb{R}^d$ and to [LMP19] for recent improved results in the one-dimensional case.

Later on, in a series of papers, among them [BP03; BPP05] extended this method to the computation of conditional expectations allowing to deal with nonlinear problems in finance and, more precisely, to the pricing and hedging of American/Bermudan options, which is the part we are interested in. These problems are of the form

$$\sup_{\tau} \mathbb{E}\left[\mathrm{e}^{-\int_0^\tau r_s^d ds} \psi_\tau(S_\tau) \right]$$

where $\left(\mathrm{e}^{-\int_0^{t_k} r_s^d ds} \psi_{t_k}(S_{t_k}) \right)_{k=0,\dots,n}$ is the obstacle function and $\tau : \Omega \to \{t_0, t_1, \dots, t_n\}$ is a stopping time for the filtration $(\mathcal{F}_{t_k})_{k\geqslant 0}$ where $\mathcal{F}_t = \sigma\big(S_s, P^d(s,T), P^f(s,T), s \leqslant t\big)$ is the natural filtration to consider because the foreign exchange rate and the zero-coupon curves are observables in the market.

In this paper, we will present two numerical solutions, motivated by the works described above, to the problem of the evaluation of Bermudan option on Foreign Exchange rate with stochastic interest rates. The paper is organised as follows. First, in Section 6.1, we introduce the diffusion models for the zero coupon curves and the foreign exchange rate we work with. In Section 6.2, we describe in details the financial product we want to evaluate: Bermudan option on foreign exchange rate. In this Section, we express the *Backward Dynamic Programming Principle* and study the regularity of the obstacle process and the value function. Then, in Section 6.3, we propose two numerical solutions for pricing the financial product defined above based on Product Quantization and we study the L^2-error induced by these numerical approximations. In Section 6.4, several examples are presented in order to compare the two methods presented in Section 6.3. First, we begin with plain European option, this test is carried out in order to benchmark the methods because a closed-form formula is known for the price of a European Call/Put in the 3-factor model. Then, we compare the two methods in the case of a Bermudan option with several exercise dates. Finally, in Appendix 6.A, we make some change of numéraire and in Appendix 6.B, we give the closed-form formula for the price of an European Call, in the 3-factor model, used in Section 6.4 as a benchmark.

6.1 Diffusion Models

Interest Rate Model. We shall denote by $P(t,T)$ the value at time t of one unit of the currency delivered (that is, paid) at time T, also known as a zero coupon price or discount factor. When t is today, this function can usually be derived from the market price of standard products, such as bonds and interest rate swaps in the market, along with an interpolation

scheme (for the dates different than the maturities of the market rates used). In a simple single-curve framework, the derivation of the initial curve, that is, the zero coupons $P(0,T)$ for $T > 0$ is rather simple, through relatively standard methods of curve stripping. In more enhanced frameworks accounting for multiple yield curves such as having different for curves for discounting and forward rates, those methods are somewhat more demanding but still relatively straightforward. We focus herein on the simple single-curve framework.

In our case we are working with financial products on Foreign Exchange (FX) rates between the domestic and the foreign currency, hence we will be working with zero coupons in the domestic currency denoted by $P^d(t,T)$ and zero coupons in the foreign currency denoted by $P^f(t,T)$. The diffusion of the domestic zero-coupon curve under the domestic risk-neutral probability $\mathbb{P}$ is given by

$$\frac{dP^d(t,T)}{P^d(t,T)} = r_t^d dt + \sigma_d(T-t)dW_t^d$$

where W^d is a $\mathbb{P}$-Brownian Motion, r_t^d is the domestic instantaneous rate at time t and σ_d is the volatility for the domestic zero coupon curve. For the foreign zero-coupon curve, the diffusion is given, under the foreign neutral probability $\widetilde{\mathbb{P}}$, by

$$\frac{dP^f(t,T)}{P^f(t,T)} = r_t^f dt + \sigma_f(T-t)d\widetilde{W}_t^f$$

where $\widetilde{W}^f$ is a $\widetilde{\mathbb{P}}$-Brownian Motion, r_t^f is the foreign instantaneous rate at time t and σ_f is the volatility for the foreign zero coupon curve. The two probabilities $\widetilde{\mathbb{P}}$ and $\mathbb{P}$ are supposed to be equivalent, i.e $\widetilde{\mathbb{P}} \sim \mathbb{P}$ and it exists ρ_{df} defined as limit of the quadratic variation $\langle W^d, \widetilde{W}^f \rangle_t = \rho_{df} t$.

Remarks. Such a framework to model random yield curves has been quite popular with practitioners due to its elegance, simplicity and intuitive understanding of rates dynamics through time yet providing a comprehensive and consistent modelling of an entire yield curve through time. Indeed, it is mathematically and numerically easily tractable. It carries no path dependency and allows the handling of multiple curves for a given currency as well as multiple currencies – and their exchange rates – as well as equities (when one wishes to account for random interest rates). It allows negative rates and can be refined by adding factors (Brownian motions).

However, it cannot easily cope with smile or non-normally distributed shocks or with internal curve "oddities" or specifics such as different volatilities for different swap tenors within the same curve dynamics. Nonetheless, our aim being to propose a model and a numerical method which make possible to produce risk computations (such as xVA's) in an efficient way, these properties are of little importance. That said, when it comes to deal with accounting for random rates in long-dated derivatives valuations, its benefits far outweigh its limitations and its use for such applications is popular, see [NP14] for the pricing of swaptions, [Pit05] for PRDCs...

Foreign Exchange Model. The diffusion of the foreign exchange (FX) rate defined under the domestic risk-neutral probability is

$$\frac{dS_t}{S_t} = (r_t^d - r_t^f)dt + \sigma_S dW_t^S$$

with W_t^S a $\mathbb{P}$-Brownian Motion under the domestic risk-neutral probability such that their exist ρ_{Sd} and ρ_{Sf} defined as limit of the quadratic variation $\langle W^S, W^d\rangle_t = \rho_{Sd}t$ and $\langle W^S, \widetilde{W}^f\rangle_t = \rho_{Sf}t$, respectively.

Finally, the processes, expressed in the domestic risk-neutral probability $\mathbb{P}$, are

$$\begin{cases} \dfrac{dP^d(t,T)}{P^d(t,T)} = r_t^d dt + \sigma_d(T-t)dW_t^d \\ \qquad\dfrac{dS_t}{S_t} = (r_t^d - r_t^f)dt + \sigma_S dW_t^S \\ \dfrac{dP^f(t,T)}{P^f(t,T)} = \big(r_t^f - \rho_{Sf}\sigma_S\sigma_f(T-t)\big)dt + \sigma_f(T-t)dW_t^f \end{cases} \tag{6.6}$$

where W^f, defined by $dW_s^f = d\widetilde{W}_s^f + \rho_{Sf}\sigma_S ds$, is a $\mathbb{P}$-Brownian motion, as shown in Appendix 6.A. Using Itô's formula, we can explicitly express the processes

$$\begin{cases} P^d(t,T) = P^d(0,T)\exp\left(\displaystyle\int_0^t \left(r_s^d - \frac{\sigma_d^2(T-s)^2}{2}\right)ds + \sigma_d\int_0^t (T-s)dW_s^d\right) \\ \qquad S_t = S_0\exp\left(\displaystyle\int_0^t \left(r_s^d - r_s^f - \frac{\sigma_S^2}{2}\right)ds + \sigma_S W_t^S\right) \\ P^f(t,T) = P^f(0,T)\exp\left(\displaystyle\int_0^t \left(r_s^f - \rho_{Sf}\sigma_S\sigma_f(T-s) - \frac{\sigma_f^2(T-s)^2}{2}\right)ds + \sigma_f\int_0^t (T-s)dW_s^f\right) \end{cases}.$$

From these equations, we deduce $\exp\left(-\int_0^t r_s^d ds\right)$ and $\exp\left(-\int_0^t r_s^f ds\right)$, by taking $T = t$ and using that $P^d(t,t) = P^f(t,t) = 1$, it follows that

$$\begin{cases} \exp\left(-\displaystyle\int_0^t r_s^d ds\right) = \varphi_d(t)\exp\left(\sigma_d\int_0^t (t-s)dW_s^d\right) \\ \exp\left(-\displaystyle\int_0^t r_s^f ds\right) = \varphi_f(t)\exp\left(\sigma_f\int_0^t (t-s)dW_s^f\right), \end{cases}$$

where

$$\varphi_d(t) = P^d(0,t)\exp\left(-\sigma_d^2\int_0^t \frac{(t-s)^2}{2}ds\right) \tag{6.7}$$

and

$$\varphi_f(t) = P^f(0,t)\exp\left(-\int_0^t \left(\rho_{Sf}\sigma_S\sigma_f(t-s) + \frac{\sigma_f^2(t-s)^2}{2}\right)ds\right). \tag{6.8}$$

These expressions for the domestic and the foreign discount factors will be useful in the following sections of the paper.

6.2 Bermudan options

6.2.1 Product Description

Let $(\Omega, \mathcal{A}, \mathbb{P})$ our domestic risk neutral probability space. We want to evaluate the price of a Bermudan option on the FX rate S_t defined by

$$S_t = \frac{1}{\exp\left(-\int_0^t r_s^d ds\right)} S_0 \varphi_f(t) \exp\left(-\frac{\sigma_S^2}{2}t + \sigma_S W_t^S + \sigma_f \int_0^t (t-s) dW_s^f\right)$$

with

$$\exp\left(-\int_0^t r_s^d ds\right) = \varphi_d(t) \exp\left(\sigma_d \int_0^t (t-s) dW_s^d\right)$$

where the owner of the financial product can exercise its option at predetermined dates $t_0, t_1, \cdots, t_n$ with payoff ψ_{t_k} at date t_k, where $t_0 = 0$.

At a given time t, the observables in the market are the foreign exchange rate S_t and the zero-coupon curves $\left(P^d(t,T)\right)_{T \geqslant t}$ and $\left(P^f(t,T)\right)_{T \geqslant t}$, hence the natural filtration to consider is

$$\mathcal{F}_t = \sigma\big(S_s, P^d(s,T), P^f(s,T), s \leqslant t\big) = \sigma\big(W_s^S, W_s^d, W_s^f, s \leqslant t\big). \tag{6.9}$$

Let $\tau : \Omega \to \{t_0, t_1, \ldots, t_n\}$ a stopping time for the filtration $(\mathcal{F}_{t_k})_{k \geqslant 0}$ and $\mathcal{T}$ the set of all stopping times for the filtration $(\mathcal{F}_{t_k})_{k \geqslant 0}$. In this paper, we consider problems where the horizon is finite then we define $\mathcal{T}_k^n$, the set of all stopping times taking finite values

$$\mathcal{T}_k^n = \big\{\tau \in \mathcal{T}, \mathbb{P}(t_k \leqslant \tau \leqslant t_n) = 1\big\}. \tag{6.10}$$

Hence, the price at time t_k of the Bermudan option is given by

$$V_k = \sup_{\tau \in \mathcal{T}_k^n} \mathbb{E}\left[\, \mathrm{e}^{-\int_0^\tau r_s^d ds}\, \psi_\tau(S_\tau) \mid \mathcal{F}_{t_k} \right]$$

and V_k is called the *Snell envelope* of the obstacle process $\left(\mathrm{e}^{-\int_0^{t_k} r_s^d ds}\, \psi_{t_k}(S_{t_k})\right)_{k=0:n}$ such that

$$\mathbb{E}\left[\psi_{t_k}(S_{t_k})^2\right] < +\infty, \quad \forall k = 0, \ldots, n. \tag{6.11}$$

Remark. The financial products we consider in the applications are PRDC. Their payoffs (see Figure 6.1) have the following expression

$$\psi_{t_k}(x) = \min\left(\max\left(\frac{C_f(t_k)}{S_0}x - C_d(t_k), \text{Floor}(t_k)\right), \text{Cap}(t_k)\right) \tag{6.12}$$

where $\text{Floor}(t_k)$ and $\text{Cap}(t_k)$ are the floor and cap values chosen at the creation of the product, as well as $C_f(t_k)$ and $C_d(t_k)$ that are the coupons value we wish to compare to the foreign and the domestic currency, respectively.

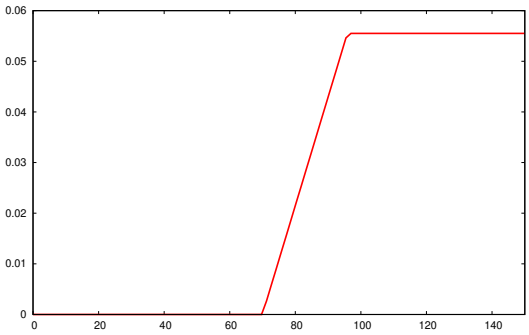


Fig. 6.1 *Example of a PRDC payoff* $\psi_{t_k}(S_{t_k}) = \min\left(\left(0.189\frac{S_{t_k}}{88.17} - 0.15\right)_+, 0.0555\right)$ *at time* t_k.

The interesting feature of such functions is that their (right) derivative have a compact support.

6.2.2 Backward Dynamic Programming Principle

V_k can also be defined recursively by

$$\begin{cases} V_n = \mathrm{e}^{-\int_0^{t_n} r_s^d ds}\,\psi_n(S_{t_n}), \\ V_k = \max\left(\mathrm{e}^{-\int_0^{t_k} r_s^d ds}\,\psi_k(S_{t_k}), \mathbb{E}[V_{k+1} \mid \mathcal{F}_{t_k}]\right), \qquad 0 \leqslant k \leqslant n-1 \end{cases} \tag{6.13}$$

and this representation is called the *Backward Dynamic Programming Principle* (BDPP).

First, noticing that the obstacle process $\mathrm{e}^{-\int_0^t r_s^d ds}\,\psi_t(S_t)$ can be rewritten as a function h_t of two processes X_t and Y_t such that

$$h_t(X_t, Y_t) = \mathrm{e}^{-\int_0^t r_s^d ds}\,\psi_t(S_t)$$

where h is given by

$$h_t(x, y) = \varphi_d(t)\,\mathrm{e}^{-y}\,\psi_t\left(S_0\frac{\varphi_f(t)}{\varphi_d(t)}\,\mathrm{e}^{-\sigma_S^2 t/2 + x + y}\right) \tag{6.14}$$

and (X, Y) is defined by

$$(X_t, Y_t) = \left(\sigma_S W_t^S + \sigma_f \int_0^t (t-s) dW_s^f, -\sigma_d \int_0^t (t-s) dW_s^d\right). \tag{6.15}$$

Now, in order to alleviate notations, we denote by $X_k = X_{t_k}$, $W_k^f = W_{t_k}^f$, $Y_k = Y_{t_k}$, $W_k^d = W_{t_k}^d$, $W_k^S = W_{t_k}^S$ and $h_k = h_{t_k}$.

Using this new form, the Snell envelope becomes

$$V_k = \sup_{\tau \in \mathcal{T}_k^n} \mathbb{E}\left[h_\tau(X_\tau, Y_\tau) \mid \mathcal{F}_{t_k}\right]$$

and the *Backward Dynamic Programming Principle* (6.13) rewrites

$$\begin{cases} V_n = h_n(X_n, Y_n), \\ V_k = \max\left(h_k(X_k, Y_k), \mathbb{E}\left[V_{k+1} \mid \mathcal{F}_{t_k}\right]\right), \qquad 0 \leqslant k \leqslant n-1. \end{cases} \tag{6.16}$$

Second, in order to solve the problem theoretically by dynamic programming it is required to associate a $\mathcal{F}_t$-Markov process to this problem and in our case, the simplest of them (i.e. of minimal dimension) is (X_t, W_t^f, Y_t, W_t^d) which is $\mathcal{F}_t$-adapted and a Markov process because

$$\begin{cases} X_{k+1} = X_k + \sigma_f \delta W_k^f + \sigma_S \displaystyle\int_{t_k}^{t_{k+1}} dW_s^S + \sigma_f \int_{t_k}^{t_{k+1}} (t_{k+1} - s) dW_s^f \\ W_{k+1}^f = W_k^f + \displaystyle\int_{t_k}^{t_{k+1}} dW_s^f \\ Y_{k+1} = Y_k - \sigma_d \delta W_k^d - \sigma_d \displaystyle\int_{t_k}^{t_{k+1}} (t_{k+1} - s) dW_s^d \\ W_{k+1}^d = W_k^d + \displaystyle\int_{t_k}^{t_{k+1}} dW_s^d \end{cases}$$

where $\delta = \frac{T}{n}$ and can be written as

$$\begin{cases} X_{k+1} = X_k + \sigma_f \delta W_k^f + G_{k+1}^1 \\ W_{k+1}^f = W_k^f + G_{k+1}^2 \\ Y_{k+1} = Y_k - \sigma_d \delta W_k^d + G_{k+1}^3 \\ W_{k+1}^d = W_k^d + G_{k+1}^4, \end{cases} \tag{6.17}$$

where the increments are normally distributed

$$\begin{pmatrix} G_{k+1}^1 \\ G_{k+1}^2 \\ G_{k+1}^3 \\ G_{k+1}^4 \end{pmatrix} \sim \mathcal{N}\left(\mu_{k+1}, \Sigma_{k+1}\right) \tag{6.18}$$

with

$$\mu_{k+1} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \qquad \text{and} \qquad \Sigma_{k+1} = \left(\mathbb{C}\mathrm{ov}\left(G_{k+1}^i, G_{k+1}^j\right) \right)_{i,j=1:4}. \tag{6.19}$$

One notices that $\left((G_k^1, G_k^2, G_k^3, G_k^4)\right)_{k=1\ldots,n}$ are i.i.d. Based on Equation (6.17), we deduce the Markov process transition of (X_k, W_k^f, Y_k, W_k^d), for any integrable function $f : \mathbb{R}^4 \to \mathbb{R}$, given by

$$Pf(x,u,y,v) = \mathbb{E}\left[f(x + \sigma_f \delta u + G_{k+1}^1, u + G_{k+1}^2, y - \sigma_d \delta v + G_{k+1}^3, v + G_{k+1}^4)\right]. \tag{6.20}$$

Remark. Using the Markov process (X, W^f, Y, W^d) newly defined, we rewrite the filtration $\mathcal{F}_t$ as

$$\mathcal{F}_t = \sigma\big(W_s^S, W_s^d, W_s^f, s \leqslant t\big) = \sigma\big(X_s, W_s^f, Y_s, W_s^d, s \leqslant t\big). \tag{6.21}$$

Then, using the new expression for the filtration and the Markov property of (X_k, W_k^f, Y_k, W_k^d), the BDPP (6.16) reads as follows,

$$\begin{cases} V_n = h_n(X_n, Y_n), \\ V_k = \max\left(h_k(X_k, Y_k), \mathbb{E}\left[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\right]\right), \qquad 0 \leqslant k \leqslant n-1. \end{cases} \tag{6.22}$$

Moreover, by backward induction we get $V_k = v_k(X_k, W_k^f, Y_k, W_k^d)$ where

$$\begin{cases} v_n(X_n, W_n^f, Y_n, W_n^d) = h_n(X_n, Y_n), \\ v_k(X_k, W_k^f, Y_k, W_k^d) = \max\left(h_k(X_k, Y_k), Pv_{k+1}(X_k, W_k^f, Y_k, W_k^d)\right), \qquad 0 \leqslant k \leqslant n-1. \end{cases} \tag{6.23}$$

Payoff regularity. First, we look at the regularity of the payoff. The next proposition will then allow us to study the regularity of the value function through the propagation of the local Lipschitz property by the transition of the Markov process.

Proposition 6.2.1. *If ψ_{t_k} is are Lipschitz continuous with Lipschitz coefficient $[\psi_{t_k}]_{Lip}$ with compactly supported (right) derivative (such as the payoff defined in (6.12)) then $h_k(x,y)$ given*

by (6.14) is locally Lipschitz continuous, for every $x, x', y, y' \in \mathbb{R}$

$$|h_k(x,y) - h_k(x',y')| \leqslant \mathrm{e}^{|y|\vee|y'|}\left([\bar{\psi}_k]_{Lip}|x-x'| + (\varphi_d(t_k)\|\psi_{t_k}\|_\infty + [\bar{\psi}_k]_{Lip})|y-y'|\right) \tag{6.24}$$

with $[\bar{\psi}_k]_{Lip} = [\psi_{t_k}]_{Lip} S_0 \varphi_f(t_k)\,\mathrm{e}^{-\sigma_S^2 t_k/2}\,\|\psi'_{t_k}\|_\infty\,\mathrm{e}^c$ *with* ψ'_{t_k} *the right derivative of* ψ_{t_k}.

Proof. Let g_k be defined by

$$g_k(x,y) = \psi_{t_k}\left(S_0 \frac{\varphi_f(t_k)}{\varphi_d(t_k)}\,\mathrm{e}^{-\sigma_S^2 t_k/2 + x + y}\right). \tag{6.25}$$

As ψ'_{t_k} has a compact support, then it exists $c \in \mathbb{R}$ such that

$$|(\psi_{t_k}(\mathrm{e}^x))'| = |\,\mathrm{e}^x\,\psi'_{t_k}(\mathrm{e}^x)| \leqslant \|\psi'_{t_k}\|_\infty \sup_{x \in \mathrm{supp}\,\psi'_{t_k}} \mathrm{e}^x \leqslant \|\psi'_{t_k}\|_\infty\,\mathrm{e}^c\,. \tag{6.26}$$

Hence

$$|g_k(x,y) - g_k(x',y')| \leqslant \frac{[\bar{\psi}_k]_{Lip}}{\varphi_d(t_k)}\left(|x-x'| + |y-y'|\right) \tag{6.27}$$

with $[\bar{\psi}_k]_{Lip} = [\psi_{t_k}]_{Lip} S_0 \varphi_f(t_k)\,\mathrm{e}^{-\sigma_S^2 t_k/2}\,\|\psi'_{t_k}\|_\infty\,\mathrm{e}^c$. Then for every $x, x', y, y' \in \mathbb{R}$, we have

$$\begin{aligned}
|h_k(x,y) - h_k(x',y')| &= \varphi_d(t_k)\big|\,\mathrm{e}^{-y}\,g_k(x,y) - \mathrm{e}^{-y'}\,g_k(x',y')\big| \\
&\leqslant \varphi_d(t_k)\Big(\big|\,\mathrm{e}^{-y}\,g_k(x,y) - \mathrm{e}^{-y'}\,g_k(x,y)\big| + \big|\,\mathrm{e}^{-u'}\,g_k(x,y) - \mathrm{e}^{-y'}\,g_k(x',y')\big|\Big) \\
&\leqslant \varphi_d(t_k)\Big(\big|\,\mathrm{e}^{-y} - \mathrm{e}^{-y'}\,\big| \cdot \|\psi_{t_k}\|_\infty + \mathrm{e}^{-y'}\,\big|g_k(x,y) - g_k(x',y')\big|\Big) \\
&\leqslant \mathrm{e}^{|y|\vee|y'|}\left([\bar{\psi}_k]_{Lip}|x-x'| + (\varphi_d(t_k)\|\psi_{t_k}\|_\infty + [\bar{\psi}_k]_{Lip})|y-y'|\right).
\end{aligned} \tag{6.28}$$

□

The next Lemma shows that the transition of the Markov process propagates the local Lipschitz continuity of a function f. This result will be helpful to estimate the error induced by the numerical approximation (6.23).

