HAL
open science

Some aspects of optimal quantization and applications to finance

Sylvain Corlay

To cite this version:

Sylvain Corlay. Some aspects of optimal quantization and applications to finance. Mathematics [math]. Université Pierre et Marie Curie - Paris VI, 2011. English. NNT : . tel-00626445

HAL Id: tel-00626445

https://theses.hal.science/tel-00626445

Submitted on 27 Sep 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse de doctorat de l'université Pierre et Marie Curie

Spécialité :
Mathématiques Appliquées

Présentée par :
Sylvain Corlay

Pour obtenir le grade de
Docteur en sciences de l'université Pierre et Marie Curie

Quelques aspects de la quantification optimale et applications à la finance

Sous la direction de Gilles Pagès.

Rapporteurs : Siegfried Graf
Benjamin Jourdain

Soutenue publiquement le 23 septembre 2011 devant le jury composé de

Frédéric Abergel	Examinateur
Michel Crouhy	Examinateur
Jean Jacod	Examinateur
Benjamin Jourdain	Rapporteur
Gilles Pagès	Directeur de thèse
Jacques Printems	Examinateur
Marc Yor	Examinateur

Remerciements

J'ai une profonde gratitude envers Gilles Pagès, qui a accompagné mes premiers pas dans la recherche en mathématiques. Je lui suis reconnaissant pour ce sujet de thèse stimulant et pour ses relectures minutieuses et exigeantes. J'ai tiré d'importants enseignements de nos discussions mathématiques tout au long de la thèse, autant d'occasions pour lui de m'aiguiller dans ma recherche. Je remercie Michel Crouhy et Frédéric Abergel, qui m'ont recruté pour cette thèse CIFRE dans l'équipe EDA (Equity Derivatives & Arbitrage) de Natixis, pour leur confiance, et Adil Reghai, qui a dirigé l'équipe de recherche quantitative pendant la majeure partie de ma présence dans l'entreprise. Cette expérience au sein d'une équipe de praticiens m'a donné les clefs pour tisser les liens entre théorie mathématique et applications pratiques. Je remercie Siegfried Graf et Benjamin Jourdain, rapporteurs de cette thèse, pour leur lecture détaillée du manuscrit et leurs remarques pertinentes. Je remercie Marc Yor pour sa grande disponibilité, par exemple quand il m'a orienté vers les travaux du Séminaire de Probabilités sur le grossissement initial de filtration. Je remercie aussi Jacques Printems et Jean Jacod d'avoir accepté de faire partie du jury.

La conclusion de cette thèse marquant la fin de mes études, il convient de remercier ici ceux qui m'ont apporté un soutien inconditionnel depuis le début, ma famille. Je remercie mes parents, ma mère pour son écoute et son affection et mon père pour ses conseils et sa clairvoyance et tous les deux pour leur soutien indéfectible. Vous demeurez un exemple à suivre dans votre façon d'aborder la vie, dans l'adversité ou le succès. Ma sœur Delphine et mes frères Vincent et Antoine m'ont entouré et aidé quelles que soient les circonstances. Je remercie Manuel Santana pour ses conseils et les nombreuses corrections en anglais. Merci aussi à Thioro, Cynthia, Léo et Martin Corlay pour tous ces bons moments que nous avons passés ensemble.

Je souhaite particulièrement remercier ici Johan Mabille. Informaticien talentueux, il a su me donner goût à la programmation en `C++` et à l'architecture logicielle. Johan m'a donné les clefs pour savoir surmonter les obstacles qui séparent les mathématiques de leur mise en pratique. Utiliser les méthodes de méta-programmation par les templates, faire appel à des librairies tierces, s'interfacer avec d'autres langages, écrire du code standard pouvant être utilisé sur différentes plateformes : tout cela n'aurait pas été possible sans cette interaction quotidienne fructueuse avec mon collègue et ami Johan.

Je suis très reconnaissant envers Marouen Messaoud, qui a relu le premier article sur la stratification fonctionnelle et a relevé de nombreuses coquilles. Je remercie aussi Adel Ben Haj Yedder pour sa disponibilité et son discernement. Adel et Marouen, les deux responsables des pôles « mono » et « multi » sous-jacents de l'équipe, sont tous les deux docteurs et ont su, malgré leurs responsabilités opérationnelles, garder un contact avec la recherche en mathématiques financières et son actualité. José Luu, responsable du calcul scientifique, m'a encouragé à explorer de nouvelles voies quand j'ai fait l'expérience d'instabilités numériques en géométrie algorithmique. C'est lui qui m'a suggéré d'utiliser le calcul en précision arbitraire et la librairie ARPREC. Il m'a beaucoup encouragé à collaborer avec son équipe et notamment avec Johan, sur le développement d'algorithmes numériques. Je lui en suis très reconnaissant.

Je remercie aussi Éric Bertrand, dont j'ai partagé le bureau pendant la première année de thèse, pour ses conseils et les discussions animées de nos déjeuners. Je remercie mon ami Claude Muller autant pour ce qu'il m'a appris sur les modèles de volatilité stochastique que pour les nombreux footings au parc de la cité universitaire et au Mont Valérien. Merci à Mathieu Decharrière pour son humour décapant et pour son recul sur le milieu professionnel.

Je tiens à remercier Gilles Boya près de qui j'ai travaillé pendant plusieurs mois, ainsi qu'Églantine Giraud. Gilles et Églantine, tous deux recrutés après moi sont très rapidement devenus les forces vives de l'équipe de quants. Gilles est très au fait des développements récents de la recherche en mathématiques financières. Ses connaissances précises en probabilités, sa curiosité et ses qualités pédagogiques font de lui un interlocuteur privilégié. J'ai eu beaucoup de plaisir à travailler avec lui et je le remercie pour ce qu'il m'a appris. Gilles, Églantine, Claude, Mathieu et Johan constituent un groupe que je retrouve avec plaisir à l'occasion des déjeuners chez « Lili et Marcel » dans une ambiance amicale. J'espère les recroiser et travailler avec eux dans l'avenir.

Enfin, je n'oublie pas les autres personnes avec qui j'ai pu travailler à Natixis : Albert Andinaik et Amine Boukhaffa, que j'ai encadrés pendant leurs stages et qui continuent aujourd'hui de travailler avec Natixis ; Geoffroy Querol, Vincent Garcin et Michael Irsutti, qui m'ont apporté beaucoup d'aide sur les langages du framework .NET et l'interface avec les langages natifs ; mais aussi Éric Cellier, Nicolas Huth, Ban Zheng, Pierre Lamy, Abdessamad Sahnoun, Loïc Grosman, Houari Houalef, Vincent Klayelé, Mounir Zeghari, Émilie Tetard, Yassine Faqri, Riadh Zaatour, Olivier Croissant, Laurent Jacquel, Bruno Fine, Christophe Héron, Fatima Elkhiari, Sanae Loulidi, Sidi Mohamed Ould Aly, Numa Lescot, Marc Souaille, Mohamed Lakhdar, Vincent Lusset, Alain Mounier, Emmanuel Candus, Gaël Riboulet et Julien Calas.

Je remercie mon ami Joachim Lebovits, qui a entamé sa thèse en même temps que moi, et avec qui je travaille depuis l'année du mastère de mathématiques financières. J'espère que notre projet de recherche aboutira bientôt. Je le remercie ici pour sa compagnie et sa clairvoyance ces trois dernières années. Je lui suis aussi très reconnaissant pour son éclairage sur les propriétés de l'intégrale de Wiener.

Merci à David Benoist, pour nos nombreuses discussions lors de nos rendez-vous hebdomadaires avec Joachim. J'espère que nous continuerons à nous réunir pour ces soirées pizzas, maths et piscine.

Pendant mon monitorat à l'ÉNS de Cachan, j'ai bénéficié des conseils et des ressources de mon tuteur pédagogique Nicolas Vayatis et du directeur du département de mathématiques, Alain Trouvé. Je les remercie pour leur implication, leur exigence et leurs conseils avisés sur l'enseignement des probabilités. J'ai toujours trouvé leur porte ouverte au CMLA. Je remercie également mes étudiants de l'ÉNS de Cachan pour leurs questions et remarques pertinentes pendant ces trois années de monitorat. Je saisis l'occasion pour remercier les professeurs qui ont marqué ma scolarité à l'ÉNS de Cachan ou à Jussieu - Laurent Desvillettes, Claudine Picaronny, Jean-Michel Morel et Stéphane Gaubert - ainsi que mes professeurs de classes préparatoires, Alexis Fagebaume et Jacques Malet.