Lemma 6.2.2. *Let* $Pf(x,u,y,v) = \mathbb{E}\left[f(x + \sigma_f\delta u + G^1, u + G^2, y - \sigma_d\delta v + G^3, v + G^4)\right]$ *be a Markov kernel. If the function* f *satisfies the following local Lipschitz property,*

$$\begin{aligned}
|f(x,u,y,v) - f(x',u',y',v')| \leqslant \big(A|x-x'| + B|u-u'| + C|y-y'| + D|v-v'|\big) \\
\times\,\mathrm{e}^{|y|\vee|y'|+b|v|\vee|v'|}
\end{aligned} \tag{6.29}$$

then

$$\begin{aligned}
|Pf(x,u,y,v) - Pf(x',u',y',v')| \leqslant \big(\widetilde{A}|x-x'| + \widetilde{B}|u-u'| + \widetilde{C}|y-y'| + \widetilde{D}|v-v'|\big) \\
\times\,\mathrm{e}^{|y|\vee|y'|+\widetilde{b}|v|\vee|v'|}\,.
\end{aligned} \tag{6.30}$$

Proof. It follows from Jensen's inequality and our assumption on f

$$\begin{aligned}
&|Pf(x,u,y,v) - Pf(x',u',y',v')| \\
&\qquad \leqslant \mathbb{E}\Big[\big|f(x+\sigma_f\delta u + G^1, u+G^2, y-\sigma_d\delta v + G^3, v+G^4) \\
&\qquad\qquad - f(x'+\sigma_f\delta u' + G^1, u'+G^2, y'-\sigma_d\delta v' + G^3, v'+G^4)\big|\Big] \\
&\qquad \leqslant \big(A|x-x'| + (B+A\sigma_f\delta)|u-u'| + C|y-y'| + (D+C\sigma_d\delta)|v-v'|\big) \\
&\qquad\qquad \times \mathrm{e}^{|y|\vee|y'| + (b+\sigma_d\delta)|v|\vee|v'|}\,\mathbb{E}\big[\,\mathrm{e}^{|G^3|+b|G^4|}\,\big] \\
&\qquad \leqslant \big(\widetilde{A}|x-x'| + \widetilde{B}|u-u'| + \widetilde{C}|y-y'| + \widetilde{D}|v-v'|\big) \\
&\qquad\qquad \times \mathrm{e}^{|y|\vee|y'| + \widetilde{b}|v|\vee|v'|}
\end{aligned} \tag{6.31}$$

where

$$\widetilde{A} = A\,\mathbb{E}[\kappa], \qquad \widetilde{B} = (B + A\sigma_f\delta)\,\mathbb{E}[\kappa] \tag{6.32}$$

and

$$\widetilde{C} = C\,\mathbb{E}[\kappa], \qquad \widetilde{D} = (D + C\sigma_d\delta)\,\mathbb{E}[\kappa], \qquad \widetilde{b} = b + \sigma_d\delta \tag{6.33}$$

with $\kappa = \exp(|G^3| + b|G^4|)$ and $\mathbb{E}[\kappa] < +\infty$. □

Value function regularity. If the functions $(\psi_{t_k})_{k=0:n}$ are defined as in Equation (6.12) then $v_n(x,u,y,v)$ preserves a local Lipschitz property. Hence, for every $x, x', u, u', y, y', v, v' \in \mathbb{R}$,

$$\begin{aligned}
|v_n(x,u,y,v) - v_n(x',u',y',v')| \leqslant \big(A_n|x-x'| + B_n|u-u'| + C_n|y-y'| + D_n|v-v'|\big) \\
\times \mathrm{e}^{|y|\vee|y'| + b_n|v|\vee|v'|}
\end{aligned} \tag{6.34}$$

where

$$A_n = [\bar{\psi}_n]_{Lip}, \qquad B_n = 0, \qquad C_n = \varphi_d(t_n)\|\psi_n\|_\infty + [\bar{\psi}_n]_{Lip}, \qquad D_n = 0, \qquad b_n = 0 \tag{6.35}$$

with $[\bar{\psi}_n]_{Lip} = [\psi_{t_n}]_{Lip} S_0 \varphi_f(t_n) \exp(-\sigma_S^2 t_n/2)\|\psi'_{t_n}\|_\infty \,\mathrm{e}^c$. Using now Lemma 6.2.2 recursively and the elementary inequality $\max(a, b+c) \leqslant \max(a,b) + c$ (as $x \mapsto \max(a,x)$ is 1-Lipschitz),

we have

$$\begin{aligned}
&|v_k(x,u,y,v) - v_k(x',u',y',v')| \\
&\qquad \leqslant \max\big(|h_k(x,y) - h_k(x',y')|, |Pv_{k+1}(x,u,y,v) - Pv_{k+1}(x',u',y',v')|\big) \\
&\qquad \leqslant \max\Bigg(\mathrm{e}^{|y|\vee|y'|}\big([\bar{\psi}_k]_{Lip}|x-x'| + (\varphi_d(t_k)\|\psi_{t_k}\|_\infty + [\bar{\psi}_k]_{Lip})|y-y'|\big) \\
&\qquad\qquad , \big(\widetilde{A}_k|x-x'| + \widetilde{B}_k|u-u'| + \widetilde{C}_k|y-y'| + \widetilde{D}_k|v-v'|\big) \\
&\qquad\qquad \times \mathrm{e}^{|y|\vee|y'|+\widetilde{b}_k|v|\vee|v'|} \Bigg) \\
&\qquad \leqslant \big(A_k|x-x'| + B_k|u-u'| + C_k|y-y'| + D_k|v-v'|\big) \\
&\qquad\qquad \times \mathrm{e}^{|y|\vee|y'|+b_k|v|\vee|v'|}
\end{aligned} \tag{6.36}$$

where

$$A_k = [\bar{\psi}_k]_{Lip} \vee \big(A_{k+1}\,\mathbb{E}[\kappa_{k+1}]\big), \qquad B_k = (B_{k+1} + A_{k+1}\sigma_f\delta)\,\mathbb{E}[\kappa_{k+1}], \qquad b_k = b_{k+1} + \sigma_d\delta \tag{6.37}$$

and

$$C_k = \big(\varphi_d(t_k)\|\psi_{t_k}\|_\infty + [\bar{\psi}_k]_{Lip}\big) \vee \big(C_{k+1}\,\mathbb{E}[\kappa_{k+1}]\big), \qquad D_k = (D_{k+1} + C_{k+1}\sigma_d\delta)\,\mathbb{E}[\kappa_{k+1}] \tag{6.38}$$

with $\kappa_{k+1} = \exp(|G^3_{k+1}| + b_{k+1}|G^4_{k+1}|)$. Or equivalently

$$A_k = \max_{l\geqslant k}\Bigg([\bar{\psi}_l]_{Lip}\prod_{j=k+1}^{l}\mathbb{E}[\kappa_j]\Bigg), \qquad B_k = \sigma_f\frac{T}{n}\sum_{l=k+1}^{n}\max_{l\leqslant i\leqslant n}\Bigg([\bar{\psi}_i]_{Lip}\prod_{j=k+1}^{i}\mathbb{E}[\kappa_j]\Bigg) \tag{6.39}$$

and

$$\begin{aligned}
C_k &= \max_{l\geqslant k}\Bigg((\varphi_d(t_l)\|\psi_l\|_\infty + [\bar{\psi}_l]_{Lip})\prod_{j=k+1}^{l}\mathbb{E}[\kappa_j]\Bigg), \\
D_k &= \sigma_d\frac{T}{n}\sum_{l=k+1}^{n}\max_{l\leqslant i\leqslant n}\Bigg((\varphi_d(t_i)\|\psi_i\|_\infty + [\bar{\psi}_i]_{Lip})\prod_{j=k+1}^{i}\mathbb{E}[\kappa_j]\Bigg)
\end{aligned} \tag{6.40}$$

with

$$b_k = \sigma_d T\Big(1 - \frac{k-1}{n}\Big). \tag{6.41}$$

6.3 Bermudan pricing using Optimal Quantization

In this section, we propose two numerical solutions based on Product Optimal Quantization for the pricing of Bermudan options on the FX rate S_t. First, we remind briefly what is an optimal quantizer and what we mean by a product quantization tree. Second, we present a first numerical solution, based on quantization of the Markovian tuple (X, W^f, Y, W^d), to

solve the numerical problem (6.22) and detail the L^2-error induced by this approximation. However, remember that we are looking for a method that makes possible to compute xVA's risk measures in a reasonable time but this solution can be too time consuming in practice due to the dimensionality of the quantized processes. That is why we present a second numerical solution which reduces the dimensionality of the problem by considering an approximate problem, based on quantization of the non-Markovian couple (X, Y), introducing a systematic error induced by the non-markovianity and we study the L^2-error produced by this approximation.

6.3.1 About Optimal Quantization

Theoretical background (the one-dimensional case). The aim of Optimal Quantization is to determine Γ_N, a set with cardinality at most N, which minimises the quantization error among all such sets Γ. We place ourselves in the one-dimensional case. Let Z be an $\mathbb{R}$-valued random variable with distribution $\mathbb{P}_Z$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that $Z \in L^2_{\mathbb{R}}$.

Definition 6.3.1. Let $\Gamma_N = \{z_1, \dots, z_N\} \subset \mathbb{R}$ be a subset of size N, called N-quantizer. A Borel partition $\big(C_i(\Gamma_N)\big)_{i\in[\![1,N]\!]}$ of $\mathbb{R}$ is a Voronoï partition of $\mathbb{R}$ induced by the N-quantizer Γ_N if, for every $i = \{1, \cdots, N\}$,

$$C_i\big(\Gamma_N\big) \subset \Big\{\xi \in \mathbb{R}, |\xi - z_i| \leqslant \min_{j\neq i} |\xi - z_j|\Big\}.$$

The Borel sets $C_i(\Gamma_N)$ are called Voronoï cells of the partition induced by Γ_N.

One can always consider that the quantizers are ordered: $z_1 < z_2 < \cdots < z_{N-1} < z_N$ and in that case the Voronoï cells are given by

$$C_k\big(\Gamma_N\big) = \big(z_{k-1/2}, z_{k+1/2}\big], \qquad k \in [\![1, N-1]\!], \qquad C_N\big(\Gamma_N\big) = \big(z_{N-1/2}, z_{N+1/2}\big)$$

where $\forall k \in \{2, \cdots, N\}$, $\quad z_{k-1/2} = \frac{z_{k-1}+z_k}{2}$ and $z_{1/2} = \inf\big(\operatorname{supp}(\mathbb{P}_Z)\big)$ and $z_{N+1/2} = \sup\big(\operatorname{supp}(\mathbb{P}_Z)\big)$.

Definition 6.3.2. Let $\Gamma_N = \{z_1, \dots, z_N\}$ be an N-quantizer. The nearest neighbour projection $\operatorname{Proj}_{\Gamma_N} : \mathbb{R} \to \{z_1, \dots, z_N\}$ induced by a Voronoï partition $\big(C_i(\Gamma_N)\big)_{i\in\{1,\cdots,N\}}$ is defined by

$$\forall \xi \in \mathbb{R}, \qquad \operatorname{Proj}_{\Gamma_N}(\xi) = \sum_{i=1}^{N} z_i \, \mathbb{1}_{\xi \in C_i(\Gamma_N)}\,.$$

Hence, we can define the quantization of Z as the nearest neighbour projection of Z onto Γ_N by composing $\operatorname{Proj}_{\Gamma_N}$ and X

$$\widehat{Z}^{\Gamma_N} = \operatorname{Proj}_{\Gamma_N}(Z) = \sum_{i=1}^{N} z_i \, \mathbb{1}_{Z \in C_i(\Gamma_N)}\,.$$

In order to alleviate notations, we write $\widehat{Z}^N$ from now on in place of $\widehat{Z}^{\Gamma_N}$.

Now that we have defined the quantization of Z, we explain where does the term "optimal" comes from in the term optimal quantization. First, we define the quadratic distortion function.

Definition 6.3.3. The L^2-mean quantization error induced by the quantizer $\widehat{Z}^N$ is defined as

$$\|Z-\widehat{Z}^N\|_2=\left(\mathbb{E}\left[\min_{i\in\{1,\cdots,N\}}|Z-z_i|^2\right]\right)^{1/2}=\left(\int_{\mathbb{R}}\min_{i\in\{1,\cdots,N\}}|\xi-z_i|^2\,\mathbb{P}_Z(d\xi)\right)^{1/2}. \tag{6.42}$$

It is convenient to define the quadratic distortion function at level N as the squared mean quadratic quantization error on $(\mathbb{R})^N$:

$$\mathcal{Q}_{2,\mathrm{N}}:z=(z_1,\ldots,z_N)\longmapsto\mathbb{E}\left[\min_{i\in\{1,\cdots,N\}}|Z-z_i|^2\right]=\|Z-\widehat{Z}^N\|_2^2.$$

Remark. All these definitions can be extended to the L^p case. For example the L^p-mean quantization error induced by a quantizer of size N is

$$\|Z-\widehat{Z}^N\|_p=\left(\mathbb{E}\left[\min_{i\in\{1,\cdots,N\}}|Z-z_i|^p\right]\right)^{1/p}=\left(\int_{\mathbb{R}}\min_{i\in\{1,\cdots,N\}}|Z-z_i|^p\,\mathbb{P}_Z(d\xi)\right)^{1/p}. \tag{6.43}$$

The existence of a N-tuple $z^{(N)}=(z_1,\ldots,z_N)$ minimizing the quadratic distortion function $\mathcal{Q}_{2,\mathrm{N}}$ at level N has been shown and its associated quantizer $\Gamma_N=\{z_i,i\in\{1,\cdots,N\}\}$ is called an optimal quadratic N-quantizer, see e.g. [Pag18] for further details and references. We now turn to the asymptotic behaviour in N of the quadratic mean quantization error. The next Theorem, known as Zador's Theorem, provides the sharp rate of convergence of the L^p-mean quantization error.

Theorem 6.3.4. *(Zador's Theorem) Let $p\in(0,+\infty)$.*

(a) Sharp rate. *Let $Z\in L^{p+\delta}_{\mathbb{R}}(\mathbb{P})$ for some $\delta>0$. Let $\mathbb{P}_Z(d\xi)=\varphi(\xi)\cdot\lambda(d\xi)+\nu(d\xi)$, where $\nu\perp\lambda$ i.e. denotes the singular part of $\mathbb{P}_Z$ with respect to the Lebesgue measure λ on $\mathbb{R}$. Then,*

$$\lim_{N\to+\infty}N\min_{\Gamma_N\subset\mathbb{R},|\Gamma_N|\leqslant N}\|Z-\widehat{Z}^N\|_p=\frac{1}{2^p(p+1)}\left[\int_{\mathbb{R}}\varphi^{\frac{1}{1+p}}d\lambda\right]^{1+\frac{1}{p}}. \tag{6.44}$$

(b) Non asymptotic upper-bound. *Let $\delta>0$. There exists a real constant $C_{1,p,\delta}\in(0,+\infty)$ such that, for every $\mathbb{R}$-valued random variable Z,*

$$\forall N\geqslant 1,\qquad\min_{\Gamma_N\subset\mathbb{R},|\Gamma_N|\leqslant N}\|Z-\widehat{Z}^N\|_p\leqslant C_{1,p,\delta}\sigma_{\delta+p}(Z)N^{-1} \tag{6.45}$$

where, for $r\in(0,+\infty)$, $\sigma_r(Z)=\min_{a\in\mathbb{R}}\|Z-a\|_r<+\infty$.

The next result answers to the following question: what can be said about the convergence rate of $\mathbb{E}\left[|Z-\widehat{Z}^N|^{2+\beta}\right]$, knowing that $\widehat{Z}^N$ is a quadratic optimal quantization?

This problem is known as the distortion mismatch problem and has been first addressed by [GLP08] and the results have been extended in Theorem 4.3 of [PS18a].

Theorem 6.3.5. *[L^r-L^s-distortion mismatch] Let $Z : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}$ be a random variable and let $r \in (0, +\infty)$. Assume that the distribution $\mathbb{P}_Z$ of Z has a non-zero absolutely continuous component with density φ. Let $(\Gamma_N)_{N\geqslant 1}$ be a sequence of L^r-optimal grids. Let $s \in (r, r+1)$. If*

$$Z \in L^{\frac{s}{1+r-s}+\delta}(\Omega, \mathcal{A}, \mathbb{P}) \tag{6.46}$$

for some $\delta > 0$, then

$$\limsup_N N\|Z - \widehat{Z}^N\|_s < +\infty. \tag{6.47}$$

Product Quantization. Now, let $Z = (Z^\ell)_{\ell=1:d}$ be an $\mathbb{R}^d$-valued random vector with distribution $\mathbb{P}_Z$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. There are two approaches if one wishes to scale to higher dimensions. Either one applies the above framework directly to the random vector Z and build an optimal quantizer of Z, or one may consider separately each component Z^ℓ independently, build a one-dimensional optimal quantization $\widehat{Z}^\ell$, of size N^ℓ, with quantizer $\Gamma_\ell^{N^\ell} = \{z_{i_\ell}^\ell, i_\ell \in \{1, \cdots, N^\ell\}\}$ and then build the product quantizer $\Gamma^N = \prod_{\ell=1}^d \Gamma_l^{N^\ell}$ of size $N = N^1 \times \cdots \times N^d$ defined by

$$\Gamma^N = \{(z_{i_1}^1, \cdots, z_{i_\ell}^\ell, \cdots, z_{i_d}^d), \quad i_\ell \in \{1, \cdots, N_\ell\}, \quad \ell \in \{1, \cdots, d\}\}. \tag{6.48}$$

In our case we chose the second approach. Indeed, it is much more flexible when dealing with normal distribution, like in our case. We do not need to solve the d-dimensional minimization problem at each time step. We only need to load precomputed optimal quantizer of standard normal distribution $\mathcal{N}(0, 1)$ and then take advantage of the stability of optimal quantization by rescaling in one dimension in the sense that if $\Gamma^N = \{z_i, 1 \leqslant i \leqslant N\}$ is optimal at level N for $\mathcal{N}(0, 1)$ then $\mu + \sigma\Gamma^N$ (with obvious notations) is optimal for $\mathcal{N}(\mu, \sigma^2)$.

Even though it exists fast methods for building optimal quantizers in two-dimension based on deterministic methods like in the one-dimensional case, when dealing with optimal quantization of bivariate Gaussian vector, we may face numerical instability when the covariance matrix is ill-conditioned: so is the case if the variance of one coordinate is relatively high compared to the second one (which is our case in this paper). This a major drawback as we are looking for a fast numerical solution able to produce prices in a few seconds and this is possible when using product optimal quantization.

Quantization Tree. Now, in place of considering a random variable Z, let $(Z_t)_{t\in[0,T]}$ be a stochastic process following a Stochastic Differential Equation (SDE)

$$Z_t = Z_0 + \int_0^t b_s(Z_s)ds + \int_0^t \sigma(s, Z_s)dW_s \tag{6.49}$$

with $Z_0 = z_0 \in \mathbb{R}^d$, W a standard Brownian motion living on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and b and σ satisfy the standard assumptions in order to ensure the existence of a strong solution of the SDE.

What we call Quantization Tree is defined, for chosen time steps $t_k = Tk/n, k = 0, \cdots, n$, by quantizers $\widehat{Z}_k$ of Z_k (Product Quantizers in our case) at dates t_k and the transition probabilities between date t_k and date t_{k+1}. Although $(\widehat{Z}_k)_k$ is no longer a Markov process we will consider the transition probabilities $\pi_{ij}^k = \mathcal{L}(\widehat{Z}_{k+1} \mid \widehat{Z}_k)$. We can apply this methodology because, with the model we consider, we know all the marginal laws of our processes at each date of interest.

In the next subsection, we present the approach based on the quantization tree previously defined that allows us to approximate the price of Bermudan options where the risk factors are driven by the 3-factor model (6.6).

6.3.2 Quantization tree approximation: Markov case

Our first idea in order to discretize (6.22) is to replace the processes by a product quantizer composed with optimal quadratic quantizers. Indeed, at each time t_k, we know the law of the processes X_k, W_k^f, Y_k and W_k^d. Then we "force" in some sense the (lost) Markov property by introducing the *Quantized Backward Dynamic Programming Principle* (QBDPP) defined by

$$\begin{cases} \widehat{V}_n = h_n(\widehat{X}_n, \widehat{Y}_n), \\ \widehat{V}_k = \max\left(h_k(\widehat{X}_k, \widehat{Y}_k), \mathbb{E}\left[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)\right]\right), \qquad 0 \leqslant k \leqslant n-1, \end{cases} \tag{6.50}$$

where for every $k = 0, \ldots, n$, $\widehat{X}_k$, $\widehat{W}_k^f$, $\widehat{Y}_k$ and $\widehat{W}_k^d$ are quadratic optimal quantizers of X_k, W_k^f, Y_k and W_k^d of size N_k^X, $N_k^{W^f}$, N_k^Y and $N_k^{W^d}$ respectively and we denote $N_k = N_k^X \times N_k^{W^f} \times N_k^Y \times N_k^{W^d}$ the size of the grid of the product quantizer.