Je n'oublierai pas le groupe d'élèves, amis, de ma promotion au département de mathématiques de l'ÉNS de Cachan, toujours prêts à plancher sur un nouvel « exo ». Guillaume Poly, Ayman Moussa, Rafik Imekraz, Dominique Malicet, et Martin Gaume. Une ambiance potache, leur talent et leur approche ludique des mathématiques ont fait de ce petit groupe le meilleur professeur de mathématiques possible. Je salue également Laetitia Borel-Mathurin, Fabrice Borel-Mathurin, Gabriel Gauguelin et Frédérique Charles.

Parmi les membres du LPMA, je remercie aussi Nicole El Karoui de m'avoir apporté son recul et sa hauteur de vue sur les mathématiques financières, les pratiques du marché et plus particulièrement sur les différents thèmes abordés dans son cours. Je remercie chaleureusement Jacques Portès qui a fourni les moyens informatiques nécessaires aux calculs que je devais effectuer, ce en plus de toutes les sollicitations auxquelles il répond quotidiennement.

Je tiens à saluer les personnes avec lesquelles j'ai partagé le quotidien au LPMA à la fin de la thèse : Reda Chhaibi avec qui j'avais déjà eu le plaisir de travailler quotidiennement à Natixis quand il y faisait son stage de mastère. Je remercie aussi Sophie Laruelle et Mathieu Richard qui sont les coorganisateurs du groupe de travail des thésards ainsi que Raoul Normand, Éric Luçon, Cécile Delaporte, Stavros Vakeroudis, Clément Foucart, Paul Bourgade, Nicolaos Karaliolios, Karim Bounebache, Sophie Dede et Noufel Frikha. Je salue Abass Sagna et Benedikt Wilbertz, qui ont travaillé sur des thématiques proches des miennes avec Gilles Pagès.

Un autre remerciement revient à l'immense communauté développant la variété de logiciels libres qui ont été utilisés pendant cette thèse. Tout ceci aurait certainement été impossible sans

utiliser LaTeX, ou GNU/Linux et GCC.

Merci à mes amis, Pablo Winant, Victorien Rami, Céline Pateron, Amélie Thévenet, Damien Lardoux, Hélène Gobin, Émilie Moreira, Jacinta Carvalho, Jérémy Ladron, Arnaud Cassan, Raphaël Rodriguez-Sierra, Aude Hofleitner, Nicolas Charles et Maëlle Nauroy pour leur affection et soutien. Enfin, je tiens tout particulièrement à remercier mon amie, Anna Mercier qui m'a accompagné, soutenu et encouragé ces deux dernières années et qui continue de le faire aujourd'hui.

Introduction

Ce document rassemble les résultats obtenus durant mes trois années de thèse sous la direction de Gilles Pagès. Il est constitué de cinq chapitres. Chaque chapitre est conçu comme un article indépendant comportant sa propre bibliographie, et est écrit en anglais.

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et E un espace de Banach réflexif séparable. La norme de E est notée $|\cdot|$. La quantification d'une variable aléatoire X prenant ses valeurs dans E consiste en son approximation par une variable aléatoire Y prenant un nombre fini de valeurs dans E. L'erreur résultant de cette discrétisation est mesurée par la norme L^p de $|X - Y|$. Si on se donne un cardinal maximal N pour $Y(\Omega)$, la minimisation de l'erreur revient au problème d'optimisation

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

Une solution au problème de minimisation (1) est appelée quantifieur optimal de X. La quantification optimale a d'abord été étudiée pour fournir une méthode de discrétisation de signal [5] et a ensuite été introduite dans le domaine des probabilités numériques pour concevoir des méthodes de cubature [12] ou pour résoudre des problèmes d'arrêt optimal multidimensionnels [3].
Le cas infini-dimensionnel est étudié depuis le début des années 2000, en particulier pour son application à la quantification fonctionnelle, autrement dit la quantification de variables aléatoires à valeurs dans des espaces fonctionnels. Cette étude a surtout porté sur le cas de la quantification L^2 sur des espaces de Hilbert [9], mais d'autres espaces de Banach ont aussi été considérés [19]. Les processus stochastiques sont vus comme des variables aléatoires prenant leurs valeurs dans les espaces de trajectoires considérés.

Cette thèse présente quelques aspects de la quantification optimale et leur application à la finance mathématique.