We are interested by the error induced by the numerical algorithm defined in (6.50) and more precisely its L^2-error, with in mind that we "lost" the Markov property in the quantization procedure. Moreover, this can be circumvented as shown below.

Theorem 6.3.6. *Let the Markov transition $Pf(x,u,y,v)$ defined in (6.20) be locally Lipschitz in the sense of Lemma 6.2.2. Assume that all the payoff functions $(\psi_{t_k})_{k=0:n}$ are Lipschitz continuous with compactly supported (right) derivative. Then the L^2-error induced by the quantization approximation $(\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)$ is upper-bounded by*

$$\left\|V_k - \widehat{V}_k\right\|_2 \leqslant \left(\sum_{l=k}^{n} C_{X_l}\left\|X_l - \widehat{X}_l\right\|_{2p}^2 + C_{Y_l}\left\|Y_l - \widehat{Y}_l\right\|_{2p}^2 + C_{W_l^d}\left\|W_l^d - \widehat{W}_l^d\right\|_{2p}^2 + C_{W_l^f}\left\|W_l^f - \widehat{W}_l^f\right\|_{2p}^2\right)^{1/2}, \tag{6.51}$$

where $1 < p < 3/2$ *and* $q \geqslant 1$ *such that* $\frac{1}{p} + \frac{1}{q} = 1$ *and*

$$\begin{aligned} C_{X_l} &= [\bar{\psi}_l]_{Lip}^2 \| \, \mathrm{e}^{|Y_l| \vee |\widehat{Y}_l|} \, \|_{2q}^2 + \widetilde{A}_l^2 K_l^2, & C_{W_l^d} &= \widetilde{B}_l^2 K_l^2, \\ C_{Y_l} &= \left(\varphi_d(t_l)\|\psi_{t_l}\|_\infty + [\bar{\psi}_l]_{Lip}\right)^2 \| \, \mathrm{e}^{|Y_l| \vee |\widehat{Y}_l|} \, \|_{2q}^2 + \widetilde{C}_l^2 K_l^2, & C_{W_l^f} &= \widetilde{D}_l^2 K_l^2 \end{aligned} \tag{6.52}$$

with

$$K_l = \| \, \mathrm{e}^{|Y_l| \vee |\widehat{Y}_l| + \widetilde{b}_l |W_l^d| \vee |\widehat{W}_l^d|} \, \|_{2q}. \tag{6.53}$$

As a consequence if $\bar{N} = \min N_k$*, we have*

$$\lim_{\bar{N} \to +\infty} \left\| V_k - \widehat{V}_k \right\|_2^2 = 0. \tag{6.54}$$

Remark. From the definition of the processes X_k, W_k^f, Y_k and W_k^d, all are Gaussian random variables hence all the L^{2q}-norms in Equations (6.52) and (6.53) are finite. Indeed, let $Z \sim \mathcal{N}(0, \sigma_Z)$ a Gaussian random variable with variance σ_Z^2 and $\widehat{Z}$ an optimal quantizer of Z with cardinality N then $\forall \lambda \in \mathbb{R}_+$

$$\| \, \mathrm{e}^{\lambda |Z| \vee |\widehat{Z}|} \, \|_{2q} = \left(\mathbb{E}\left[\mathrm{e}^{2q\lambda |Z| \vee |\widehat{Z}|} \right] \right)^{\frac{1}{2q}} \leqslant \left(2\, \mathbb{E}\left[\mathrm{e}^{2q\lambda |Z|} \right] \right)^{\frac{1}{2q}} \leqslant 2^{\frac{1}{2q}} \, \mathrm{e}^{q^2 \lambda^2 \sigma_Z^2}. \tag{6.55}$$

Proof. The error between the Snell envelope and its approximation is given by

$$\begin{aligned} |V_k - \widehat{V}_k| \leqslant \max \Big(& \left| h_k(X_k, Y_k) - h_k(\widehat{X}_k, \widehat{Y}_k) \right|, \\ & \left| \mathbb{E}\left[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d) \right] - \mathbb{E}\left[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d) \right] \right| \Big) \end{aligned} \tag{6.56}$$

thus, using the local Lipschitz property of h_k established in Proposition 6.2.1 and Hölder's inequality with $p, q \geqslant 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$, the L^2-error is upper-bounded by

$$\begin{aligned} \left\| V_k - \widehat{V}_k \right\|_2^2 &\leqslant \left\| h_k(X_k, Y_k) - h_k(\widehat{X}_k, \widehat{Y}_k) \right\|_2^2 \\ &\qquad + \left\| \mathbb{E}\left[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d) \right] - \mathbb{E}\left[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d) \right] \right\|_2^2. \\ &\leqslant \| \, \mathrm{e}^{|Y_k| \vee |\widehat{Y}_k|} \, \|_{2q}^2 \left(\left(\varphi_d(t_k)\|\psi_{t_k}\|_\infty + [\bar{\psi}_k]_{Lip} \right)^2 \left\| Y_k - \widehat{Y}_k \right\|_{2p}^2 + [\bar{\psi}_k]_{Lip}^2 \left\| X_k - \widehat{X}_k \right\|_{2p}^2 \right) \\ &\qquad + \left\| \mathbb{E}\left[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d) \right] - \mathbb{E}\left[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d) \right] \right\|_2^2. \end{aligned} \tag{6.57}$$

Looking at the last term, we have

$$\begin{aligned} & \mathbb{E}\left[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d) \right] - \mathbb{E}\left[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d) \right] \\ &\qquad = \mathbb{E}\left[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d) \right] - \mathbb{E}\left[V_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d) \right] \\ &\qquad + \mathbb{E}\left[V_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d) \right] - \mathbb{E}\left[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d) \right]). \end{aligned} \tag{6.58}$$

Now, we inspect the L^2-error of each term on the right-hand side of the equality.

- For the first term, notice that

$$\mathbb{E}\left[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\right] = Pv_{k+1}(X_k, W_k^f, Y_k, W_k^d) \tag{6.59}$$

and

$$\mathbb{E}\left[V_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)\right] = Pv_{k+1}(\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d) \tag{6.60}$$

then, we directly apply Lemma 6.2.2 on the function v_{k+1} and obtain

$$\begin{aligned}
&|Pv_{k+1}(X_k, W_k^f, Y_k, W_k^d) - Pv_{k+1}(\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)| \\
&\leqslant \left(\widetilde{A}_k|X_k - \widehat{X}_k| + \widetilde{B}_k|W_k^f - \widehat{W}_k^f| + \widetilde{C}_k|Y_k - \widehat{Y}_k| + \widetilde{D}_k|W_k^d - \widehat{W}_k^d|\right) \mathrm{e}^{|Y_k| \vee |\widehat{Y}_k| + \widetilde{b}_k |W_k^d| \vee |\widehat{W}_k^d|}
\end{aligned} \tag{6.61}$$

with $\widetilde{A}_k$, $\widetilde{B}_k$, $\widetilde{C}_k$, $\widetilde{D}_k$ and $\widetilde{b}_k$ defined by (6.32) and (6.33). Hence, using Hölder's inequality with $p, q \geqslant 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$,

$$\begin{aligned}
&\left\| \mathbb{E}\left[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\right] - \mathbb{E}\left[V_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)\right] \right\|_2^2 \\
&\leqslant \left(\widetilde{A}_k^2 \|X_k - \widehat{X}_k\|_{2p}^2 + \widetilde{B}_k^2 \|W_k^f - \widehat{W}_k^f\|_{2p}^2 + \widetilde{C}_k^2 \|Y_k - \widehat{Y}_k\|_{2p}^2 + \widetilde{D}_k^2 \|W_k^d - \widehat{W}_k^d\|_{2p}^2\right) \\
&\qquad \times \left\| \mathrm{e}^{|Y_k| \vee |\widehat{Y}_k| + \widetilde{b}_k |W_k^d| \vee |\widehat{W}_k^d|} \right\|_{2q}^2.
\end{aligned} \tag{6.62}$$

- The last one is useful for the induction, indeed

$$\left\| \mathbb{E}\left[V_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)\right] - \mathbb{E}\left[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)\right] \right\|_2^2 \leqslant \left\| V_{k+1} - \widehat{V}_{k+1} \right\|_2^2. \tag{6.63}$$

Finally, using the L^r-L^s mismatch theorem for the quadratic optimal quantizers $\widehat{X}_k$ and $\widehat{Y}_k$, if $1 < p < 3/2$, then

$$\begin{aligned}
&\limsup_{N_k^X} N_k^X \|X_k - \widehat{X}_k\|_{2p} < +\infty, \qquad \limsup_{N_k^Y} N_k^Y \|Y_k - \widehat{Y}_k\|_{2p} < +\infty, \\
&\limsup_{N_k^{W^f}} N_k^{W^f} \|W_k^f - \widehat{W}_k^f\|_{2p} < +\infty \quad \text{and} \quad \limsup_{N_k^{W^d}} N_k^{W^d} \|W_k^d - \widehat{W}_k^d\|_{2p} < +\infty
\end{aligned} \tag{6.64}$$

this yields

$$\begin{aligned}
&\left\|V_k - \widehat{V}_k\right\|_2^2 \\
&\leqslant \left\|X_k - \widehat{X}_k\right\|_{2p}^2 \left([\bar{\psi}_k]_{Lip}^2 \left\| \mathrm{e}^{|Y_k| \vee |\widehat{Y}_k|} \right\|_{2q}^2 + \widetilde{A}_k^2 K_k^2 \right) \\
&\qquad + \left\|Y_k - \widehat{Y}_k\right\|_{2p}^2 \left(\left(\varphi_d(t_k) \|\psi_{t_k}\|_\infty + [\bar{\psi}_k]_{Lip} \right)^2 \left\| \mathrm{e}^{|Y_k| \vee |\widehat{Y}_k|} \right\|_{2q}^2 + \widetilde{C}_k^2 K_k^2 \right) \\
&\qquad\qquad + \widetilde{B}_k^2 K_k^2 \left\|W_k^f - \widehat{W}_k^f\right\|_{2p}^2 + \widetilde{D}_k^2 K_k^2 \left\|W_k^d - \widehat{W}_k^d\right\|_{2p}^2 + \left\|V_{k+1} - \widehat{V}_{k+1}\right\|_2^2 \qquad (6.65) \\
&\leqslant \sum_{l=k}^{n} C_{X_l} \left\|X_l - \widehat{X}_l\right\|_{2p}^2 + C_{Y_l} \left\|Y_l - \widehat{Y}_l\right\|_{2p}^2 + C_{W_l^d} \left\|W_l^d - \widehat{W}_l^d\right\|_{2p}^2 + C_{W_l^f} \left\|W_l^f - \widehat{W}_l^f\right\|_{2p}^2 \\
&\quad \xrightarrow{\bar{N} \to +\infty} 0
\end{aligned}$$

where $K_k = \left\| \mathrm{e}^{|Y_k| \vee |\widehat{Y}_k| + \widetilde{b}_k |W_k^d| \vee |\widehat{W}_k^d|} \right\|_{2q}$ and $\forall k = 1, \ldots, n, \quad C_{X_k}, C_{Y_k}, C_{W_k^d}, C_{W_k^f} < +\infty.$ □

Remark. The same result can be obtained if we relax the assumption on the payoff ψ_k. If we only assume the payoff Lipschitz continuous, we have the same limit with the same rate of convergence, however the constants $C_{X_l}, C_{Y_l}, C_{W_l^d}, C_{W_l^f}$ are not the same.

To conclude this section, although considering product optimal quantizer in four dimensions for (X_k, W_k^f, Y_k, W_k^d) seems to be natural, the computational cost associated to the resulting QBDPP is too high, of order $O(n \times (\max N_k)^2)$. Moreover the computation of the transition probabilities needed for the evaluation of the terms $\mathbb{E}\left[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d)\right]$ are challenging. These transition probabilities cannot be computed using deterministic numerical integration methods and we have to use Monte Carlo estimators. Even though it is feasible, it is a drawback for the method since it increases drastically the computation time for calibrating the quantization tree. In the next section we provide a solution to these problems which consists in reducing the dimension of the problem at the price of adding a systematic error, which turns out to be quite small in practice.

6.3.3 Quantization tree approximation: Non Markov case

In this part, we want to reduce the dimension of the problem in order to scale down the numerical complexity of the pricer. For that we discard the processes W^d and W^f in the tree and only keep X and Y. Doing so, we loose the Markovian property of our original model but we drastically reduce the numerical complexity of the problem. Thence, (6.22) is approximated by

$$\begin{cases} \widehat{V}_n = h_n(\widehat{X}_n, \widehat{Y}_n), \\ \widehat{V}_k = \max\left(h_k(\widehat{X}_k, \widehat{Y}_k), \mathbb{E}\left[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{Y}_k)\right] \right), \qquad 0 \leqslant k \leqslant n-1 \end{cases} \qquad (6.66)$$

where for every $k = 0, \dots, n$, $\widehat{X}_k$ and $\widehat{Y}_k$ are quadratic optimal quantizers of X_k and Y_k of size N_k^X and N_k^Y, respectively and we denote $N_k = N_k^X \times N_k^Y$ the size of the grid of the product quantizer.

Theorem 6.3.7. *Let the Markov transition $Pf(x,u,y,v)$ be defined by (6.20) be locally Lipschitz in the sense of Lemma 6.2.2. Assume that all the payoff functions $(\psi_{t_k})_{k=0:n}$ are Lipschitz continuous with compactly supported (right) derivative. Then the L^2-error, induced by the quantization approximation $(\widehat{X}_k, \widehat{Y}_k)$ is upper-bounded by*

$$\begin{aligned}\left\|V_k - \widehat{V}_k\right\|_2 \leqslant \Bigg(\sum_{l=k}^{n-1} C_{W_{l+1}^f} \left\|W_{l+1}^f - \mathbb{E}[W_{l+1}^f \mid (X_l, Y_l)]\right\|_{2p}^2 + C_{W_{l+1}^d} \left\|W_{l+1}^d - \mathbb{E}[W_{l+1}^d \mid (X_l, Y_l)]\right\|_{2p}^2 \\ + C_{X_l} \left\|X_l - \widehat{X}_l\right\|_{2p}^2 + C_{Y_l} \left\|Y_l - \widehat{Y}_l\right\|_{2p}^2 \Bigg)^{1/2}\end{aligned} \tag{6.67}$$

where $1 < p < 3/2$ and $q \geqslant 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$, moreover

$$\begin{aligned} C_{X_l} &= [\bar{\psi}_l]_{Lip}^2 \left\| \mathrm{e}^{|Y_l| \vee |\widehat{Y}_l|} \right\|_{2q}^2 + \bar{A}_l^2 \left\| \mathrm{e}^{\bar{b}_l |Y_l| \vee |\widehat{Y}_l|} \right\|_{2q}^2, & C_{W_{l+1}^f} &= B_{l+1}^2 \left\|\widetilde{\kappa}_{k+1}\right\|_{2q}^2, \\ C_{Y_l} &= \left(\varphi_d(t_l) \|\psi_{t_l}\|_\infty + [\bar{\psi}_l]_{Lip}\right)^2 \left\| \mathrm{e}^{|Y_l| \vee |\widehat{Y}_l|} \right\|_{2q}^2 + \bar{C}_l^2 \left\| \mathrm{e}^{\bar{b}_l |Y_l| \vee |\widehat{Y}_l|} \right\|_{2q}^2, & C_{W_{l+1}^d} &= D_{l+1}^2 \left\|\widetilde{\kappa}_{k+1}\right\|_{2q}^2. \end{aligned} \tag{6.68}$$

Taking the limit in $\bar{N} = \min N_k$, the size of the quadratic optimal quantizers, we have

$$\lim_{\bar{N} \to +\infty} \left\|V_k - \widehat{V}_k\right\|_2^2 = \sum_{l=k}^{n-1} C_{W_{l+1}^f} \left\|W_{l+1}^f - \mathbb{E}[W_{l+1}^f \mid (X_l, Y_l)]\right\|_{2p}^2 + C_{W_{l+1}^d} \left\|W_{l+1}^d - \mathbb{E}[W_{l+1}^d \mid (X_l, Y_l)]\right\|_{2p}^2. \tag{6.69}$$

Proof. We apply the same methodology as in the proof for the Markov case. The error between the Snell envelope and its approximation is given by

$$|V_k - \widehat{V}_k| \leqslant \max\left(\left|h_k(X_k, Y_k) - h_k(\widehat{X}_k, \widehat{Y}_k)\right|, \left| \mathbb{E}\left[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\right] - \mathbb{E}\left[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{Y}_k)\right] \right| \right) \tag{6.70}$$

thus, using Proposition 6.2.1 and Hölder's inequality with $p, q \geqslant 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$, the L^2-error is given by

$$\begin{aligned} \left\|V_k - \widehat{V}_k\right\|_2^2 &\leqslant \left\|h_k(X_k, Y_k) - h_k(\widehat{X}_k, \widehat{Y}_k)\right\|_2^2 \\ &\qquad + \left\| \mathbb{E}\left[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\right] - \mathbb{E}\left[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{Y}_k)\right] \right\|_2^2 \\ &\leqslant \left([\bar{\psi}_k]_{Lip}^2 \left\|X_k - \widehat{X}_k\right\|_{2p}^2 + \left(\varphi_d(t_k) \|\psi_{t_k}\|_\infty + [\bar{\psi}_k]_{Lip}\right)^2 \left\|Y_k - \widehat{Y}_k\right\|_{2p}^2 \right) \left\| \mathrm{e}^{|Y_k| \vee |\widehat{Y}_k|} \right\|_{2q}^2 \\ &\qquad + \left\| \mathbb{E}\left[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\right] - \mathbb{E}\left[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{Y}_k)\right] \right\|_2^2. \end{aligned} \tag{6.71}$$

The last term in Equation (6.71) can be decomposed as follows

$$\begin{aligned}\mathbb{E}\left[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\right] - \mathbb{E}\left[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{Y}_k)\right] \\ = \mathbb{E}\left[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\right] - \mathbb{E}\left[V_{k+1} \mid (X_k, Y_k)\right] \\ + \mathbb{E}\left[V_{k+1} \mid (X_k, Y_k)\right] - \mathbb{E}\left[V_{k+1} \mid (\widehat{X}_k, \widehat{Y}_k)\right] \\ + \mathbb{E}\left[V_{k+1} \mid (\widehat{X}_k, \widehat{Y}_k)\right] - \mathbb{E}\left[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{Y}_k)\right].\end{aligned} \tag{6.72}$$

And again, each term can be upper-bounded.