- Le premier chapitre porte sur l'application de la quantification optimale à la réduction de variance par stratification. En effet, des aspects théoriques de la stratification montrent un lien fort entre le problème de la quantification quadratique d'une variable aléatoire et la réduction de variance qui peut être atteinte par cette méthode. Pour commencer, nous soulignons la pertinence de la quantification pour définir les strates pour les méthodes d'échantillonnage stratifié dans les cas fini-dimensionnels et infini-dimensionnels. Ensuite, nous abordons le cas de la stratification fonctionnelle de processus gaussiens bi-mesurables. À cet effet, nous proposons un algorithme de simulation de complexité linéaire pour la loi conditionnelle des marginales d'un processus gaussien dans la cellule de Voronoi d'un quantifieur stationnaire de ce processus. La méthode est complètement spécifiée dans les cas du mouvement brownien, du pont brownien et des processus d'Ornstein-Uhlenbeck. Comme la quantification optimale effective d'un processus gaussien requiert la connaissance de sa base de Karhunen-Loève, bien connue dans les cas du mouvement brownien et du pont brownien, nous détaillons en annexe le calcul complet de la base de Karhunen-Loève des processus d'Ornstein-Uhlenbeck. Des tests numériques sont effectués sur des problèmes de valorisation d'options. Ce chapitre est le résultat d'un travail conjoint avec Gilles Pagès.

- Comme nous l'avons souligné plus haut, la connaissance de la base de Karhunen-Loève est nécessaire pour la construction effective d'un quantifieur quadratique de processus gaussien.

Le second chapitre souligne la possibilité d'utiliser des méthodes numériques d'approximation des solutions d'équations intégrales pour le calcul des bases de Karhunen-Loève de processus pour lesquels on ne dispose pas de formule fermée. Nous proposons d'utiliser la méthode dite de Nyström pour le problème de la quantification optimale de processus gaussiens. Dans les cas où on dispose d'une formule fermée de référence, nous montrons que la méthode de Nyström permet d'obtenir une précision proche de l'erreur machine. Ensuite, le cas particulier du mouvement brownien fractionnaire est traité. La cohérence des valeurs obtenues est vérifiée numériquement grâce à une méthode de « reconstruction » du processus gaussien. Enfin, cela nous permet d'appliquer la méthode de stratification fonctionnelle développée au premier chapitre au problème de la valorisation d'une option asiatique dans le modèle de Black & Scholes fractionnaire.

- Dans le troisième chapitre, nous proposons une nouvelle approche de la quantification fonctionnelle dans le cas d'une semimartingale gaussienne continue X que nous baptisons la « quantification partielle ». Cette approche consiste pour l'essentiel à ne quantifier que certaines coordonnées de X (en nombre fini) sur sa base de Karhunen-Loève. Le principal résultat est que conditionnellement à ces coordonnées, X reste une semimartingale par rapport à sa propre filtration. Ce résultat est établi en utilisant des techniques de grossissement de filtration. Ceci nous permet notamment de véritablement définir la stratification fonctionnelle d'une solution d'équation différentielle stochastique dirigée par la semimartingale X et de légitimer la méthode numérique utilisée dans le premier chapitre dans le cas des équations différentielles stochastiques. Nous prouvons également plusieurs résultats de convergence de la quantification partielle d'EDS.

- Dans le quatrième chapitre, nous proposons une méthode de cubature basée sur la quantification fonctionnelle pour la valorisation d'options vanilles dans le cas de modèles à volatilité stochastique. On se place d'abord dans le même cadre que dans l'article [14]. Ensuite, la méthode est étendue aux cas de modèles comportant un terme de volatilité locale, souvent appelés « modèles à volatilité locale stochastique ». Pour cela, nous proposons une nouvelle approximation que nous appelons la « quantification normale ». Cette méthode est basée sur les résultats relatifs à la quantification partielle de processus gaussiens introduite dans le chapitre précédent. Nous effectuons des tests numériques dans le cas du modèle SABR. Ce chapitre est le résultat d'un travail conjoint avec Gilles Pagès.

- Les recherches de plus proche voisin représentent une part critique de la plupart des algorithmes d'optimisation de grilles de quantification, ainsi que des algorithmes de réduction de variance utilisant un quantifieur Voronoi comme variable de contrôle. Dans le cinquième chapitre, nous proposons un nouvel algorithme de recherche de plus proche voisin lui-même basé sur une méthode de quantification vectorielle. La description complète de la méthode requiert l'exposé de quelques résultats de géométrie algorithmique relatifs aux diagrammes de Voronoi.

- Un intérêt de la quantification optimale, tant pour son application à la cubature que pour ses autres usages est que, une fois les quantifieurs calculés on peut les conserver pour un usage futur. Sur le site web `www.quantize.maths-fi.com` [15], une grande base de données de grilles de quantification de variables aléatoires gaussiennes est disponible au téléchargement. Les grilles gaussiennes unidimensionnelles sont calculées avec une précision relative de 10^{-32}, autrement dit, elles peuvent être considérées comme exactes dans leur représentation en simple, double et quadruple précision. En appendice, nous détaillons les méthodes utilisées pour obtenir ces grilles « sur-optimisées » pouvant s'appliquer pour obtenir des grilles de quantification en précision arbitraire.

0.1 Principaux résultats du chapitre 1

Le principe de l'échantillonnage stratifié est de localiser la méthode de Monte-Carlo sur les éléments d'une partition de l'espace d'état d'une variable aléatoire L^2, $X : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$. On se donne une partition $\mathcal{E}$-mesurable $(A_i)_{i\in I}$ de E, et nous appelons les éléments A_i strates. On suppose que les poids $p_i := \mathbb{P}[X \in A_i]$ sont connus et strictement positifs. On considère ensuite une famille de variables aléatoires indépendantes $(X_i)_{i\in I}$ de distribution $X_i \overset{\mathcal{L}}{\sim} \mathcal{L}(X|X \in A_i)$.

On suppose qu'on sait simuler les variables aléatoires X_i, ce qui équivaut à supposer qu'on peut écrire $X_i = \phi_i(U)$, où U est uniformément distribuée sur $[0,1]^{r_i}$ avec $r_i \in \mathbb{N}^*$ et $\phi : [0,1]^{r_i} \to \mathbb{R}$ est calculable facilement.

L'idée de la stratification est d'utiliser l'estimateur suivant pour calculer $\mathbb{E}[F(X)]$, où F est une fonctionnelle à valeurs réelles telle que $F(X) \in L^2$:

$$\overline{F(X)}_M^I := \sum_{i\in I} p_i \frac{1}{M_i} \sum_{k=1}^{M_i} F(X_i^k), \tag{2}$$

où M est le budget global de simulations de Monte-Carlo et $M_i := q_i M$ est le budget alloué pour le calcul de $\mathbb{E}[F(X_i)]$ dans chaque strate, et $(X_i^k)_{k\leq 1\leq M_i}$ sont M_i réalisations indépendantes selon $\mathcal{L}(X|X \in A_i)$. On impose naturellement que $\sum_{i\in I} q_i = 1$.

Cet estimateur est sans biais et sa variance est donnée par

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

où

$$\sigma_{F,i}^2 := \mathrm{Var}\left(F(X)|X \in A_i\right) = \mathrm{Var}\left(F(X_i)\right), \quad i \in I.$$

Choix des budgets $(q_i)_{i\in I}$ de tirages alloués à chaque strate

- Le choix naturel, mais « sous-optimal » est de fixer $q_i = p_i$ pour tout $i \in I$. Deux raisons pour un tel choix sont d'une part que la variance de l'estimateur obtenu est toujours inférieure à celle de l'estimateur standard de Monte-Carlo, et d'autre part que les poids p_i des strates sont généralement connus.

- Une autre possibilité est l'optimisation sous contrainte de la variance de l'estimateur (2), dont la solution est
$$q_i^* = \frac{p_i \sigma_{F,i}}{\sum_{j\in I} p_j \sigma_{F,j}}, \quad i \in I.$$
À ce stade, le problème est qu'on ne connaît pas explicitement les inerties locales $\sigma_{F,i}^2$. Dans l'article [18], Étoré et Jourdain proposent un algorithme qui modifie les proportions des simulations futures dans chaque strate de façon adaptative et qui converge vers l'allocation optimale.

Géométrie des strates

Maintenant, la principale inconnue est le choix de la partition $(A_i)_{i\in I}$. Ce choix est guidé par l'objectif de réduction de la variance, mais aussi par la nécessité de pouvoir simuler les lois conditionnelles $\mathcal{L}(X|X \in A_i)$. C'est l'objet de la section 0.1.1.

0.1.1 Pertinence de la quantification optimale quadratique pour concevoir les strates des méthodes d'échantillonnage stratifié

Le théorème 0.1.1 réunit les précédents résultats énoncés sur la stratification et met en avant un lien avec les notions d'inertie locale et d'inertie interclasse issues de la quantification quadratique.

Il suggère de choisir pour partition $(A_i)_{i\in I}$ les cellules de Voronoi associées à une quantification quadratique optimisée de X, et surtout une troisième possibilité pour le choix d'allocation des tirages entre les strates, qui a une efficacité uniforme sur les fonctionnelles F lipschitziennes.