- The first can be upper-bounded using what we did above on the value function v_k and Hölder's inequality with $p, q \geqslant 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$

$$\begin{aligned}&\left\| \mathbb{E}\big[V_{k+1} \mid (X_k, W_k^f, Y_k, W_k^d)\big] - \mathbb{E}\left[V_{k+1} \mid (X_k, Y_k)\right]\right\|_2^2 \\ &\quad \leqslant \left\|V_{k+1} - \mathbb{E}\left[V_{k+1} \mid (X_k, Y_k)\right]\right\|_2^2 \\ &\quad \leqslant \Big\|v_{k+1}(X_{k+1}, W_{k+1}^f, Y_{k+1}, W_{k+1}^d) \\ &\qquad\qquad - v_{k+1}\big(X_{k+1}, \mathbb{E}\left[W_{k+1}^f \mid (X_k, Y_k)\right], Y_{k+1}, \mathbb{E}\left[W_{k+1}^d \mid (X_k, Y_k)\right]\big)\Big\|_2^2 \\ &\quad \leqslant \left\| \Big(B_{k+1} \big| W_{k+1}^f - \mathbb{E}\left[W_{k+1}^f \mid (X_k, Y_k)\right]\big| + D_{k+1} \big| W_{k+1}^d - \mathbb{E}\left[W_{k+1}^d \mid (X_k, Y_k)\right]\big| \Big) \widetilde{\kappa}_{k+1} \right\|_2^2 \\ &\quad \leqslant \left\|\widetilde{\kappa}_{k+1}\right\|_{2q}^2 \Big(B_{k+1}^2 \left\|W_{k+1}^f - \mathbb{E}[W_{k+1}^f \mid (X_k, Y_k)]\right\|_{2p}^2 + D_{k+1}^2 \left\|W_{k+1}^d - \mathbb{E}\left[W_{k+1}^d \mid (X_k, Y_k)\right]\right\|_{2p}^2 \Big)\end{aligned} \tag{6.73}$$

with coefficients b_{k+1}, B_{k+1} and D_{k+1} defined in (6.39) and (6.40) and

$$\widetilde{\kappa}_{k+1} = \mathrm{e}^{|Y_{k+1}| + b_{k+1}|W_{k+1}^d| \vee |\mathbb{E}[W_{k+1}^d | (X_k, Y_k)]|}. \tag{6.74}$$

- For the second, we define

$$\widetilde{v}_k(X_k, Y_k) = \mathbb{E}\left[v_{k+1}(X_{k+1}, W_{k+1}^f, Y_{k+1}, W_{k+1}^d) \mid (X_k, Y_k)\right]. \tag{6.75}$$

Indeed, $\mathbb{E}\left[V_{k+1} \mid (X_k, Y_k)\right]$ is only a function of X_k and Y_k, as shown below

$$\begin{aligned}\mathbb{E}\left[V_{k+1} \mid (X_k, Y_k)\right] &= \mathbb{E}\left[v_{k+1}(X_{k+1}, W_{k+1}^f, Y_{k+1}, W_{k+1}^d) \mid (X_k, Y_k)\right] \\ &= \mathbb{E}\Big[\mathbb{E}\left[v_{k+1}(X_{k+1}, W_{k+1}^f, Y_{k+1}, W_{k+1}^d) \mid (X_k, W_k^f, Y_k, W_k^d)\right] \mid (X_k, Y_k)\Big] \\ &= \mathbb{E}\left[P v_{k+1}(X_k, W_k^f, Y_k, W_k^d) \mid (X_k, Y_k)\right].\end{aligned} \tag{6.76}$$

Moreover, we can rewrite $W_k^f = \lambda_k X_k \overset{\perp\!\!\!\perp}{+} \xi_k$ and $W_k^d = \widetilde{\lambda}_k Y_k \overset{\perp\!\!\!\perp}{+} \chi_k$ where

$$\lambda_k = \frac{\mathbb{C}\mathrm{ov}(X_k, W_k^f)}{\mathbb{V}\mathrm{ar}(X_k)}, \qquad \widetilde{\lambda}_k = \frac{\mathbb{C}\mathrm{ov}(Y_k, W_k^d)}{\mathbb{V}\mathrm{ar}(Y_k)}$$

and $\xi_k \sim \mathcal{N}(0, \sigma^2_{\xi_k})$ and $\chi_k \sim \mathcal{N}(0, \sigma^2_{\chi_k})$ with $\sigma^2_{\xi_k} = \mathbb{V}\mathrm{ar}(W^f_k - \lambda_k X_k)$ and $\sigma^2_{\chi_k} = \mathbb{V}\mathrm{ar}(W^d_k - \widetilde{\lambda}_k Y_k)$, then

$$\begin{aligned}\mathbb{E}\left[Pv_{k+1}(X_k, W^f_k, Y_k, W^d_k) \mid (X_k, Y_k) = (x, y)\right] \\ = \mathbb{E}\left[Pv_{k+1}(x, \lambda_k x + \xi_k, y, \widetilde{\lambda}_k y + \chi_k)\right]\big|_{(x,y)=(X_k,Y_k)}\end{aligned} \tag{6.77}$$

yielding

$$\widetilde{v}_k(x, y) = \mathbb{E}\left[Pv_{k+1}(x, \lambda_k x + \xi_k, y, \widetilde{\lambda}_k y + \chi_k)\right]. \tag{6.78}$$

Now, using Lemma 6.2.2 on $\widetilde{v}_k$, we have

$$\begin{aligned}&\left|\widetilde{v}_k(x, y) - \widetilde{v}_k(x', y')\right| \\ &= \left|\mathbb{E}\left[Pv_{k+1}(x, \lambda_k x + \xi_k, y, \widetilde{\lambda}_k y + \chi_k) - Pv_{k+1}(x', \lambda_k x' + \xi_k, y', \widetilde{\lambda}_k y' + \chi_k)\right]\right| \\ &\leqslant \mathbb{E}\left[\left|\left((\widetilde{A}_k + \widetilde{B}_k|\lambda_k|)|x - x'| + (1 + \widetilde{C}_k|\widetilde{\lambda}_k|)|y - y'|\right) \mathrm{e}^{(1+\widetilde{b}_k|\widetilde{\lambda}_k|)|y|\vee|y'|+\widetilde{b}_k|\chi_k|}\right|\right] \\ &\leqslant \left(\bar{A}_k|x - x'| + \bar{C}_k|y - y'|\right) \mathrm{e}^{\bar{b}_k|y|\vee|y'|}\end{aligned} \tag{6.79}$$

where

$$\bar{A}_k = (\widetilde{A}_k + \widetilde{B}_k|\lambda_k|)\,\mathbb{E}\left[\,\mathrm{e}^{\widetilde{b}_k|\chi_k|}\,\right], \qquad \bar{C}_k = 1 + \widetilde{C}_k|\widetilde{\lambda}_k|, \tag{6.80}$$

$$\bar{b}_k = 1 + \widetilde{b}_k|\widetilde{\lambda}_k| \tag{6.81}$$

with $\widetilde{A}_k$, $\widetilde{B}_k$, $\widetilde{C}_k$ and $\widetilde{b}_k$ defined in (6.32) and (6.33). Hence, using Hölder's inequality with $p, q \geqslant 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$

$$\begin{aligned}&\left\|\mathbb{E}\left[V_{k+1} \mid (X_k, Y_k)\right] - \mathbb{E}\left[V_{k+1} \mid (\widehat{X}_k, \widehat{Y}_k)\right]\right\|^2_2 \\ &\qquad = \left\|\widetilde{v}_k(X_k, Y_k) - \widetilde{v}_k(\widehat{X}_k, \widehat{Y}_k)\right\|^2_2 \\ &\qquad \leqslant \left\|\left(\bar{A}_k|X_k - \widehat{X}_k| + \bar{C}_k|Y_k - \widehat{Y}_k|\right) \mathrm{e}^{\bar{b}_k|Y_k|\vee|\widehat{Y}_k|}\right\|^2_2 \\ &\qquad \leqslant \left\|\,\mathrm{e}^{\bar{b}_k|Y_k|\vee|\widehat{Y}_k|}\right\|^2_{2q}\left(\bar{A}^2_k\|X_k - \widehat{X}_k\|^2_{2p} + \bar{C}^2_k\|Y_k - \widehat{Y}_k\|^2_{2p}\right).\end{aligned} \tag{6.82}$$

- The last one is useful for the induction, indeed

$$\left\|\mathbb{E}\left[V_{k+1} \mid (\widehat{X}_k, \widehat{Y}_k)\right] - \mathbb{E}\left[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{Y}_k)\right]\right\|^2_2 \leqslant \left\|V_{k+1} - \widehat{V}_{k+1}\right\|^2_2. \tag{6.83}$$

Finally, using the L^r-L^s mismatch theorem on the quadratic optimal quantizers $\widehat{X}_k$ and $\widehat{Y}_k$, if $1 < p < 3/2$, then

$$\limsup_{N^X_k} N^X_k\|X_k - \widehat{X}_k\|_{2p} < +\infty \qquad \text{and} \qquad \limsup_{N^Y_k} N^Y_k\|Y_k - \widehat{Y}_k\|_{2p} < +\infty \tag{6.84}$$

and

$$
\begin{aligned}
&\big\|V_k-\widehat{V}_k\big\|_2^2\\
&\quad\leqslant \big\|X_k-\widehat{X}_k\big\|_{2p}^2\Big([\bar{\psi}_k]_{Lip}^2\big\|\,\mathrm{e}^{|Y_k|\vee|\widehat{Y}_k|}\,\big\|_{2q}^2+\bar{A}_k^2\big\|\,\mathrm{e}^{\bar{b}_k|Y_k|\vee|\widehat{Y}_k|}\,\big\|_{2q}^2\Big)\\
&\qquad+\big\|Y_k-\widehat{Y}_k\big\|_{2p}^2\Big(\big(\varphi_d(t_k)\|\psi_{t_k}\|_\infty+[\bar{\psi}_k]_{Lip}\big)^2\big\|\,\mathrm{e}^{|Y_k|\vee|\widehat{Y}_k|}\,\big\|_{2q}^2+\bar{C}_k^2\big\|\,\mathrm{e}^{\bar{b}_k|Y_k|\vee|\widehat{Y}_k|}\,\big\|_{2q}^2\Big)\\
&\qquad\quad+B_{k+1}^2\big\|\widetilde{\kappa}_{k+1}\big\|_{2q}^2\big\|W_{k+1}^f-\mathbb{E}[W_{k+1}^f\mid(X_k,Y_k)]\big\|_{2p}^2\\
&\qquad\qquad+D_{k+1}^2\big\|\widetilde{\kappa}_{k+1}\big\|_{2q}^2\big\|W_{k+1}^d-\mathbb{E}\left[W_{k+1}^d\mid(X_k,Y_k)\right]\big\|_{2p}^2+\big\|V_{k+1}-\widehat{V}_{k+1}\big\|_2^2\\
&\quad\leqslant\sum_{l=k}^{n-1}C_{W_{l+1}^f}\big\|W_{l+1}^f-\mathbb{E}[W_{l+1}^f\mid(X_l,Y_l)]\big\|_{2p}^2+C_{W_{l+1}^d}\big\|W_{l+1}^d-\mathbb{E}[W_{l+1}^d\mid(X_l,Y_l)]\big\|_{2p}^2\\
&\qquad\qquad+C_{X_l}\big\|X_l-\widehat{X}_l\big\|_{2p}^2+C_{Y_l}\big\|Y_l-\widehat{Y}_l\big\|_{2p}^2\\
&\quad\xrightarrow{\bar{N}\to+\infty}\sum_{l=k}^{n-1}C_{W_{l+1}^f}\big\|W_{l+1}^f-\mathbb{E}[W_{l+1}^f\mid(X_l,Y_l)]\big\|_{2p}^2+C_{W_{l+1}^d}\big\|W_{l+1}^d-\mathbb{E}[W_{l+1}^d\mid(X_l,Y_l)]\big\|_{2p}^2.
\end{aligned}
\tag{6.85}
$$

□

Practitioner's corner. Market implied values of σ_f, σ_d and σ_S used for the numerical computations are usually of order

$$
\sigma_f\approx 0.005,\qquad \sigma_d\approx 0.005,\qquad \sigma_S\approx 0.5 \tag{6.86}
$$

and in the most extreme cases, we compute Bermudan options on foreign exchange with maturity 20 years ($T=20$). Thus, we can estimate the order of the induced systematic error. First, we recall the expression of the related coefficients which depends of

$$
\begin{aligned}
B_k&=\sigma_f\frac{T}{n}\sum_{l=k+1}^{n}\max_{l\leqslant i\leqslant n}\Big([\bar{\psi}_i]_{Lip}\prod_{j=k+1}^{i}\mathbb{E}[\kappa_j]\Big),\\
D_k&=\sigma_d\frac{T}{n}\sum_{l=k+1}^{n}\max_{l\leqslant i\leqslant n}\Big((\varphi_d(t_i)\|\psi_i\|_\infty+[\bar{\psi}_i]_{Lip})\prod_{j=k+1}^{i}\mathbb{E}[\kappa_j]\Big)
\end{aligned}
\tag{6.87}
$$

with

$$
\kappa_j=\mathrm{e}^{|G_j^3|+b_j|G_j^4|},\qquad \widetilde{\kappa}_{l+1}=\mathrm{e}^{|Y_{l+1}|+b_{l+1}|W_{l+1}^d|\vee|\,\mathbb{E}[W_{l+1}^d|(X_l,Y_l)]|} \tag{6.88}
$$

and

$$
b_k=\sigma_d T\Big(1-\frac{k-1}{n}\Big). \tag{6.89}
$$

Now, considering the case where the payoffs are the same at each exercise date, the Lipschitz constants can be upper-bounded by $[\bar{\psi}]_{Lip}$:

$$[\bar{\psi}_k]_{Lip} = [\psi_{t_k}]_{Lip} S_0 \varphi_f(t_k)\, \mathrm{e}^{-\sigma_S^2 t_k/2}\, \|\psi'_{t_k}\|_\infty\, \mathrm{e}^c \leqslant S_0 [\psi_{t_k}]_{Lip} \|\psi'_{t_k}\|_\infty\, \mathrm{e}^c =: [\bar{\psi}]_{Lip} \tag{6.90}$$

and let κ defined by

$$\kappa = \max_k \mathbb{E}[\kappa_k] = \mathbb{E}\left[\, \mathrm{e}^{|G_0^3| + b_0 |G_0^4|}\, \right] \leqslant \frac{1}{2}\, \mathbb{E}\left[\, \mathrm{e}^{2|G_0^3|} + \mathrm{e}^{2b_0|G_0^4|}\, \right] \tag{6.91}$$

moreover, if $Z \sim \mathcal{N}(0, \sigma^2)$ then $\mathbb{E}\left[\, \mathrm{e}^{\lambda |Z|}\, \right] = \mathrm{e}^{\lambda^2 \sigma^2/2}$, thence we can upper-bound κ

$$\kappa \leqslant \frac{1}{2}\, \mathbb{E}\left[\, \mathrm{e}^{\sigma_3^2/2} + \mathrm{e}^{b_0^2/2}\, \right] = \frac{1}{2}\, \mathbb{E}\left[\, \mathrm{e}^{\sigma_d^2/96} + \mathrm{e}^{\sigma_d^2 T^2/2}\, \right] \approx 1. \tag{6.92}$$

κ being bounded, we notice that the main constants B_k^2 and D_k^2 in the remaining error are of order σ_d^2 or σ_f^2, indeed

$$\begin{aligned} B_k &\leqslant \sigma_f \frac{T}{n} [\bar{\psi}]_{Lip} (n-k) \kappa^{n-k} \approx \sigma_f \frac{T}{n} [\bar{\psi}]_{Lip} (n-k), \\ D_k &\leqslant \sigma_d \frac{T}{n} \big(\max_l \varphi_d(t_l) \|\psi\|_\infty + [\bar{\psi}]_{Lip} \big) (n-k) \kappa^{n-k} \approx \sigma_d \frac{T}{n} (\|\psi\|_\infty + [\bar{\psi}]_{Lip}) (n-k). \end{aligned} \tag{6.93}$$

Furthermore

$$\begin{aligned} \mathbb{E}\left[\widetilde{\kappa}_{k+1}^{2q}\right] &= \mathbb{E}\left[\, \mathrm{e}^{2q|Y_{k+1}| + 2qb_{k+1}|W_{k+1}^d| \vee |\,\mathbb{E}[W_{k+1}^d|(X_k, Y_k)]|}\, \right] \\ &\leqslant \frac{1}{2} \left(\mathbb{E}\left[\, \mathrm{e}^{4q|Y_{k+1}|}\, \right] + \mathbb{E}\left[\, \mathrm{e}^{4qb_{k+1}|W_{k+1}^d| \vee |\,\mathbb{E}[W_{k+1}^d|(X_k, Y_k)]|}\, \right] \right) \\ &\leqslant \frac{1}{2} \left(\mathbb{E}\left[\, \mathrm{e}^{4q|Y_{k+1}|}\, \right] + \mathbb{E}\left[\, \mathrm{e}^{4q\sigma_d(T-t_k)|W_{k+1}^d| \vee |\,\mathbb{E}[W_{k+1}^d|(X_k, Y_k)]|}\, \right] \right) \\ &\leqslant \frac{1}{2} \left(\mathrm{e}^{8q^2\sigma_d^2 T^3/3} + 2\, \mathrm{e}^{8q^2\sigma_d^2 (T-t_k)^2 t_{k+1}} \right) \end{aligned} \tag{6.94}$$

and from elementary inequality $(a+b)^{1/q} \leqslant a^{1/q} + b^{1/q}$, $a, b \geqslant 0$, $q \geqslant 1$

$$\begin{aligned} \left\| \widetilde{\kappa}_{k+1} \right\|_{2q}^2 = \mathbb{E}\left[\widetilde{\kappa}_{k+1}^{2q}\right]^{\frac{1}{q}} &\leqslant \left(\frac{1}{2}\, \mathrm{e}^{8q^2\sigma_d^2 T^3/3} + \mathrm{e}^{8q^2\sigma_d^2 (T-t_k)^2 t_{k+1}} \right)^{\frac{1}{q}} \\ &\leqslant \left(\frac{1}{2}\, \mathrm{e}^{8q^2\sigma_d^2 T^3/3} \right)^{\frac{1}{q}} + \left(\mathrm{e}^{8q^2\sigma_d^2 (T-t_k)^2 t_{k+1}} \right)^{\frac{1}{q}} \\ &\leqslant \frac{1}{2^{1/q}}\, \mathrm{e}^{8q\sigma_d^2 T^3/3} + \mathrm{e}^{8q\sigma_d^2 (T-t_k)^2 t_{k+1}}. \end{aligned} \tag{6.95}$$

The two terms on the right-hand side of the inequality do not explode. Indeed, the function $g : t \mapsto (T-t)^2 t$, defined for $t \in [0, T]$ with $T = 20$, attains its maximum on $t = 20/3$ and

$g(20/3) \approx 1185$, hence for the considered values

$$\forall k = 1, \dots, n, \qquad \left\|\widetilde{\kappa}_{k+1}\right\|_{2q}^2 \leqslant C_{\widetilde{\kappa}} \approx 6. \tag{6.96}$$

Finally, rewriting the obtained systematic error induced by the approximation with this new informations in (6.69) we have

$$\begin{aligned}
\left\|V_k - \widehat{V}_k\right\|_2^2 \xrightarrow{N\to+\infty} & \sum_{l=k}^{n-1} B_{l+1}^2 \left\|\widetilde{\kappa}_{l+1}\right\|_{2q}^2 \left\|W_{l+1}^d - \mathbb{E}[W_{l+1}^d \mid (X_l, Y_l)]\right\|_{2p}^2 \\
& \qquad + D_{l+1}^2 \left\|\widetilde{\kappa}_{l+1}\right\|_{2q}^2 \left\|W_{l+1}^f - \mathbb{E}[W_{l+1}^f \mid (X_l, Y_l)]\right\|_{2p}^2 \\
\leqslant & \sigma_f^2 \Big(\frac{T}{n}\Big)^2 [\bar{\psi}]_{Lip}^2 \sum_{l=k}^{n-1} (n-l)^2 \kappa^{2(n-l)} C_{\widetilde{\kappa}} \left\|W_{l+1}^d - \mathbb{E}[W_{l+1}^d \mid (X_l, Y_l)]\right\|_{2p}^2 \\
& + \sigma_d^2 \Big(\frac{T}{n}\Big)^2 \big(\max_l \varphi_d(t_l) \|\psi\|_\infty + [\bar{\psi}]_{Lip}\big)^2 \\
& \qquad \times \sum_{l=k}^{n-1} (n-l)^2 \kappa^{2(n-l)} C_{\widetilde{\kappa}} \left\|W_{l+1}^f - \mathbb{E}[W_{l+1}^f \mid (X_l, Y_l)]\right\|_{2p}^2 \\
\leqslant & 2\sigma_f^2 \Big(\frac{T}{n}\Big)^2 [\bar{\psi}]_{Lip}^2 \sum_{l=k}^{n-1} (n-l)^2 \kappa^{2(n-l)} C_{\widetilde{\kappa}} \left\|W_{l+1}^d\right\|_{2p}^2 \\
& + 2\sigma_d^2 \Big(\frac{T}{n}\Big)^2 \big(\max_l \varphi_d(t_l) \|\psi\|_\infty + [\bar{\psi}]_{Lip}\big)^2 \sum_{l=k}^{n-1} (n-l)^2 \kappa^{2(n-l)} C_{\widetilde{\kappa}} \left\|W_{l+1}^f\right\|_{2p}^2 \\
\leqslant & \Big(\sigma_f^2 [\bar{\psi}]_{Lip}^2 + \sigma_d^2 \big(\max_l \varphi_d(t_l) \|\psi\|_\infty + [\bar{\psi}]_{Lip}\big)^2\Big) 4 \frac{C_{\widetilde{\kappa}}}{\pi^{1/3}} \Big(\frac{T}{n}\Big)^2 \sum_{l=k}^{n-1} t_{l+1} (n-l)^2 \kappa^{2(n-l)}.
\end{aligned} \tag{6.97}$$

Hence, the systematic error is upper-bounded by the squared volatilities σ_d^2 and σ_f^2. These parameters being of order 5×10^{-3} at most, the systematic error is negligible as long as these volatilities stay reasonably small.

Remark. As in the Markov case, we can extend this result to the case where the payoffs $(\psi_k)_k$ are Lipschitz continuous, however the residual error can not be as easily estimated and controlled.

6.4 Numerical experiments

In this section, we illustrate the theoretical results found in Section 6.3 regarding the pricing of Bermudan options in the 3-factor model described in Section 6.1. First, we detail both algorithms and how to compute the quantities that appear in them (conditional expectation, conditional probabilities, ...). Then, we test our two numerical solutions for the pricing of European options, whose price is known in closed form. European options are Bermudan options with only one date of exercise, hence when using the non-Markovian approximate we do

not introduce the systematic error shown in Theorem 6.3.7 but pricing these kind of options is a good benchmark in order to test our methodologies. Finally, we evaluate Bermudan options and compare our two solutions, the Markovian and the non-Markovian approximation.

We have to keep in mind that the computation time is crucial because these pricers are only a small block in the complex computation of xVA's. Indeed, they will be called hundreds of thousands of time each time these risks measures are needed.

All the numerical tests have been carried out in C++ on a laptop with a 2,4 GHz 8-Core Intel Core i9 CPU. The computations of the transition probabilities and the computations of the conditional expectations are parallelized on the CPU.

Remark. The computation times given below measure the time needed for loading the pre-computed optimal grids from files, rescaling the optimal quantizers in order to get the right variance, computing the conditional probabilities (the part that demands the most in term of computing power) and finally computing the expectations for the pricing. One has to keep in mind that the complexity is linear in function of n, the number of exercise dates. Indeed, if we double the number of exercise dates, we double the number of conditional probability matrices and expectations to compute.

Characterisation of the Quantization Tree. In what follows, we describe the choice of parameters we made when building the quantization tree: the time discretisation and the size of each grid at each time.

- The time discretisation is an easy choice because it is decided by the characteristics of the financial product. Indeed, we take only one date (and today's date) in the tree if we want to evaluate European options and if we want to evaluate Bermudan options we take as many discretisation dates (plus today's date) in the tree as there are exercise dates in the description of the product.

- Then, we have to decide the size of each grid at each date in the tree. In our case, we consider grids of same size at each date hence $N_k = N$, $k = 1 \dots, n$ and then we take $N^X = 10N^Y$ for both trees. This choice seems to be reasonable because the risk factor X_k is prominent, due to the value of σ_S compare to σ_d. Now, in the Markovian case, we take $N^X = 4N^{W_f}$ and $N^Y = 4N^{W_d}$, indeed the two Brownian Motions are important only when we compute the conditional expectation but not when we want to evaluate the payoffs, hence we want to give as much as possible of the budget N to N^X and N^Y.