Théorème 0.1.1 (Stratification universelle). *Soit $A = (A_i)_{i\in I}$ une partition de E et $\mathrm{Proj}_{A,Z}$ la projection barycentrique associée à la partition A pour la variable aléatoire Z (définition 1.1.4).*

1. *Pour tout $i \in I$, considérons l'inertie locale de la variable aléatoire X,*

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

Alors pour toute fonction lipschitzienne $F : E \to \mathbb{R}$,

$$\forall i \in I, \qquad \sigma_{F,i} \leq [F]_{\mathrm{Lip}}\sigma_i \qquad \text{de sorte que} \qquad \sup_{[F]_{\mathrm{Lip}}\leq 1} \sigma_{F,i} \leq \sigma_i. \tag{3}$$

2. *Dans le cas du choix sous-optimal,*

$$\begin{aligned} \sup_{[F]_{\mathrm{Lip}}\leq 1} \Big(\sum_{i\in I} p_i\sigma_{F,i}^2\Big) \leq \sum_{i\in I} p_i\sigma_i^2 &= \Big\|X - \mathbb{E}[X|\sigma(\{X \in A_i\}, i \in I)]\Big\|_2^2 \\ &= \Big\|X - \mathrm{Proj}_{A,X}(X)\Big\|_2^2. \end{aligned} \tag{4}$$

3. *Dans le cas du choix optimal,*

$$\sup_{[F]_{\mathrm{Lip}}\leq 1} \Big(\sum_{i\in I} p_i\sigma_{F,i}^2\Big) \leq \Big(\sum_{i\in I} p_i\sigma_i\Big)^2, \tag{5}$$

et

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

4. *Si on considère des fonctions lipschitziennes à valeurs vectorielles $F : E \to E$, alors les inégalités (3), (4) et (5) sont en fait des égalités.*

0.1.2 Quantification et simulabilité dans les strates

Dans le cas particulier où X est une variable aléatoire gaussienne de loi $\mathcal{N}(0, D)$, où D est une matrice diagonale, on montre qu'il est possible de simuler exactement la distribution conditionnelle de X dans un hyperrectangle avec un coût constant (voir section 1.3.3). Ce n'est pas le cas de la distribution conditionnelle dans un polytope quelconque. Pour cette raison, une stratification par quantification produit - résultant en des strates sous forme d'hyperrectangles - est préférable à la quantification optimale dont les cellules de Voronoi sont des polytopes plus généraux.

Nous abordons maintenant le problème de la stratification fonctionnelle de processus gaussiens bimesurables sur un intervalle $[0, T]$. Nous supposons que le processus considéré est L^2 et a une fonction de covariance continue sur l'intervalle considéré. Les quantifieurs quadratiques optimaux de tels processus se trouvent sur l'espace engendré par leurs premières fonctions propres de Karhunen-Loève. La forme des cellules de Voronoi associées dans $L^2([0, T])$ est donc très simple quand elle est exprimée dans cette base, en particulier dans le cas de la quantification produit. La simulation de la loi conditionnelle de marginales de X, $(X_{t_0}, \cdots X_{t_n})$ pour une subdivision $0 = t_0 \leq \cdots \leq t_n = T$ consiste alors

- tout d'abord à simuler la loi conditionnelle (fini-dimensionnelle) des premières coordonnées de Karhunen-Loève,
- puis la loi conditionnelle (gaussienne) des marginales de X connaissant ses premières coordonnées de Karhunen-Loève.

Dans le cas de la quantification produit, la première étape a déjà été traitée. La seconde étape consiste en un simple conditionnement gaussien, qui peut donc être effectué avec une complexité de $O(n^2)$ en utilisant une factorisation de Cholesky. Ce coût quadratique en le nombre de pas de temps n'étant pas satisfaisant, nous proposons un nouvel algorithme de complexité $O(d \times n)$ où d est la dimension de quantification. Cette dimension de quantification d est plutôt faible dans le cas des processus gaussiens considérés, pour lesquels elle est asymptotiquement équivalente au logarithme du nombre de strates. Par exemple, dans le cas où X est un mouvement brownien standard, la dimension de quantification pour une quantification de niveau $N = 10^4$ est de 9. C'est cet algorithme de simulation, utilisant une approche bayésienne qui rend la stratification fonctionnelle de processus gaussiens utilisable en pratique.

Sur la figure 1, on représente quelques trajectoires de la loi conditionnelle de 500 marginales d'un mouvement brownien standard sachant que ce mouvement brownien appartient à la cellule de Voronoi de la courbe épaissie sur le graphique. L'apparence des trajectoires obtenues suggère de considérer la méthode comme une « méthode de Monte-Carlo guidée ».