The algorithm: Markovian Case. Let $(x^k_{i_1})_{i_1=1:N^X}$, $(u^k_{i_2})_{i_2=1:N^{W^f}}$, $(y^k_{i_3})_{i_3=1:N^Y}$ and $(v^k_{i_4})_{i_4=1:N^{W^d}}$ be the associated centroids of $\widehat{X}_k$, $\widehat{W}^f_k$, $\widehat{Y}_k$ and $\widehat{W}^d_k$ respectively, at a given time t_k with $0 \leqslant k \leqslant n$. Using the discrete property of the optimal quantizers, the conditional expectation appearing in

(6.50) can be rewritten as

$$\begin{aligned}\mathbb{E}\big[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d) = (x_{i_1}^k, u_{i_2}^k, y_{i_3}^k, v_{i_4}^k)\big] \\ = \mathbb{E}\big[\widehat{v}_{k+1}(\widehat{X}_{k+1}, \widehat{W}_{k+1}^f, \widehat{Y}_{k+1}, \widehat{W}_{k+1}^d) \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d) = (x_{i_1}^k, u_{i_2}^k, y_{i_3}^k, v_{i_4}^k)\big] \\ = \sum_{j_1,j_2,j_3,j_4} \pi_{i,j}^{(\mathrm{M}),k}\, \widehat{v}_{k+1}(x_{j_1}^{k+1}, u_{j_2}^{k+1}, y_{j_3}^{k+1}, v_{j_4}^{k+1})\end{aligned} \tag{6.98}$$

where $\pi_{i,j}^{(\mathrm{M}),k}$, with $i = (i_1, i_2, i_3, i_4)$ and $j = (j_1, j_2, j_3, j_4)$, is the conditional probability defined by

$$\begin{aligned}\pi_{i,j}^{(\mathrm{M}),k} = \mathbb{P}\Big((\widehat{X}_{k+1}, \widehat{W}_{k+1}^f, \widehat{Y}_{k+1}, \widehat{W}_{k+1}^d) = (x_{j_1}^{k+1}, u_{j_2}^{k+1}, y_{j_3}^{k+1}, v_{j_4}^{k+1}) \\ \mid (\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d) = (x_{i_1}^k, u_{i_2}^k, y_{i_3}^k, v_{i_4}^k)\Big).\end{aligned}$$

Due to the dimension of the problem (4 in this case), we cannot compute these probabilities using deterministic methods, hence one has to simulate trajectories of the processes in order to evaluate them. We refer the reader to [BPP05; BP03; PPP04b] for details on the methodology.

A way to reduce the complexity of the problem is to approximate these probabilities by $\widetilde{\pi}_{i,j}^{(\mathrm{M}),k}$, where the conditional part $\{(\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d) = (x_{i_1}^k, u_{i_2}^k, y_{i_3}^k, v_{i_4}^k)\}$ is replaced by $\{(X_k, W_k^f, Y_k, W_k^d) = (x_{i_1}^k, u_{i_2}^k, y_{i_3}^k, v_{i_4}^k)\}$, yielding

$$\begin{aligned}\widetilde{\pi}_{i,j}^{(\mathrm{M}),k} = \mathbb{P}\Big((\widehat{X}_{k+1}, \widehat{W}_{k+1}^f, \widehat{Y}_{k+1}, \widehat{W}_{k+1}^d) = (x_{j_1}^{k+1}, u_{j_2}^{k+1}, y_{j_3}^{k+1}, v_{j_4}^{k+1}) \\ \mid (X_k, W_k^f, Y_k, W_k^d) = (x_{i_1}^k, y_{i_2}^k, u_{i_3}^k, v_{i_4}^k)\Big).\end{aligned} \tag{6.99}$$

The reason for replacing $\{(\widehat{X}_k, \widehat{W}_k^f, \widehat{Y}_k, \widehat{W}_k^d) = (x_{i_1}^k, u_{i_2}^k, y_{i_3}^k, v_{i_4}^k)\}$ by $\{(X_k, W_k^f, Y_k, W_k^d) = (x_{i_1}^k, u_{i_2}^k, y_{i_3}^k, v_{i_4}^k)\}$ is explained in the next paragraph dealing with the Non-Markovian case with lighter notations (see Equation (6.102) and (6.104)). Although, these probabilities are easier to calculate, one still has to devise a Monte Carlo simulation in order to evaluate them. This simplification will be useful later in the uncorrelated case.

These remarks allow us to rewritte the QBDPP in the Markovian case (6.50) as

$$\begin{cases}\widehat{v}_n(x_{i_1}^n, u_{i_2}^n, y_{i_3}^n, v_{i_4}^n) = h_n(x_{i_1}^n, y_{i_3}^n), \\ \widehat{v}_k(x_{i_1}^k, u_{i_2}^k, y_{i_3}^k, v_{i_4}^k) = \max\Big(h_k(x_{i_1}^k, y_{i_3}^k), \sum_{j_1,j_2,j_3,j_4} \widetilde{\pi}_{i,j}^{(\mathrm{M}),k}\, \widehat{v}_{k+1}(x_{j_1}^{k+1}, u_{j_2}^{k+1}, y_{j_3}^{k+1}, v_{j_4}^{k+1})\Big).\end{cases} \tag{6.100}$$

The algorithm: Non-Markovian case. Let $(x_{i_1}^k)_{i_1=1:N^X}$ and $(y_{i_3}^k)_{i_3=1:N^Y}$ be the associated centroids of $\widehat{X}_k$ and $\widehat{Y}_k$ respectively, at a given time t_k with $0 \leqslant k \leqslant n$. Again, as in the Markovian case, using the discrete property of the optimal quantizers, the conditional

expectation appearing in (6.66) can be rewritten as

$$\begin{aligned}\mathbb{E}\left[\widehat{V}_{k+1} \mid (\widehat{X}_k, \widehat{Y}_k) = \left(x_{i_1}^k, y_{i_2}^k\right)\right] &= \mathbb{E}\left[\widehat{v}_{k+1}(\widehat{X}_{k+1}, \widehat{Y}_{k+1}) \mid (\widehat{X}_k, \widehat{Y}_k) = \left(x_{i_1}^k, y_{i_2}^k\right)\right] \\ &= \sum_{j_1, j_2} \pi_{i,j}^{(\mathrm{NM}),k}\, \widehat{v}_{k+1}\left(x_{j_1}^{k+1}, y_{j_2}^{k+1}\right)\end{aligned} \tag{6.101}$$

where $\pi_{i,j}^{(\mathrm{NM}),k}$, with $i = (i_1, i_2)$ and $j = (j_1, j_2)$, is the conditional probability defined by

$$\pi_{i,j}^{(\mathrm{NM}),k} = \mathbb{P}\left(\left(\widehat{X}_{k+1}, \widehat{Y}_{k+1}\right) = \left(x_{j_1}^{k+1}, y_{j_2}^{k+1}\right) \mid \left(\widehat{X}_k, \widehat{Y}_k\right) = \left(x_{i_1}^k, y_{i_2}^k\right)\right).$$

This probability can be computed by numerical integration, ie

$$\begin{aligned}\pi_{i,j}^{(\mathrm{NM}),k} &= \mathbb{P}\left(\left(\widehat{X}_{k+1}, \widehat{Y}_{k+1}\right) = \left(x_{j_1}^{k+1}, y_{j_2}^{k+1}\right) \mid \left(\widehat{X}_k, \widehat{Y}_k\right) = \left(x_{i_1}^k, y_{i_2}^k\right)\right) \\ &= \mathbb{P}\left(\left(\widehat{X}_{k+1}, \widehat{Y}_{k+1}\right) = \left(x_{j_1}^{k+1}, y_{j_2}^{k+1}\right) \mid X_k \in \left(x_{i_1-1/2}^k, x_{i_1+1/2}^k\right), Y_k \in \left(y_{i_2-1/2}^k, y_{i_2+1/2}^k\right)\right) \\ &= \int_{x_{i_1-1/2}^k}^{x_{i_1+1/2}^k} \int_{y_{i_2-1/2}^k}^{y_{i_2+1/2}^k} \mathbb{P}\left(\left(\widehat{X}_{k+1}, \widehat{Y}_{k+1}\right) = \left(x_{j_1}^{k+1}, y_{j_2}^{k+1}\right) \mid (X_k, Y_k) = (x, y)\right) f_\Sigma(x, y)\, dx\, dy\end{aligned} \tag{6.102}$$

where $f_\Sigma(x, y)$ is the joint density of a centered bivariate Gaussian vector with covariance matrix Σ given by

$$\Sigma = \begin{pmatrix} \mathbb{V}\mathrm{ar}(X_k) & \mathbb{C}\mathrm{ov}(X_k, Y_k) \\ \mathbb{C}\mathrm{ov}(X_k, Y_k) & \mathbb{V}\mathrm{ar}(Y_k) \end{pmatrix}. \tag{6.103}$$

However, computing the probability in Equation (6.102) can be too time consuming, hence once again, we approximate this probability by $\widetilde{\pi}_{i,j}^{(\mathrm{NM}),k}$, where the conditional part $\left\{\left(\widehat{X}_k, \widehat{Y}_k\right) = \left(x_{i_1}^k, y_{i_2}^k\right)\right\}$ is replaced by $\left\{(X_k, Y_k) = \left(x_{i_1}^k, y_{i_2}^k\right)\right\}$, yielding

$$\widetilde{\pi}_{i,j}^{(\mathrm{NM}),k} = \mathbb{P}\left(\left(\widehat{X}_{k+1}, \widehat{Y}_{k+1}\right) = \left(x_{j_1}^{k+1}, y_{j_2}^{k+1}\right) \mid (X_k, Y_k) = \left(x_{i_1}^k, y_{i_2}^k\right)\right). \tag{6.104}$$

From the definition of an optimal quantizer and Equation (6.17), this probability can be rewritten as the probability that a correlated bivariate normal distribution lies in a rectangular domain

$$\begin{aligned}\widetilde{\pi}_{i,j}^{(\mathrm{NM}),k} &= \mathbb{P}\left(\widehat{X}_{k+1} = x_{j_1}^{k+1}, \widehat{Y}_{k+1} = y_{j_2}^{k+1} \mid X_k = x_{i_1}^k, Y_k = y_{i_2}^k\right) \\ &= \mathbb{P}\left(X_{k+1} \in \left(x_{j_1-1/2}^{k+1}, x_{j_1+1/2}^{k+1}\right), Y_{k+1} \in \left(y_{j_2-1/2}^{k+1}, y_{j_2+1/2}^{k+1}\right) \mid X_k = x_{i_1}^k, Y_k = y_{i_2}^k\right) \\ &= \mathbb{P}\left(x_{i_1}^k + \sigma_f \delta W_k^f + G_{k+1}^1 \in \left(x_{j_1-1/2}^{k+1}, x_{j_1+1/2}^{k+1}\right), y_{i_2}^k - \sigma_d \delta W_k^d + G_{k+1}^3 \in \left(y_{j_2-1/2}^{k+1}, y_{j_2+1/2}^{k+1}\right)\right) \\ &= \mathbb{P}\left(Z^1 \in \left(x_{j_1-1/2}^{k+1} - x_{i_1}^k, x_{j_1+1/2}^{k+1} - x_{i_1}^k\right), Z^2 \in \left(y_{j_2-1/2}^{k+1} - y_{i_2}^k, y_{j_2+1/2}^{k+1} - y_{i_2}^k\right)\right)\end{aligned} \tag{6.105}$$

where

$$\begin{pmatrix} Z^1 \\ Z^2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{Z^1} & \rho_{Z^1,Z^2}\sigma_{Z^1}\sigma_{Z^2} \\ \rho_{Z^1,Z^2}\sigma_{Z^1}\sigma_{Z^2} & \sigma^2_{Z^2} \end{pmatrix} \right) \tag{6.106}$$

with $\sigma^2_{Z^1} = \mathbb{V}\mathrm{ar}(\sigma_f \delta W^f_k + G^1_{k+1})$, $\sigma^2_{Z^2} = \mathbb{V}\mathrm{ar}(-\sigma_d \delta W^d_k + G^3_{k+1})$ and $\rho_{Z^1,Z^2} = \mathbb{C}\mathrm{orr}(\sigma_f \delta W^f_k + G^1_{k+1}, -\sigma_d \delta W^d_k + G^3_{k+1})$.

The advantage of expressing (6.105) as the probability that a bivariate Gaussian vector lies in a rectangular domain is that it can be rewritten as a linear combination of bivariate cumulative distribution functions.

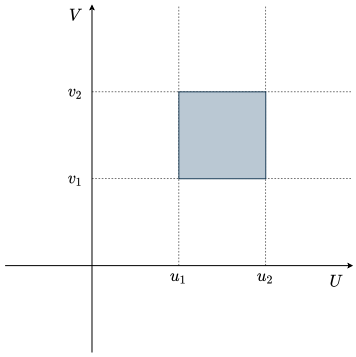


Fig. 6.2

Indeed, let (U, V) a two-dimensional correlated and standardized normal distribution with correlation ρ and cumulative distribution function (CDF) given by $F^\rho_{U,V}(u, v) = \mathbb{P}(U \leqslant u, V \leqslant v)$. Fast and efficient numerical implementation of such function exists (for example, a C++ implementation of the upper right tail of a correlated bivariate normal distribution can be found in John Burkardt's website, see [Bur12], which is based on the work of [Don73] and [Owe58]. In our case, we are interested in the computation of probabilities of the form

$$\mathbb{P}\left(U \in (u_1, u_2), V \in (v_1, v_2)\right). \tag{6.107}$$

This probability is represented graphically as the integral of the two-dimensional density over the rectangular domain in grey in Figure 6.2. Now, using $F^\rho_{U,V}(u, v)$, the probability (6.107) is given by

$$\mathbb{P}\left(U \in (u_1, u_2), V \in (v_1, v_2)\right) = F^\rho_{U,V}(u_2, v_2) - F^\rho_{U,V}(u_1, v_2) - F^\rho_{U,V}(u_2, v_1) + F^\rho_{U,V}(u_1, v_1). \tag{6.108}$$

This remark will allow us to reduce drastically the computation time induced by the evaluation of the conditional probabilities and so, of the conditional expectations.

Now, going back to our problem, the QBDPP in the non-Markovian case rewrites (6.66)

$$\begin{cases} \widehat{v}_n\left(x^n_{i_1}, y^n_{i_2}\right) = h_n\left(x^n_{i_1}, y^n_{i_2}\right), \qquad 1 \leqslant i_1 \leqslant N^X_n, \qquad 1 \leqslant i_2 \leqslant N^Y_n, \\ \widehat{v}_k\left(x^k_{i_1}, y^k_{i_2}\right) = \max\left(h_k\left(x^k_{i_1}, y^k_{i_2}\right), \sum_{j_1, j_2} \pi^{(\mathrm{NM}),k}_{i,j}\, \widehat{v}_{k+1}\left(x^{k+1}_{j_1}, y^{k+1}_{j_2}\right) \right). \end{cases} \tag{6.109}$$

In order to test numerically the two methods, we will evaluate PRDC European and Bermudan options with maturities $2Y$, $5Y$ and $10Y$. We describe below the market and products parameters we consider. The volatilities of the domestic and the foreign interest rates are not detailed below because we investigate the behaviour of the methods with respect to σ_d and σ_f.

$P_d(0,t)$	$\exp(-r_d t)$	r_d	0.015	ρ_{Sd}	0
$P_f(0,t)$	$\exp(-r_f t)$	r_f	0.01	ρ_{Sf}	0
S_0	88.17	σ_S	0.5	ρ_{df}	0

Table 6.1 *Market values.*

$\forall k \in 1,\dots,n,\quad C_d(t_k)$	15%	$\forall k \in 1,\dots,n,\quad C_f(t_k)$	18.9%
$\forall k \in 1,\dots,n,\quad \mathrm{Cap}(t_k)$	5.55%	$\forall k \in 1,\dots,n,\quad \mathrm{Floor}(t_k)$	0%
Exercise date (EU): t_n	T	Exercise dates (US): t_k	Tk/n

Table 6.2 *Product description.*

Remark. When the correlations ρ_{df} and ρ_{Sd} are equal to zero, the numerical computation of probabilities $\widetilde{\pi}_{i,j}^{(\mathrm{M}),k}$ and $\widetilde{\pi}_{i,j}^{(\mathrm{NM}),k}$ can be accelerated. Indeed, in the Markovian case, (6.99) can be rewritten as

$$\begin{aligned}\widetilde{\pi}_{i,j}^{(\mathrm{M}),k} = \mathbb{P}\left((\widehat{X}_{k+1}, \widehat{W}^f_{k+1}) = (x^{k+1}_{j_1}, u^{k+1}_{j_2}) \mid (X_k, W^f_k) = (x^k_{i_1}, u^k_{i_2})\right) \\ \times \mathbb{P}\left((\widehat{Y}_{k+1}, \widehat{W}^d_{k+1}) = (y^{k+1}_{j_3}, v^{k+1}_{j_4}) \mid (Y_k, W^d_k) = (y^k_{i_3}, v^k_{i_4})\right).\end{aligned} \tag{6.110}$$

In that case, we can use the CDF of a correlated bivariate normal distribution, as detailed above for the non-Markovian case in (6.108), for computing these probabilities in a very effective and faster way rather than performing a Monte Carlo simulation.

In the non-Markovian case, (6.105) can be rewritten as

$$\begin{aligned}\widetilde{\pi}_{i,j}^{(\mathrm{NM}),k} &= \mathbb{P}\left(Z^1 \in (x^{k+1}_{j_1-1/2} - x^k_{i_1}, x^{k+1}_{j_1+1/2} - x^k_{i_1})\right)\mathbb{P}\left(Z^2 \in (y^{k+1}_{j_2-1/2} - y^k_{i_2}, y^{k+1}_{j_2+1/2} - y^k_{i_2})\right) \\ &= \left(F_Z\left(\frac{x^{k+1}_{j_1+1/2} - x^k_{i_1}}{\sigma_{Z^1}}\right) - F_Z\left(\frac{x^{k+1}_{j_1-1/2} - x^k_{i_1}}{\sigma_{Z^1}}\right)\right)\left(F_Z\left(\frac{y^{k+1}_{j_2+1/2} - y^k_{i_2}}{\sigma_{Z^2}}\right) - F_Z\left(\frac{y^{k+1}_{j_2-1/2} - y^k_{i_2}}{\sigma_{Z^2}}\right)\right)\end{aligned} \tag{6.111}$$

where $F_Z(\cdot)$ is the CDF of a one-dimensional normal distribution, σ_{Z^1} is the standard deviation of Z^1 and σ_{Z^2} is the standard deviation of Z^2. This remark allows us to drastically reduce the computation time of the conditional probabilities in the case of zero correlations.

6.4.1 European Option

First of all, we compare the asymptotic behaviour of the Markovian and the non-Markovian approaches when pricing European PRDC Options with different volatilities and maturities. In this case, we consider only two dates in the tree: $t_0 = 0$ an $t_n = T$, the algorithm is a regular cubature formula and no systematic error is induced by the non-markovianity of the couple (X_k, Y_k). These numerical tests confirm that both approaches give the same value, however the non-Markovian approach converges much faster due to the dimension of the

product quantization, 2 for the first one and 4 for the last one. Indeed, the complexity of the 2 dimensional pricer is of order of $N = N^X \times N^Y$ while the complexity of the 4 dimensional pricer is of order $N = N^X \times N^Y \times N^{W^d} \times N^{W^f}$. N being the size of the product quantizer at each date (in two dimensions: $N = N^X \times N^Y$ and in four dimensions $N = N^X \times N^{W^f} \times N^Y \times N^{W^d}$).

In the case of the European options, we have a closed-form formula for the price of (6.12). The benchmark price is computed using the rewriting of (6.12) as a sum of Calls: at a time t_k, the payoff can be expressed as

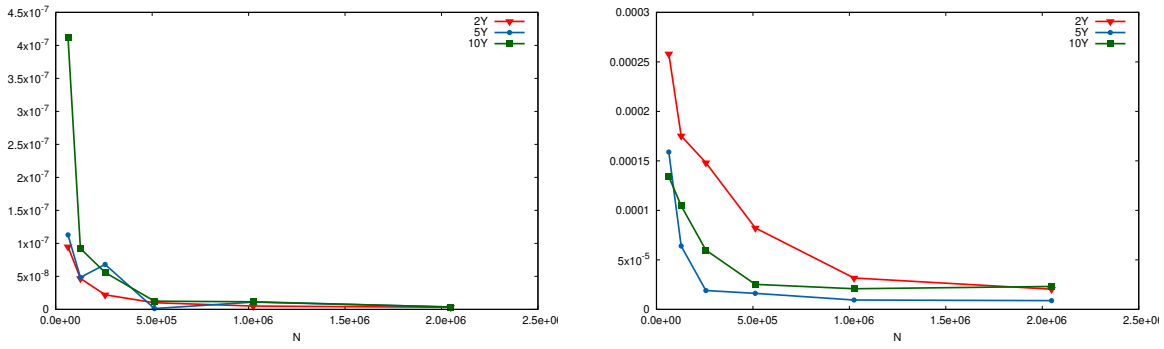
$$\begin{aligned}\psi_{t_k}(S_{t_k}) &= \min\left(\max\left(\frac{C_f(t_k)}{S_0}S_{t_k} - C_d(t_k), \text{Floor}(t_k)\right), \text{Cap}(t_k)\right) \\ &= \text{Floor}(t_k) - a_k(S_{t_k} - K_k^1)_+ + a_k(S_{t_k} - K_k^2)_+\end{aligned}$$

with $a_k = \dfrac{C_f(t_k)}{S_0}$, $K_k^1 = \dfrac{\text{Cap}(t_k) + C_d(t_k)}{C_f(t_k)} \times S_0$ and $K_k^2 = \dfrac{\text{Floor}(t_k) + C_d(t_k)}{C_f(t_k)} \times S_0$ and the closed-form formula for the price of a Call is detailed in Appendix 6.B. The prices given by the closed-form formula of the European options we consider (different values of volatilities and different maturities) are given in Table 6.3.