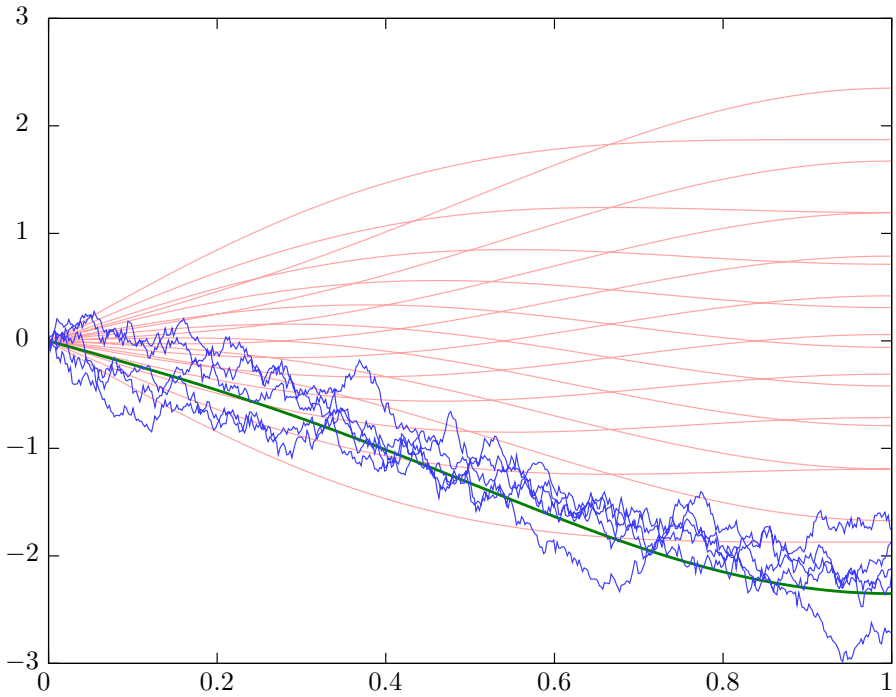

FIGURE 1 – Tracé de quelques réalisations de la loi conditionnelle du mouvement brownien standard sachant qu'il tombe dans la cellule de Voronoi L^2 de la courbe surlignée dans le quantifieur.

La méthode est appliquée à d'autres processus gaussiens, dans le chapitre 1 pour le pont brownien et les processus d'Ornstein-Uhlenbeck, et dans le chapitre 2 pour le mouvement brownien fractionnaire.

0.1.3 Quantification et stratification fonctionnelle des processus d'Ornstein-Uhlenbeck

Les fonctions propres de Karhunen-Loève ont une forme explicite dans les cas particuliers du mouvement brownien standard et du pont brownien. Le cas particulier du processus d'Ornstein-Uhlenbeck stationnaire avec paramètre de retour à la moyenne et volatilité égaux à 1 est traité dans le livre [6, p.195]. En appendice du premier chapitre, nous calculons la décomposition de Karhunen-Loève des processus d'Ornstein-Uhlenbeck de paramètres quelconques (variance initiale,

paramètre de retour à la moyenne et volatilité). Une procédure détaillée décrivant l'implémentation de la méthode est aussi fournie.

Grâce à ce calcul, la quantification quadratique optimale et la stratification fonctionnelle des processus d'Ornstein-Uhlenbeck sont possibles. Dans la suite, on vérifiera systématiquement si les résultats de cette thèse portant sur la quantification des processus gaussiens sont applicables aux processus suivants : le mouvement brownien standard, le pont brownien et les processus d'Ornstein-Uhlenbeck.