	Exact price	
T \ σ	50*bp*	500*bp*
2Y	2.171945242	2.159404007
5Y	1.630435483	1.539295559
10Y	1.127330259	0.8013151892

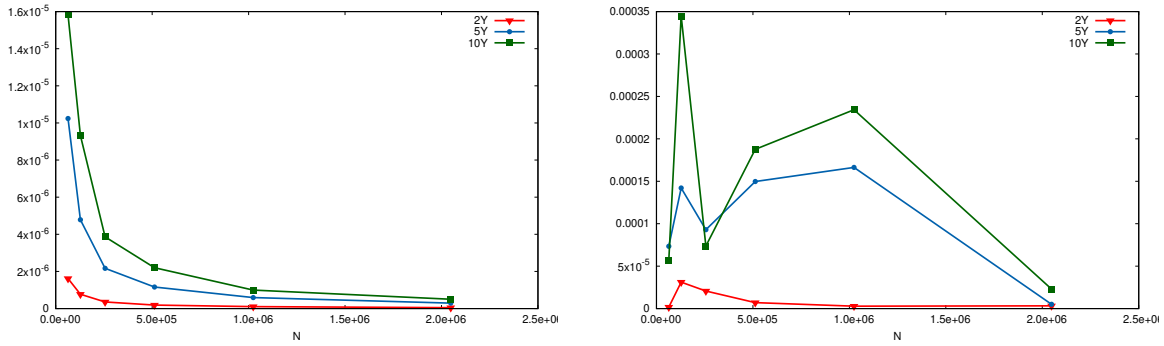
Table 6.3 *Prices given by closed-form formula of European options with zero correlations.* ($\sigma_d = \sigma_f = \sigma$)

The difference of speed of convergence between the two methods is illustrated in Figures 6.3 and 6.4 for the relative errors for both methods compared to the benchmark. N in the label of each graphic represents the size of the product quantizer ($N^X \times N^{W^f} \times N^Y \times N^{W^d}$ in the Markovian case and $N^X \times N^Y$ in the other case), hence the complexity of both trees are the same.



(a) *Non-Markovian – 2d* (b) *Markovian – 4d*

Fig. 6.3 $\sigma_d = \sigma_f = 50bp$ – *Relative errors for both methods for 2Y, 5Y and 10Y European options pricing (with zero correlations).*



(a) *Non-Markovian – 2d* (b) *Markovian – 4d*

Fig. 6.4 $\sigma_d = \sigma_f = 500bp$ – *Relative errors for both methods for 2Y, 5Y and 10Y European options pricing (with zero correlations).*

For both methods, a relative error of $1bp$ is quickly reached, even for high values of σ_d and σ_f. Indeed, the time needed in order to achieve a $1bp$ precision for building a quantization tree with 2 dates, computing the probabilities and pricing a European option is at most 6 ms for the non-Markovian method and at most 85ms for the Markovian one when the correlations are equal to zero. The computation times needed for a $1bp$ relative error are summarised in Table 6.4.

	Non-Markovian – 2d		Markovian – 4d	
T \ σ	$50bp$	$500bp$	$50bp$	$500bp$
2Y	1 ms (32000)	4 ms (32000)	24 ms (512000)	4 ms (64000)
5Y	4 ms (32000)	6 ms (32000)	4 ms (64000)	85 ms (2048000)
10Y	4 ms (32000)	3 ms (32000)	14 ms (256000)	83 ms (2048000)

Table 6.4 *Times in milliseconds needed for reaching a* $1bp$ *precision for European options pricing with zero correlations using both methods with, in parenthesis, the size* N *of the grid at each time step.* ($\sigma_d = \sigma_f = \sigma$)

Remark. Of course, the pricers can be used even when we consider non-zero correlations. We choose to show only the asymptotic behaviour of the non-Markovian method because it converges much faster and the computations of the probabilities can be made deterministically using the CDF of a correlated bivariate normal distribution. However, if we want to use the Markovian approach, we need to compute the transition probabilities using a Monte Carlo simulation, which is a drawback for the method as it increases its computation time. We consider the following correlations

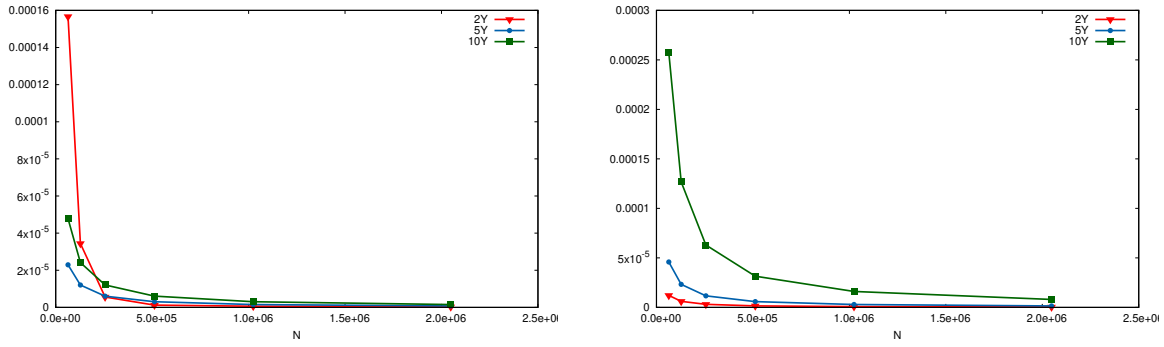
$$\rho_{Sf} = -0.0272, \qquad \rho_{Sd} = 0.1574, \qquad \rho_{df} = 0.6558.$$

Table 6.5 summarises the prices given by the closed-form formula.

	Exact price	
T \ σ	$50bp$	$500bp$
2Y	2.173803852	2.185536786
5Y	1.636518082	1.652226813
10Y	1.141944391	1.103531914

Table 6.5 *Prices given by closed-form formula of European options with correlations.* ($\sigma_d = \sigma_f = \sigma$)

Figures 6.5a and 6.5b display the relative error induced by the numerical method as a function of N. And in Table 6.6, we summarise the computation needed in order to reach a $1bp$ relative error.



(a) $\sigma_d = \sigma_f = 50bp$ (b) $\sigma_d = \sigma_f = 500bp$

Fig. 6.5 *Relative errors for the non-Markovian method for 2Y, 5Y and 10Y European options pricing (with correlations).*

	Non-Markovian – 2d	
T \ σ	$50bp$	$500bp$
2Y	71 ms (64000)	34 ms (32000)
5Y	31 ms (32000)	31 ms (32000)
10Y	32 ms (32000)	139 ms (128000)

Table 6.6 *Times in milliseconds needed for reaching a* $1bp$ *relative error of European options pricing with correlations using the non-Markovian method with, in parenthesis, the size* N *of the grid at each time step.* ($\sigma_d = \sigma_f = \sigma$)

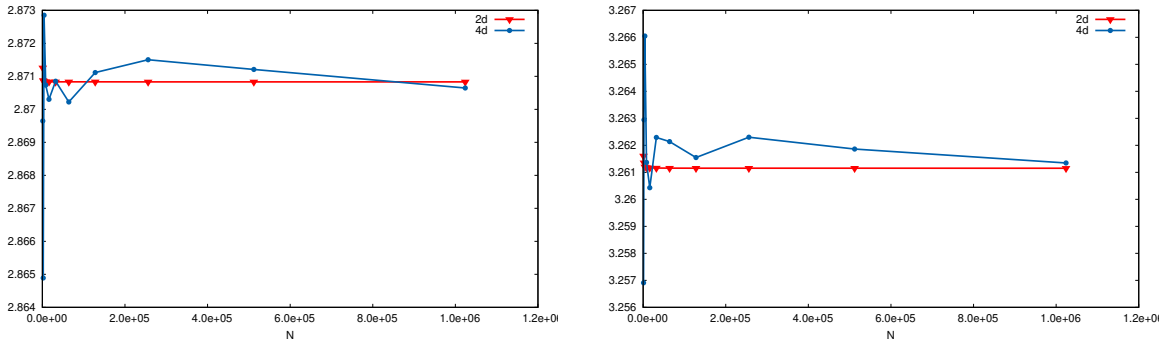
It is clear that one should prefer the non-Markovian methodology to the Markovian one for the evaluation of European options as it is a fast and accurate method for producing prices in the 3-factor model.

6.4.2 Bermudan option

Now, we compare the asymptotic behaviour of both approaches when pricing true Bermudan PRDC options. The following figures represent the price and the rescaled difference of the prices given by the two approaches as a function of N, which is the size of the product quantizer at each date (in two dimensions: $N = N^X \times N^Y$ and in four dimensions $N = N^X \times N^{W^f} \times N^Y \times N^{W^d}$). The financial products we consider are yearly exercisable Bermudan options with different values for the maturity date (2 years, 5 years and 10 years) and the domestic/foreign volatilities ($50bp$ and $500bp$).

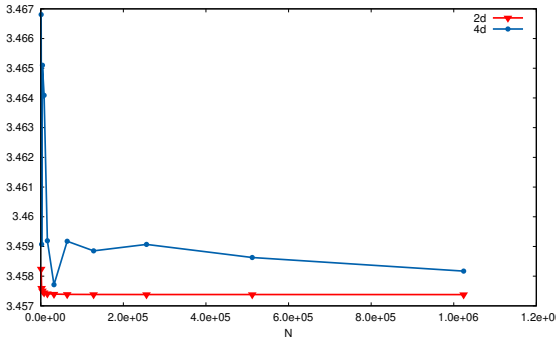
When using domestic and foreign volatilities close to market values, we observe numerically that the non-Markovian method converges a lot faster than the Markovian one for a given

complexity. However both methods do not converge to the same value (see Figures 6.6a, 6.6b, 6.6c), which is consistent with the results we found in Theorems 6.3.6 and 6.3.7. As in the European case, N in the label of each graph represents the size of the product quantizer ($N^X \times N^{W^f} \times N^Y \times N^{W^d}$ in the Markovian case and $N^X \times N^Y$ in the other case), hence the complexity of both trees are the same.



(a) *2Y*

(b) *5Y*



(c) *10Y*

Fig. 6.6 $\sigma_d = \sigma_f = 50bp$ – *Price with the two methods for 2Y, 5Y and 10Y yearly exercisable Bermudan options (with zero correlations).*

However, the relative systematic error induced by the non-Markovian methodology is negligible as can be seen in Figure 6.7, at most $5bp$ for a 10-year annual Bermudan option. Hence, one should prefer, again, the non-Markovian methodology when considering to evaluate Bermudan options.

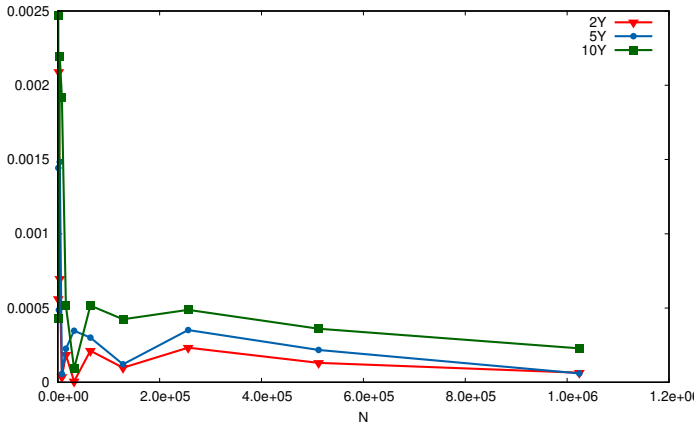


Fig. 6.7 $\sigma_d = \sigma_f = 50bp$ – *Relative differences between the two methods for 2Y, 5Y and 10Y yearly exercisable Bermudan options (with zero correlations).*

Remark. If we consider more exercise dates for the Bermudan option, the systematic errors increase, as shown in Figure 6.8 where we considered Bermudan options exercisable every 6 months and the same parameters as before with zero correlations and $\sigma_d = \sigma_f = 50bp$. However, even-though the error is higher for small N, when the non-Markovian pricer has converged, the relative difference between both methods is still acceptable (lower than $5bp$).

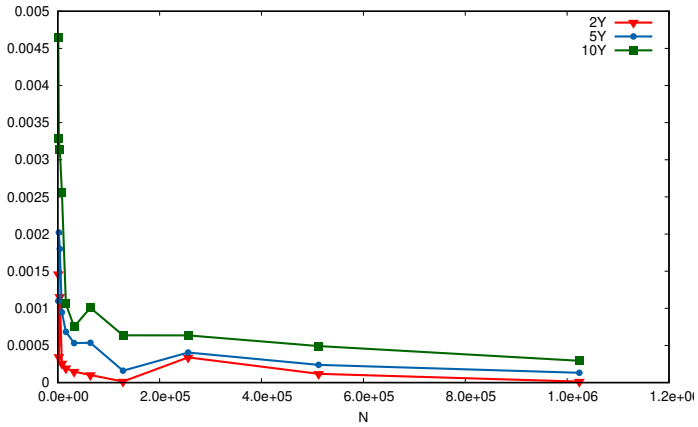
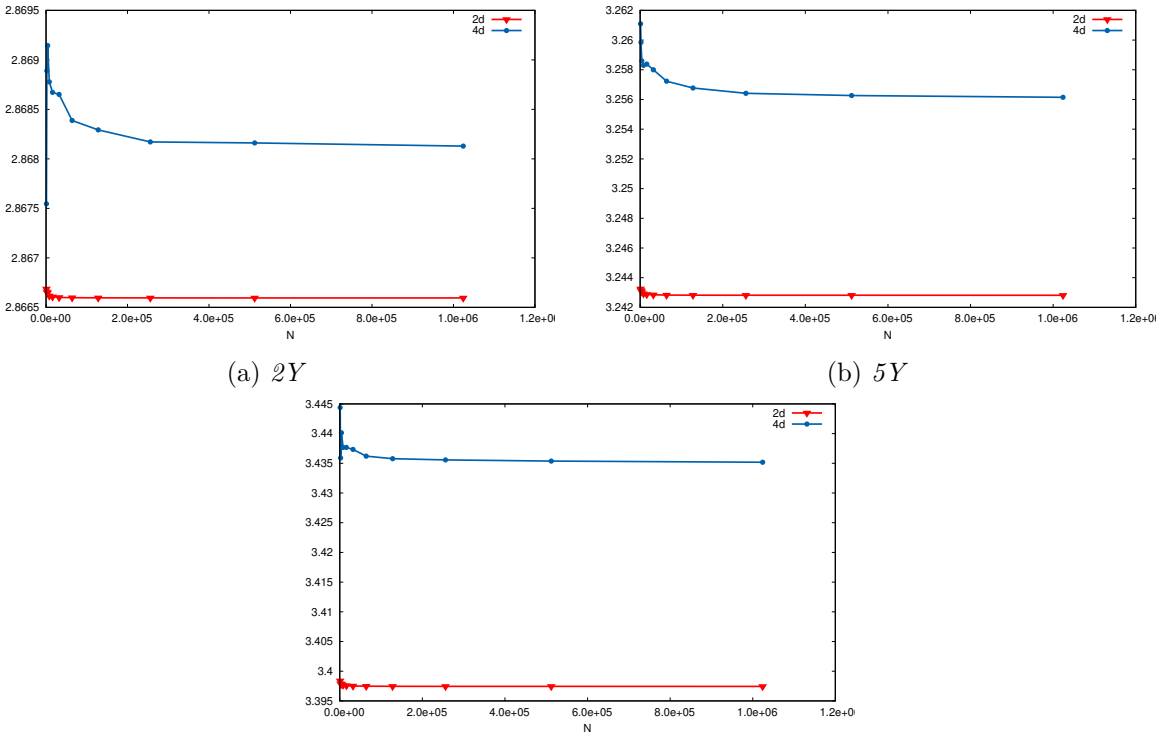


Fig. 6.8 $\sigma_d = \sigma_f = 50bp$ – *Relative differences between the two methods for 2Y, 5Y and 10Y bi-annual exercisable Bermudan options (with zero correlations).*

When we consider higher values the volatilities, $\sigma_d = \sigma_f = 500bp$, as expected the prices produced by the non-Markovian methodology produce a systematic error bigger than the case where $\sigma_d = \sigma_f = 50bp$ (see Figures 6.9a, 6.9b, 6.9c and 6.10). However, the relative difference between the two methods after convergence is reasonable: less than 0.1% for expiry 2 years, 0.4% for 5 years and around 1.1% for 10 years.



(a) *2Y* (b) *5Y*

(c) *10Y*

Fig. 6.9 $\sigma_d = \sigma_f = 500bp$ – *Price with the two methods for 2Y, 5Y and 10Y yearly exercisable Bermudan options (with zero correlations).*

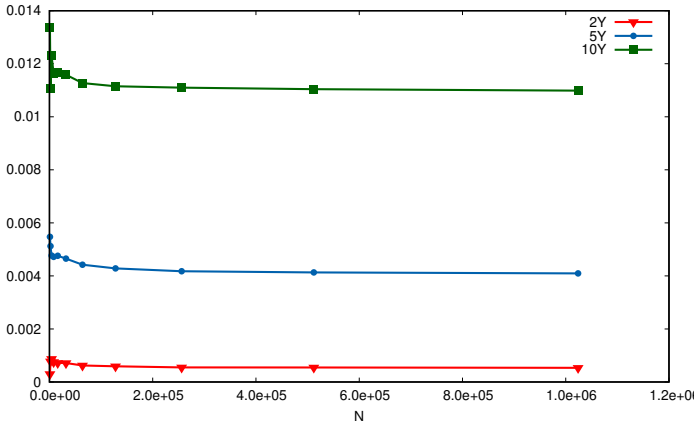


Fig. 6.10 $\sigma_d = \sigma_f = 500bp$ – *Relative differences between the two methods for 2Y, 5Y and 10Y yearly exercisable Bermudan options (with zero correlations).*

In Figure 6.7, we reference the time needed for reaching a $5bp$ relative precision (we compare the price given by grids of size N to the "asymptotic", which is the price given by the same method with a very large N) for the pricing of Bermudan options in a scenario of zero correlations. The non-Markovian method attains better precision than a relative precision

of $5bp$ in a few milliseconds, at most 7 ms where the Markovian one can need 4 seconds for reaching that precision. Hence, the 2 dimensional approximation seems again to be the better choice.

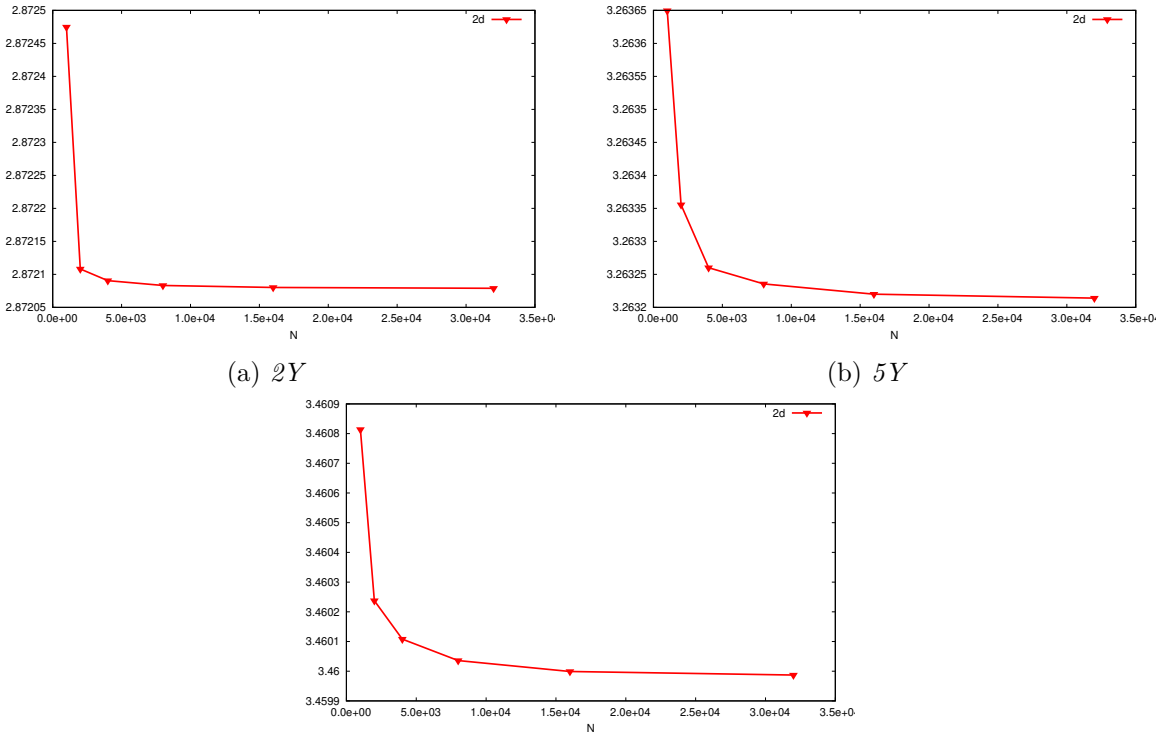
	Non-Markovian – 2d		Markovian – 4d	
T \ σ	$50bp$	$500bp$	$50bp$	$500bp$
2Y	1 ms (1000)	1 ms (1000)	25 ms (8000)	4 ms (1000)
5Y	3 ms (1000)	4 ms (1000)	98 ms (8000)	1903 ms (64000)
10Y	7 ms (1000)	7 ms (1000)	468 ms (16000)	3850 ms (64000)

Table 6.7 *Times in milliseconds needed for reaching a 5bp relative precision for Bermudan options pricing using both methods with zero correlation and, in parenthesis, the size N of the grid at each time step. ($\sigma_d = \sigma_f = \sigma$)*

Remark. Again, the pricers can even be used when we consider non-zero correlations and we choose to show only the asymptotic behaviour of the non-Markovian method, for the same reasons as the European case. We consider the same correlations as in the European case

$$\rho_{Sf} = -0.0272, \qquad \rho_{Sd} = 0.1574, \qquad \rho_{df} = 0.6558.$$

Figures 6.11a, 6.11b and 6.11c display the price given by the numerical method as a function of N and Table 6.8 summarises the computation time needed in order to do better than a $3bp$ precision.



(a) *2Y* (b) *5Y*

(c) *10Y*

Fig. 6.11 $\sigma_d = \sigma_f = 50bp$ – *Price of 2Y, 5Y and 10Y yearly exercisable Bermudan options using the non-Markovian method (with correlations).*

	Non-Markovian – 2d
T \ σ	$50bp$
2Y	122 ms (1000)
5Y	553 ms (1000)
10Y	1283 ms (1000)

Table 6.8 *Times in milliseconds needed for reaching a 3bp relative precision for Bermudan yearly exercisable options pricing with correlations using the non-Markovian method with, in parenthesis, the size N of the grid at each time step. ($\sigma_d = \sigma_f = \sigma$)*

Conclusion

We were looking for a numerical method able to produce accurate prices of Bermudan PRDC options with a 3-factor model in a very short time because the pricing of such products arises in a more complex framework: the computation of counterparty risk measures, also called

xVA's. We proposed two numerical methods based on product optimal quantization with a preference for the non-Markovian one. Indeed, even if we introduce a systematic error with our approximation, the error is controlled, as long as the volatilities of the domestic and foreign interest rates stay reasonable. Moreover, the numerical tests we conducted confirmed that idea: we are able to produce prices of Bermudan options in the 3-factor model in a fast and accurate way.

Appendix 6.A W^f is a Brownian motion under the domestic risk-neutral measure

Let $(\widetilde{W}^f)$ a $\widetilde{\mathbb{P}}$-Brownian motion. In this section, we show that the process W^f defined by

$$dW_s^f = d\widetilde{W}_s^f + \rho_{Sf}\sigma_S ds \tag{6.112}$$

is a $\mathbb{P}$-Brownian motion.

First, we define the following change of numéraire, where $\widetilde{\mathbb{P}}$ is the foreign risk-neutral probability and $\mathbb{P}$ is the domestic risk-neutral probability,r

$$\begin{aligned} d\widetilde{\mathbb{P}} &= \frac{S_T}{S_0}\exp\left(-\int_0^T r_s^d ds\right)\exp\left(\int_0^T r_s^f ds\right)d\,\mathbb{P} \\ &= \exp\left(\sigma_S W_T^S - \frac{\sigma_S^2}{2}T\right)d\,\mathbb{P} \end{aligned}$$

or equivalently

$$\begin{aligned} d\,\mathbb{P} &= \exp\left(-\sigma_S W_T^S + \frac{\sigma_S^2}{2}T\right)d\widetilde{\mathbb{P}} \\ &= \exp\left(-\sigma_S(W_T^S - \sigma_S T) - \frac{\sigma_S^2}{2}T\right)d\widetilde{\mathbb{P}} \\ &= \exp\left(-\sigma_S\widetilde{W}_T^S - \frac{\sigma_S^2}{2}T\right)d\widetilde{\mathbb{P}} \end{aligned} \tag{6.113}$$

where $\widetilde{W}^S$ is a $\widetilde{\mathbb{P}}$-Brownian motion defined by $d\widetilde{W}_t^S = dW_t^S - \sigma_S dt$. More details concerning the definition of the foreign risk-neutral probability can be found in the Chapter 9 of [Shr04].

Now, we are looking for $q \in \mathbb{R}$ such that $dW_s^f = d\widetilde{W}_s^f - qdt$ is a $\mathbb{P}$-Brownian motion. Let $\lambda \in \mathbb{R}$ and $\forall t > s$

$$\begin{aligned} \mathbb{E}\left[\,\mathrm{e}^{\lambda\left((\widetilde{W}_t^f - qt) - (\widetilde{W}_s^f - qs)\right)} \mid \mathcal{F}_s\,\right] &= \widetilde{\mathbb{E}}\left[\,\mathrm{e}^{\lambda\left((\widetilde{W}_t^f - qt) - (\widetilde{W}_s^f - qs)\right) - \sigma_S(\widetilde{W}_T^S - \widetilde{W}_s^S) - \frac{\sigma_S^2}{2}(T-s)} \mid \mathcal{F}_s\,\right] \\ &= \widetilde{\mathbb{E}}\left[\,\mathrm{e}^{\lambda\left((\widetilde{W}_t^f - qt) - (\widetilde{W}_s^f - qs)\right) - \sigma_S(\widetilde{W}_t^S - \widetilde{W}_s^S) - \frac{\sigma_S^2}{2}(t-s)} \mid \mathcal{F}_s\,\right] \\ &= \mathrm{e}^{-\lambda q(t-s) - \frac{\sigma_S^2}{2}(t-s)}\,\widetilde{\mathbb{E}}\left[\,\mathrm{e}^{\lambda(\widetilde{W}_t^f - \widetilde{W}_s^f) - \sigma_S(\widetilde{W}_t^S - \widetilde{W}_s^S)} \mid \mathcal{F}_s\,\right] \\ &= \mathrm{e}^{-\lambda q(t-s) - \frac{\sigma_S^2}{2}(t-s)}\,\mathrm{e}^{\frac{\lambda^2}{2}(t-s) - \lambda\sigma_S\rho_{Sf}(t-s) + \frac{\sigma_S^2}{2}(t-s)} \\ &= \mathrm{e}^{\frac{\lambda^2}{2}(t-s)}\,\mathrm{e}^{-\lambda q(t-s) - \lambda\sigma_S\rho_{Sf}(t-s)} \\ &= \mathrm{e}^{\frac{\lambda^2}{2}(t-s)} \end{aligned} \tag{6.114}$$

the last equality is ensured if and only if

$$0 = -\lambda q(t-s) - \lambda\sigma_S\rho_{Sf}(t-s) \quad \Longleftrightarrow \quad q = -\sigma_S\rho_{Sf}. \tag{6.115}$$

Hence, W^f defined by

$$dW_s^f = d\widetilde{W}_s^f + \rho_{Sf}\sigma_S ds$$

is a $\mathbb{P}$-Brownian motion.

Appendix 6.B FX Derivatives - European Call

The payoff at maturity t of a European Call on FX rate is given by

$$(S_t - K)_+$$

with K the strike and S_t the FX rate at time t.

Our aim will be to evaluate V_0

$$V_0 = \mathbb{E}\left[\mathrm{e}^{-\int_0^t r_s^d ds}(S_t - K)_+\right].$$

Proposition 6.B.1. *If we consider a 3-factor model on Foreign Exchange and Zero-coupon as defined in (6.6), V_0 is given by*[2]

$$V_0 = S_0 P^f(0,t)\,\mathcal{N}\left(\frac{\log\left(\frac{S_0 P^f(0,t)}{K P^d(0,t)}\right) + \mu(0,t)}{\sigma(0,t)}\right) - K P^d(0,t)\,\mathcal{N}\left(\frac{\log\left(\frac{S_0 P^f(0,t)}{K P^d(0,t)}\right) - \mu(0,t)}{\sigma(0,t)}\right)$$

with

$$\begin{aligned}\mu(0,t) = &\int_0^t \frac{1}{2}\big(\sigma_S^2(s) + \sigma_f^2(s,t) + \sigma_d^2(s,t)\big)ds \\ &+ \int_0^t \big(\rho_{Sf}\sigma_S(s)\sigma_f(s,t) - \rho_{Sd}\sigma_S(s)\sigma_d(s,t)ds - \rho_{fd}\sigma_f(s,t)\sigma_d(s,t)\big)ds\end{aligned}$$

and

$$\sigma^2(0,t) = 2\mu(0,t).$$

Proof. In this part, we want to evaluate

$$V_0 = \mathbb{E}\left[\mathrm{e}^{-\int_0^t r_s^d ds}(S_t - K)_+\right].$$

[2] We ignore the settlements details in the present paper in order to alleviate the notations but the formula can easily be extended to take them into account.

If we consider a 3-factor model on Foreign Exchange and Zero-coupon as defined in (6.6), we have

$$\begin{aligned}
V_0 &= \mathbb{E}\left[\mathrm{e}^{-\int_0^t r_s^d ds}(S_t - K)_+ \right] \\
&= \mathbb{E}\left[\left(\mathrm{e}^{-\int_0^t r_s^d ds}\, S_t - \mathrm{e}^{-\int_0^t r_s^d ds}\, K \right)_+ \right] \\
&= \mathbb{E}\left[\left(\mathrm{e}^{-\int_0^t r_s^d ds}\, S_t - \mathrm{e}^{-\int_0^t r_s^d ds}\, K \right) \mathbb{1}_{\{S_t \geqslant K\}} \right] \\
&= \mathbb{E}\left[\mathrm{e}^{-\int_0^t r_s^d ds}\, S_t\, \mathbb{1}_{\{S_t \geqslant K\}} \right] - K\, \mathbb{E}\left[\mathrm{e}^{-\int_0^t r_s^d ds}\, \mathbb{1}_{\{S_t \geqslant K\}} \right].
\end{aligned}$$

We focus on the first term

$$K\, \mathbb{E}\left[\mathrm{e}^{-\int_0^t r_s^d ds}\, \mathbb{1}_{\{S_t \geqslant K\}} \right]. \tag{6.116}$$

We do the following change of numéraire:

$$\frac{d\widetilde{\mathbb{Q}}}{d\,\mathbb{P}} = \frac{\widetilde{Z}_t}{\widetilde{Z}_0}$$

with

$$\begin{cases} \widetilde{Z}_t = \exp\left(\widetilde{Y}_t - \dfrac{1}{2} < \widetilde{Y}, \widetilde{Y} >_t \right), \\ \widetilde{Z}_0 = 1 \end{cases}$$

where $\widetilde{Y}_t = \int_0^t \sigma_d(s,t) dW_s^d$ and $< \widetilde{Y}, \widetilde{Y} >_t = \int_0^t \sigma_d^2(s,t) ds$.

Hence, we can define the following Brownian Motions $\widetilde{W}^d$, $\widetilde{W}^f$, $\widetilde{W}^S$ under $\widetilde{\mathbb{Q}}$:

$$\begin{aligned}
d\widetilde{W}_s^d &= dW_s^d - d < Y, W^d >_s &&= dW_s^d - \sigma_d(s,t) ds, \\
d\widetilde{W}_s^f &= dW_s^f - d < Y, W^f >_s &&= dW_s^f - \rho_{fd}\sigma_d(s,t) ds, \\
d\widetilde{W}_s^S &= dW_s^S - d < Y, W^S >_s &&= dW_s^S - \rho_{Sd}\sigma_d(s,t) ds
\end{aligned}$$

and S_t becomes

$$\begin{aligned}
S_t &= S_0 \exp\left(\int_0^t \left(r_s^d - r_s^f - \frac{\sigma_S^2(s)}{2}\right)ds + \int_0^t \sigma_S(s)dW_s^S\right) \\
&= \frac{S_0 P^f(0,t)}{P^d(0,t)} \exp\left(\int_0^t -\frac{1}{2}\big(\sigma_S^2(s) + \sigma_f^2(s,t) - \sigma_d^2(s,t)\big) - \rho_{Sf}\sigma_S(s)\sigma_f(s,t)\ ds\right) \\
&\qquad \times \exp\left(\int_0^t \sigma_S(s)dW_s^S + \int_0^t \sigma_f(s,t)dW_s^f - \int_0^t \sigma_d(s,t)dW_s^d\right) \\
&= \frac{S_0 P^f(0,t)}{P^d(0,t)} \exp\left(-\int_0^t \frac{1}{2}\big(\sigma_S^2(s) + \sigma_f^2(s,t) + \sigma_d^2(s,t)\big)\ ds\right) \\
&\qquad \times \exp\left(-\int_0^t \big(\rho_{Sf}\sigma_S(s)\sigma_f(s,t) - \rho_{Sd}\sigma_S(s)\sigma_d(s,t) - \rho_{fd}\sigma_f(s,t)\sigma_d(s,t)\big)ds\right) \\
&\qquad\qquad \times \exp\left(\int_0^t \sigma_S(s)d\widetilde{W}_s^S + \int_0^t \sigma_f(s,t)d\widetilde{W}_s^f - \int_0^t \sigma_d(s,t)d\widetilde{W}_s^d\right) \\
&= \frac{S_0 P^f(0,t)}{P^d(0,t)} \exp\left(-\mu(0,t) + \int_0^t \sigma_S(s)d\widetilde{W}_s^S + \int_0^t \sigma_f(s,t)d\widetilde{W}_s^f - \int_0^t \sigma_d(s,t)d\widetilde{W}_s^d\right).
\end{aligned}$$

Hence, as $\exp\left(-\int_0^t r_s^d ds\right) = P^d(0,t) \times \widetilde{Z}_t$, (6.116) becomes

$$\begin{aligned}
K\,\mathbb{E}\left[\mathrm{e}^{-\int_0^t r_s^d ds}\,\mathbb{1}_{\{S_t \geqslant K\}}\right] &= K P^d(0,t)\,\mathbb{E}^{\tilde{\mathbb{Q}}}\left[\mathbb{1}_{\{S_t \geqslant K\}}\right] \\
&= K P^d(0,t)\tilde{\mathbb{Q}}(S_t \geqslant K) \\
&= K P^d(0,t)\tilde{\mathbb{Q}}\left(Z \geqslant \frac{\log\left(\frac{K P^d(0,t)}{S_0 P^f(0,t)}\right) + \mu(0,t)}{\sigma(0,t)}\right) \\
&= K P^d(0,t)\tilde{\mathbb{Q}}\left(Z \leqslant \frac{\log\left(\frac{S_0 P^f(0,t)}{K P^d(0,t)}\right) - \mu(0,t)}{\sigma(0,t)}\right) \\
&= K P^d(0,t)\,\mathcal{N}\left(\frac{\log\left(\frac{S_0 P^f(0,t)}{K P^d(0,t)}\right) - \mu(0,t)}{\sigma(0,t)}\right)
\end{aligned}$$

where $Z \sim \mathcal{N}(0,1)$ with

$$\begin{aligned}
\mu(0,t) &= \int_0^t \frac{1}{2}\big(\sigma_S^2(s) + \sigma_f^2(s,t) + \sigma_d^2(s,t)\big)ds \\
&\quad + \int_0^t \big(\rho_{Sf}\sigma_S(s)\sigma_f(s,t) - \rho_{Sd}\sigma_S(s)\sigma_d(s,t)ds - \rho_{fd}\sigma_f(s,t)\sigma_d(s,t)\big)ds, \\
\sigma^2(0,t) &= \mathbb{V}\mathrm{ar}\left(\int_0^t \sigma_S(s)d\widetilde{W}_s^S + \int_0^t \sigma_f(s,t)d\widetilde{W}_s^f - \int_0^t \sigma_d(s,t)d\widetilde{W}_s^d\right) \\
&= \mathbb{V}\mathrm{ar}\left(\int_0^t \sigma_S(s)d\widetilde{W}_s^S\right) + \mathbb{V}\mathrm{ar}\left(\int_0^t \sigma_f(s,t)d\widetilde{W}_s^f\right) + \mathbb{V}\mathrm{ar}\left(\int_0^t \sigma_d(s,t)d\widetilde{W}_s^d\right) \\
&\quad + 2\,\mathbb{C}\mathrm{ov}\left(\int_0^t \sigma_S(s)d\widetilde{W}_s^S, \int_0^t \sigma_f(s,t)d\widetilde{W}_s^f\right) - 2\,\mathbb{C}\mathrm{ov}\left(\int_0^t \sigma_S(s)d\widetilde{W}_s^S, \int_0^t \sigma_d(s,t)d\widetilde{W}_s^d\right) \\
&\quad - 2\,\mathbb{C}\mathrm{ov}\left(\int_0^t \sigma_f(s,t)d\widetilde{W}_s^f, \int_0^t \sigma_d(s,t)d\widetilde{W}_s^d\right) \\
&= \int_0^t \big(\sigma_S^2(s) + \sigma_f^2(s,t) + \sigma_d^2(s,t)\big)ds \\
&\quad + 2\int_0^t \big(\rho_{Sf}\sigma_S(s)\sigma_f(s,t) - \rho_{Sd}\sigma_S(s)\sigma_d(s,t) - \rho_{fd}\sigma_f(s,t)\sigma_d(s,t)\big)ds.
\end{aligned}$$

Now, we deal with the term

$$\mathbb{E}\left[\,\mathrm{e}^{-\int_0^t r_s^d ds}\, S_t\, \mathbb{1}_{\{S_t \geqslant K\}}\,\right] = P^d(0,t)\,\mathbb{E}^{\tilde{\mathbb{Q}}}\left[S_t\, \mathbb{1}_{\{S_t \geqslant K\}}\,\right] \tag{6.117}$$

using directly the formula of the first partial moment of a log-normal random variable. Let $X \sim \text{Log-}\mathcal{N}(\mu, \sigma^2)$, then

$$\mathbb{E}\left[X\, \mathbb{1}_{\{X \geqslant x\}}\,\right] = \mathrm{e}^{\mu + \frac{\sigma^2}{2}}\, \mathcal{N}\left(\frac{\mu + \sigma^2 - \log(x)}{\sigma}\right).$$

Finally, as $S_t = \frac{S_0 P^f(0,t)}{P^d(0,t)} X$ with $X \overset{\tilde{\mathbb{Q}}}{\sim} \text{Log-}\mathcal{N}(-\mu(0,t), \sigma^2(0,t))$, we get

$$\begin{aligned}
(6.117) &= S_0 P^f(0,t)\,\mathbb{E}^{\tilde{\mathbb{Q}}}\left[X\, \mathbb{1}_{\left\{X \geqslant \frac{K P^d(0,t)}{S_0 P^f(0,t)}\right\}}\right] \\
&= S_0 P^f(0,t)\,\mathrm{e}^{-\mu(0,t) + \frac{\sigma^2(0,t)}{2}}\, \mathcal{N}\left(\frac{-\mu(0,t) + \sigma^2(0,t) - \log\left(\frac{K P^d(0,t)}{S_0 P^f(0,t)}\right)}{\sigma(0,t)}\right) \\
&= S_0 P^f(0,t)\, \mathcal{N}\left(\frac{\log\left(\frac{S_0 P^f(0,t)}{K P^d(0,t)}\right) + \mu(0,t)}{\sigma(0,t)}\right)
\end{aligned}$$

noticing that $\mu(0,t) = \frac{\sigma^2(0,t)}{2}$.

Finally, we get

$$\begin{aligned}
V_0 &= \mathbb{E}\left[\, \mathrm{e}^{-\int_0^t r_s^d ds}(S_t - K)_+ \right] \\
&= \mathbb{E}\left[\, \mathrm{e}^{-\int_0^t r_s^d ds}\, S_t\, \mathbb{1}_{\{S_t \geqslant K\}} \right] - K\, \mathbb{E}\left[\, \mathrm{e}^{-\int_0^t r_s^d ds}\, \mathbb{1}_{\{S_t \geqslant K\}} \right] \\
&= S_0 P^f(0,t)\, \mathcal{N}\left(\frac{\log\left(\frac{S_0 P^f(0,t)}{K P^d(0,t)}\right) + \mu(0,t)}{\sigma(0,t)} \right) - K P^d(0,t)\, \mathcal{N}\left(\frac{\log\left(\frac{S_0 P^f(0,t)}{K P^d(0,t)}\right) - \mu(0,t)}{\sigma(0,t)} \right).
\end{aligned}$$

Special case of constant volatility: $\sigma_S(s) = \sigma_S$, $\sigma_d(s,t) = \sigma_d \times (t-s)$ $\sigma_f(s,t) = \sigma_f \times (t-s)$

$$\begin{aligned}
\mu(0,t) &= \int_0^t \frac{1}{2}\big(\sigma_S^2(s) + \sigma_f^2(s,t) + \sigma_d^2(s,t)\big) ds \\
&\qquad + \int_0^t \big(\rho_{Sf}\sigma_S(s)\sigma_f(s,t) - \rho_{Sd}\sigma_S(s)\sigma_d(s,t) - \rho_{fd}\sigma_f(s,t)\sigma_d(s,t)\big) ds \\
&= \int_0^t \frac{1}{2}\big(\sigma_S^2 + \sigma_f^2 (t-s)^2 + \sigma_d^2 (t-s)^2\big) ds \\
&\qquad + \int_0^t \rho_{Sf}\sigma_S\sigma_f(t-s) - \rho_{Sd}\sigma_S\sigma_d(t-s) - \rho_{fd}\sigma_f\sigma_d(t-s)^2 ds \\
&= \frac{1}{2}\left(\sigma_S^2 t + \sigma_f^2 \frac{t^3}{3} + \sigma_d^2 \frac{t^3}{3}\right) + \rho_{Sf}\sigma_S\sigma_f \frac{t^2}{2} - \rho_{Sd}\sigma_S\sigma_d \frac{t^2}{2} - \rho_{fd}\sigma_f\sigma_d \frac{t^3}{3}, \\
\sigma^2(0,t) &= \int_0^t \big(\sigma_S^2(s) + \sigma_f^2(s,t) + \sigma_d^2(s,t)\big) ds \\
&\qquad + 2\int_0^t \big(\rho_{Sf}\sigma_S(s)\sigma_f(s,t) - \rho_{Sd}\sigma_S(s)\sigma_d(s,t) - \rho_{fd}\sigma_f(s,t)\sigma_d(s,t)\big) ds \\
&= 2\mu(0,t).
\end{aligned}$$

□

References

[Alb+07] Hansjörg Albrecher, Philipp Arnold Mayer, Wim Schoutens, and Jurgen Tistaert. "The little Heston trap". English. In: *Wilmott* 1 (2007), pp. 83–92. ISSN: 1540-6962.

[Alf05] Aurélien Alfonsi. "On the discretization schemes for the CIR (and Bessel squared) processes". In: *Monte Carlo Methods and Applications mcma* 11.4 (2005), pp. 355–384. DOI: 10.1515/156939605777438569.

[And07] Leif BG Andersen. "Efficient simulation of the Heston stochastic volatility model". In: *SSRN Electronic Journal* (2007). DOI: 10.2139/ssrn.946405.

[And65] Donald G Anderson. "Iterative procedures for nonlinear integral equations". In: *Journal of the ACM* 12.4 (1965), pp. 547–560. DOI: 10.1145/321296.321305.

[BP03] Vlad Bally and Gilles Pagès. "A quantization algorithm for solving multidimensional discrete-time optimal stopping problems". In: *Bernoulli* 9.6 (2003), pp. 1003–1049. DOI: 10.3150/bj/1072215199.

[BPP05] Vlad Bally, Gilles Pagès, and Jacques Printems. "A quantization tree method for pricing and hedging multi-dimensional American options". In: *Mathematical Finance* 15.1 (2005), pp. 119–168. DOI: 10.1111/j.0960-1627.2005.00213.x.

[BPP01] Vlad Bally, Gilles Pagès, and Jacques Printems. "A Stochastic Quantization Method for Nonlinear Problems". In: *Monte Carlo Methods and Applications* 7 (2001), pp. 21–34. DOI: 10.1515/mcma.2001.7.1-2.21.

[Bar+96] Bradford C Barber, David P Dobkin, David P Dobkin, and Hannu Huhdanpaa. "The quickhull algorithm for convex hulls". In: *ACM Transactions on Mathematical Software (TOMS)* 22.4 (1996), pp. 469–483. DOI: 10.1145/235815.235821.

[BBP09] Olivier Bardou, Sandrine Bouthemy, and Gilles Pagès. "Optimal quantization for the pricing of swing options". In: *Applied Mathematical Finance* 16.2 (2009), pp. 183–217. DOI: 10.1080/13504860802453218.

[BBP10] Olivier Bardou, Sandrine Bouthemy, and Gilles Pagès. "When are swing options bang-bang?" In: *International Journal of Theoretical and Applied Finance* 13.06 (2010), pp. 867–899. DOI: 10.1142/S0219024910006030.

[BB88] Jonathan Barzilai and Jonathan M Borwein. "Two-point step size gradient methods". In: *IMA journal of numerical analysis* 8.1 (1988), pp. 141–148. DOI: 10.1093/imanum/8.1.141.

[BS73] Fischer Black and Myron Scholes. "The pricing of options and corporate liabilities". In: *Journal of political economy* 81.3 (1973), pp. 637–654. DOI: 10.1086/260062.

[Bon+06] Joseph-Frédéric Bonnans, Jean Charles Gilbert, Claude Lemaréchal, and Claudia A Sagastizábal. *Numerical optimization: theoretical and practical aspects*. Springer Science & Business Media, 2006. DOI: 10.1007/978-3-540-35447-5.

[BGM97] Alan Brace, Dariusz Gatarek, and Marek Musiela. "The market model of interest rate dynamics". In: *Mathematical finance* 7.2 (1997), pp. 127–155. DOI: 10.1111/1467-9965.00028.

[BDD13] Adrien Brandejsky, Benoîte De Saporta, and François Dufour. "Optimal stopping for partially observed piecewise-deterministic Markov processes". In: *Stochastic Processes and their Applications* 123.8 (2013), pp. 3201–3238. DOI: 10.1016/j.spa.2013.03.006.

[BSD12] Adrien Brandejsky, Benoîte de Saporta, and François Dufour. "Numerical method for expectations of piecewise deterministic Markov processes". In: *Communications in Applied Mathematics and Computational Science* 7.1 (2012), pp. 63–104. DOI: 10.2140/camcos.2012.7.63.

[BZ13] Claude Brezinski and Michela Redivo Zaglia. *Extrapolation methods: theory and practice*. Vol. 2. Elsevier, 2013. DOI: 10.1007/BF02144109.

[BMP13] Damiano Brigo, Massimo Morini, and Andrea Pallavicini. *Counterparty credit risk, collateral and funding: with pricing cases for all asset classes*. Vol. 478. John Wiley & Sons, 2013. DOI: 10.1002/9781118818589.

[BW82] James Bucklew and Gary Wise. "Multidimensional asymptotic quantization theory with r th power distortion measures". In: *IEEE Transactions on Information Theory* 28.2 (1982), pp. 239–247. DOI: 10.1109/TIT.1982.1056486.

[Bur12] John Burkardt. *C++ library which evaluates the upper right tail of the bivariate normal distribution*. 2012. URL: https://people.sc.fsu.edu/~jburkardt/cpp_src/toms462/toms462.html (visited on 04/13/2012).

[CFG18] Giorgia Callegaro, Lucio Fiorin, and Martino Grasselli. "American Quantized Calibration in Stochastic Volatility". In: *Risk Magazine* (2018).

[CFG17] Giorgia Callegaro, Lucio Fiorin, and Martino Grasselli. "Pricing via recursive quantization in stochastic volatility models". In: *Quantitative Finance* 17.6 (2017), pp. 855–872. DOI: 10.1080/14697688.2016.1255348.

[CFG19] Giorgia Callegaro, Lucio Fiorin, and Martino Grasselli. "Quantization meets Fourier: a new technology for pricing options". In: *Annals of Operations Research* 282.1-2 (2019), pp. 59–86. DOI: 10.1007/s10479-018-3048-z.

[CM01] Peter Carr and Dilip Madan. "Optimal positioning in derivative securities". In: *Quantitative Finance* 1.1 (2001), pp. 19–37. DOI: 10.1080/713665549.

[CM99] Peter Carr and Dilip Madan. "Option valuation using the fast Fourier transform". In: *Journal of computational finance* 2.4 (1999), pp. 61–73. DOI: 10.21314/JCF.1999.043.

[CP15] Sylvain Corlay and Gilles Pagès. "Functional quantization-based stratified sampling methods". In: *Monte Carlo Methods and Applications* 21.1 (2015), pp. 1–32. DOI: 10.1515/mcma-2014-0010.

[CIR05] John C Cox, Jonathan E Ingersoll Jr, and Stephen A Ross. "A theory of the term structure of interest rates". In: *Theory of Valuation*. World Scientific, 2005, pp. 129–164. DOI: 10.2307/1911242.

[CBB14] Stéphane Crépey, Tomasz R Bielecki, and Damiano Brigo. *Counterparty risk and funding: A tale of two puzzles*. Chapman and Hall/CRC, 2014. DOI: 10.1201/9781315373621.

[CGM97] Juan Cuesta-Albertos, Alfonso Gordaliza, and Carlos Matrán. "Trimmed k-means: an attempt to robustify quantizers". In: *The Annals of Statistics* 25.2 (1997), pp. 553–576. DOI: 10.1214/aos/1031833664.

[DD12] Benoîte De Saporta and François Dufour. "Numerical method for impulse control of piecewise deterministic Markov processes". In: *Automatica* 48.5 (2012), pp. 779–793. DOI: 10.1016/j.automatica.2012.02.031.

[DFP04] Sylvain Delattre, Jean-Claude Fort, and Gilles Pagès. "Local Distortion and μ-Mass of the Cells of One Dimensional Asymptotically Optimal Quantizers". In: *Communications in Statistics - Theory and Methods* 33.5 (2004), pp. 1087–1117. DOI: 10.1081/STA-120029827.

[Del+04] Sylvain Delattre, Siegfried Graf, Harald Luschgy, and Gilles Pagès. "Quantization of probability distributions under norm-based distortion measures". In: *Statistics & Decisions* 22.4 (2004), pp. 261–282. DOI: 10.1016/j.jmaa.2005.06.022.

[Den10] Shaozhong Deng. *Quadrature formulas in two dimensions*. 2010. URL: https://pdfs.semanticscholar.org/4c92/2ee4effb71a78d9680a8646056e129d22cf1.pdf.

[Don73] Thomas G Donnelly. "Algorithm 462: Bivariate normal distribution". In: *Communications of the ACM* 16.10 (1973), p. 638. DOI: 10.1145/362375.362414.

[EFG96] Nicole El Karoui, Antoine Frachot, and Hélyette Geman. "A Note on the Behavior of Long Zero Coupon Rates in a No Arbitrage Framework". In: *Preprint* (1996).

[EMV92] Nicole El Karoui, Ravi Myneni, and Ramanarayanan Viswanathan. "Arbitrage pricing and hedging of interest rate claims with state variables, theory and applications". In: *Preprint* (1992).

[Fay+19] Jean-Michel Fayolle, Vincent Lemaire, Thibaut Montes, and Gilles Pagès. *Quantization-based Bermudan option pricing in the FX world*. 2019. arXiv: 1911.05462.

[FSP18] Lucio Fiorin, Abass Sagna, and Gilles Pagès. "Product Markovian Quantization of a Diffusion Process with Applications to Finance". In: *Methodology and Computing in Applied Probability* (2018), pp. 1–32. DOI: 10.1007/s11009-018-9652-1.

[FP02] Jean-Claude Fort and Gilles Pagès. "Asymptotics of optimal quantizers for some scalar distributions". In: *Journal of Computational and Applied Mathematics* 146.2 (2002), pp. 253–275. DOI: 10.1016/S0377-0427(02)00359-X.

[Gat11] Jim Gatheral. *The volatility surface: a practitioner's guide*. Vol. 357. John Wiley & Sons, 2011. DOI: 10.1002/9781119202073.

[GR09] Pierre Gauthier and Pierre-Yves Henri Rivaille. "Fitting the smile, smart parameters for SABR and Heston". In: *SSRN Electronic Journal* (2009). DOI: 10.2139/ssrn.1496982.

[GER95] Hélyette Geman, Nicole El Karoui, and Jean-Charles Rochet. "Changes of numeraire, changes of probability measure and option pricing". In: *Journal of Applied probability* (1995), pp. 443–458. DOI: 10.2307/3215299.

[GG82] Allen Gersho and Robert M Gray. "Special issue on Quantization". In: *IEEE Transactions on Information Theory* 29 (1982).

[Gla13] Paul Glasserman. *Monte Carlo methods in financial engineering*. Vol. 53. Springer Science & Business Media, 2013. DOI: 10.1007/978-0-387-21617-1.

[Gob+05] Emmanuel Gobet, Gilles Pagès, Huyên Pham, and Jacques Printems. "Discretization and simulation for a class of SPDEs with applications to Zakai and McKean-Vlasov equations". In: *Preprint, LPMA-958, Univ. Paris* 6 (2005).

[GL00] Siegfried Graf and Harald Luschgy. *Foundations of Quantization for Probability Distributions*. Berlin, Heidelberg: Springer-Verlag, 2000. DOI: 10.1007/BFb0103945.

[GLP08] Siegfried Graf, Harald Luschgy, and Gilles Pagès. "Distortion mismatch in the quantization of probability measures". In: *ESAIM: Probability and Statistics* 12 (2008), pp. 127–153. DOI: 10.1051/ps:2007044.

[Gre15] Jon Gregory. *The xVA Challenge: counterparty credit risk, funding, collateral and capital*. John Wiley & Sons, 2015. DOI: 10.1002/9781119161233.

[HJM92] David Heath, Robert Jarrow, and Andrew Morton. "Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation". In: *Econometrica: Journal of the Econometric Society* (1992), pp. 77–105. DOI: 10.2307/2951677.

[Hes93] Steven L Heston. "A closed-form solution for options with stochastic volatility with applications to bond and currency options". In: *The review of financial studies* 6.2 (1993), pp. 327–343. DOI: 10.1093/rfs/6.2.327.

[HW93] John Hull and Alan White. "One-factor interest-rate models and the valuation of interest-rate derivative securities". In: *Journal of financial and quantitative analysis* 28.2 (1993), pp. 235–254. DOI: 10.2307/2331288.

[HKA12] Farzana Hussain, MS Karim, and Razwan Ahamad. "Appropriate Gaussian quadrature formulae for triangles". In: *International Journal of Applied Mathematics and Computation* 4.1 (2012), pp. 24–38.

[IW81] Nobuyuki Ikeda and Shinzo Watanabe. *Stochastic differential equations and diffusion processes*. Vol. 24. North Holland, 1981. DOI: 10.1112/blms/14.5.449.

[JS17] Antoine Jacquier and Fangwei Shi. "The Randomized Heston Model". In: *SIAM Journal on Financial Mathematics* 10.1 (2017), pp. 89–129. DOI: 10.1137/18M1166420.

[Kie82] John C Kieffer. "Exponential rate of convergence for Lloyd's method I". In: *IEEE Transactions on Information Theory* 28.2 (1982), pp. 205–210. DOI: 10.1109/TIT.1982.1056482.

[Kie83] John C Kieffer. "Uniqueness of locally optimal quantizer for log-concave density and convex error weighting function". In: *IEEE Transactions on Information Theory* 29.1 (1983), pp. 42–47. DOI: 10.1109/TIT.1983.1056622.

[LL11] Damien Lamberton and Bernard Lapeyre. *Introduction to stochastic calculus applied to finance*. Chapman and Hall/CRC, 2011. DOI: 10.1155/S1048953398000094.

[LMP19] Vincent Lemaire, Thibaut Montes, and Gilles Pagès. "New weak error bounds and expansions for optimal quantization". In: *Journal of Computational and Applied Mathematics* (2019), p. 112670. ISSN: 0377-0427. DOI: 10.1016/j.cam.2019.112670.

[LMP20] Vincent Lemaire, Thibaut Montes, and Gilles Pagès. *Stationary Heston model: Calibration and Pricing of exotics using Product Recursive Quantization*. 2020. arXiv: 2001.03101.

[Llo82] Stuart Lloyd. "Least squares quantization in PCM". In: *IEEE transactions on information theory* 28.2 (1982), pp. 129–137. DOI: 10.1109/TIT.1982.1056489.

[Mac67] James MacQueen. "Some methods for classification and analysis of multivariate observations". In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Vol. 1. 14. Oakland, CA, USA. 1967, pp. 281–297.

[McW+18] Thomas A McWalter, Ralph Rudd, Jörg Kienitz, and Eckhard Platen. "Recursive marginal quantization of higher-order schemes". In: *Quantitative Finance* 18.4 (2018), pp. 693–706. DOI: 10.1080/14697688.2017.1402125.

[NM65] John A Nelder and Roger Mead. "A simplex method for function minimization". In: *The computer journal* 7.4 (1965), pp. 308–313.

[NP14] João Pedro Vidal Nunes and Pedro Miguel Silva Prazeres. "Pricing swaptions under multifactor Gaussian HJM models". In: *Mathematical Finance* 24.4 (2014), pp. 762–789. DOI: 10.1111/mafi.12019.

[Owe58] Donald B Owen. *Tables for computing bivariate normal probabilities*. Sandia Corporation, 1958.

[Pag98] Gilles Pagès. "A space quantization method for numerical integration". In: *Journal of computational and applied mathematics* 89.1 (1998), pp. 1–38. DOI: 10.1016/S0377-0427(97)00190-8.

[Pag15] Gilles Pagès. "Introduction to vector quantization and its applications for numerics". In: *ESAIM: proceedings and surveys* 48 (2015), pp. 29–79. DOI: 10.1051/proc/201448002.

[Pag07] Gilles Pagès. "Multi-step Richardson-Romberg extrapolation: remarks on variance control and complexity". In: *Monte Carlo Methods and Applications* 13.1 (2007), pp. 37–70. DOI: 10.1515/MCMA.2007.003.

[Pag18] Gilles Pagès. *Numerical Probability: An Introduction with Applications to Finance*. Springer, 2018. DOI: 10.1007/978-3-319-90276-0.

[PP09] Gilles Pagès and Fabien Panloup. "Approximation of the distribution of a stationary Markov process with application to option pricing". In: *Bernoulli* 15.1 (2009), pp. 146–177. DOI: 10.3150/08-BEJ142.

[PP05] Gilles Pagès and Huyên Pham. "Optimal quantization methods for nonlinear filtering with discrete-time observations". In: *Bernoulli* 11.5 (2005), pp. 893–932. DOI: 10.3150/bj/1130077599.

[PPP04a] Gilles Pagès, Huyên Pham, and Jacques Printems. "An optimal Markovian quantization algorithm for multi-dimensional stochastic control problems". In: *Stochastics and dynamics* 4.04 (2004), pp. 501–545. DOI: 10.1142/S0219493704001231.

[PPP04b] Gilles Pagès, Huyên Pham, and Jacques Printems. "Optimal Quantization Methods and Applications to Numerical Problems in Finance". In: *Handbook of computational and numerical methods in finance*. Ed. by Svetlozar T Rachev. Birkhäuser Boston, 2004, pp. 253–297. DOI: 10.1007/978-0-8176-8180-7_7.

[PP03] Gilles Pagès and Jacques Printems. "Optimal quadratic quantization for numerics: the Gaussian case". In: *Monte Carlo Methods and Applications* 9.2 (2003), pp. 135–165. DOI: 10.1515/156939603322663321.

[PS12] Gilles Pagès and Abass Sagna. "Asymptotics of the maximal radius of an L^r-optimal sequence of quantizers". In: *Bernoulli* 18.1 (2012), pp. 360–389. DOI: 10.3150/10-BEJ333.

[PS18a] Gilles Pagès and Abass Sagna. "Improved error bounds for quantization based numerical schemes for BSDE and nonlinear filtering". In: *Stochastic Processes and their Applications* 128.3 (2018), pp. 847–883. DOI: 10.1016/j.spa.2017.05.009.

[PS15] Gilles Pagès and Abass Sagna. "Recursive marginal quantization of the Euler scheme of a diffusion process". In: *Applied Mathematical Finance* 22.5 (2015), pp. 463–498. DOI: 10.1080/1350486X.2015.1091741.

[PS18b] Gilles Pagès and Abass Sagna. "Weak and strong error analysis of recursive quantization: a general approach with an application to jump diffusions". In: *arXiv preprint arXiv:1808.09755* (2018).

[PY16] Gilles Pagès and Jun Yu. "Pointwise convergence of the Lloyd I algorithm in higher dimension". In: *SIAM Journal on Control and Optimization* 54.5 (2016), pp. 2354–2382. DOI: 10.1137/151005622.

[PCR09] Huyên Pham, Marco Corsi, and Wolfgang Runggaldier. "Numerical approximation by quantization of control problems in finance under partial observations". In: *Handbook of Numerical Analysis*. Vol. 15. Elsevier, 2009, pp. 325–360. DOI: 10.1016/S1570-8659(08)00009-4.

[PRS05] Huyên Pham, Wolfgang Runggaldier, and Afef Sellami. "Approximation by quantization of the filter process and applications to optimal stopping problems under partial observation". In: *Monte Carlo Methods and Applications* 11.1 (2005), pp. 57–81. DOI: 10.1515/1569396054027283.

[Pit05] Vladimir Piterbarg. "A multi-currency model with FX volatility skew". In: *SSRN Electronic Journal* (2005). DOI: 10.2139/ssrn.685084.

[RH15] Isabelle Ramière and Thomas Helfer. "Iterative residual-based vector methods to accelerate fixed point iterations". In: *Computers & Mathematics with Applications* 70.9 (2015), pp. 2210–2226. DOI: 10.1016/j.camwa.2015.08.025.

[Ric10] Lewis Fry Richardson and Richard Tetley Glazebrook. "On the approximate arithmetical solution by finite differences of physical problems involving differential equations, with an application to the stresses in a masonry dam". In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 83.563 (1910), pp. 335–336. DOI: 10.1098/rspa.1910.0020.

[Rom55] Werner Romberg. "Vereinfachte numerische integration". In: *Norske Vid. Selsk. Forh.* 28 (1955), pp. 30–36.

[Rud+17] Ralph Rudd, Thomas McWalter, Jörg Kienitz, and Eckhard Platen. "Fast quantization of stochastic volatility models". In: *SSRN Electronic Journal* (2017). DOI: 10.2139/ssrn.2956168.

[Sag10] Abass Sagna. "Pricing of barrier options by marginal functional quantization". In: *Monte Carlo Methods and Applications* 17.4 (2010), pp. 371–398. DOI: 10.1515/mcma.2011.015.

[SST04] Wim Schoutens, Erwin Simons, and Jurgen Tistaert. "A Perfect calibration! Now what?" eng. In: *Wilmott Magazine* (2004). ISSN: 1540-6962. DOI: 10.1002/wilm.42820040216.

[She97] William Fleetwood Sheppard. "On the Calculation of the most Probable Values of Frequency-Constants, for Data arranged according to Equidistant Division of a Scale". In: *Proceedings of the London Mathematical Society* 1.1 (1897), pp. 353–380. DOI: 10.1112/plms/s1-29.1.353.

[Shr04] Steven E Shreve. *Stochastic calculus for finance II: Continuous-time models*. Vol. 11. Springer Science & Business Media, 2004. DOI: 10.1007/978-0-387-22527-2.

[Ste56] Hugo Steinhaus. "Sur la division des corps materiels en parties". In: *Bulletin de l'académie polonaise des sciences* 1.804 (1956), p. 801. DOI: 10.1371/journal.pone.0024999.

[Swa69] William Henry Swann. "A survey of non-linear optimization techniques". In: *FEBS letters* 2.S1 (1969), S39–S55. DOI: 10.1016/0014-5793(69)80075-X.

[TT90] Denis Talay and Luciano Tubaro. "Romberg extrapolations for numerical schemes solving stochastic differential equations". In: *Structural Safety* 8.1-4 (1990), pp. 143–150. DOI: 10.1016/0167-4730(90)90036-O.

[Vas77] Oldrich Vasicek. "An equilibrium characterization of the term structure". In: *Journal of financial economics* 5.2 (1977), pp. 177–188. DOI: 10.1016/0304-405X(77)90016-2.

[WN11] Homer F Walker and Peng Ni. "Anderson acceleration for fixed-point iterations". In: *SIAM Journal on Numerical Analysis* 49.4 (2011), pp. 1715–1735. DOI: 10.1137/10078356X.

[Wys17] Uwe Wystup. *FX options and structured products*. John Wiley & Sons, 2017. DOI: 10.1002/9781118673355.

[Zad82] Paul Zador. "Asymptotic quantization error of continuous signals and the quantization dimension". In: *IEEE Transactions on Information Theory* 28.2 (1982), pp. 139–149. DOI: 10.1109/TIT.1982.1056490.