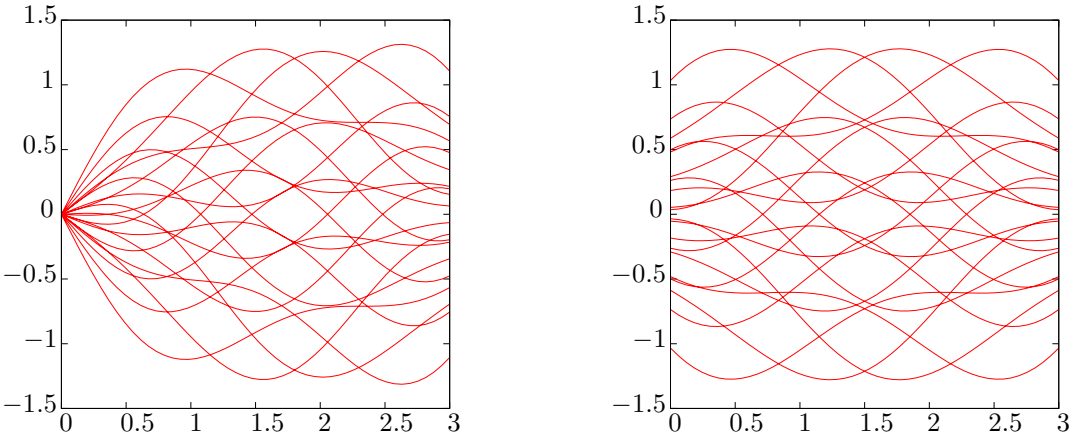

FIGURE 2 – Quantifieur produit optimal d'un processus d'Ornstein-Uhlenbeck centré partant de $r_0 = 0$ (à gauche) et stationnaire (à droite) défini par l'EDS $dr_t = -r_t dt + dW_t$, sur $[0, 3]$.

0.1.4 Application à la stratification de solutions d'équations différentielles stochastiques

Dans le cas plus restrictif où X est en fait une semimartingale gaussienne centrée partant de 0, on peut utiliser la stratification des marginales de X et les insérer dans le schéma d'Euler d'une équation différentielle stochastique pour obtenir une stratification fonctionnelle de la solution de l'équation différentielle stochastique considérée.

Cette approche de la stratification fonctionnelle pour les diffusions browniennes est justifiée à plusieurs égards :

- Sous certaines conditions sur les coefficients de l'EDS considérée, l'application qui au processus gaussien initial associe la solution de l'EDS est en fait une application lipschitzienne de $L^p([0,T])$ dans $L^p([0,T])$. De plus l'application qui à des marginales $(X_{t_0}, \cdots, X_{t_n})$ associe la suite des différences adjacentes $\left(X_{t_1} - X_{t_0}, \cdots, X_{t_n} - X_{t_{n-1}}\right)$ est aussi lipschitzienne. On reste ainsi dans le cadre du théorème 0.1.1 sur la stratification universelle.

- De plus, on verra au chapitre 3 que dans les cas cités précédemment, la loi conditionnelle de X sachant que X tombe dans une cellule de Voronoi donnée est une semimartingale par rapport à sa filtration naturelle. Cette propriété, démontrée plus loin dans la thèse par des arguments de grossissement de filtration permet de définir la stratification *continue* de la solution d'une EDS. Ainsi, la loi conditionnelle de la solution de l'EDS associée à X est en fait la solution de l'EDS associée à cette semimartingale. Utiliser ainsi les marginales conditionnelles dans le schéma d'Euler correspond en fait à implémenter le schéma d'Euler de l'EDS initiale conditionnée.

La fin de l'article présente quelques résultats numériques de cette méthode de réduction de variance appliquée à des problèmes de valorisation d'options.

0.2 Principaux résultats du chapitre 2

Comme souligné précédemment, le quantifieur quadratique optimal d'un processus gaussien bimesurable se trouve dans un plan principal de son opérateur de covariance, autrement dit, engendré par les premières fonctions propres de Karhunen-Loève. Pour cette raison, l'utilisation numérique d'un tel quantifieur nécessite d'avoir à sa disposition une méthode d'évaluation (rapide) de ses fonctions propres de Karhunen-Loève, ou au moins de celles qui correspondent aux plus grandes valeurs propres. Les bases de Karhunen-Loève du mouvement brownien, du pont brownien et des processus d'Ornstein-Uhlenbeck sont connues, mais on ne dispose pas de formule fermée dans le cas général.

Avoir à sa disposition une méthode numérique approchant les fonctions de base de Karhunen-Loève est le lien manquant pour la quantification d'autres processus gaussiens, comme le mouvement brownien fractionnaire. La quantification du mouvement brownien fractionnaire est pourtant intéressante en pratique car on dispose de beaucoup moins de méthodes numériques efficaces pour ce processus que pour le mouvement brownien standard, dont la simulation est beaucoup moins coûteuse.

Dans le chapitre 2, on applique la méthode dite « de Nyström » pour résoudre l'équation intégrale définissant le développement de Karhunen-Loève. La méthode est tout d'abord testée dans le cas des processus gaussiens cités précédemment pour lesquels on dispose de formules fermées pour leur base de Karhunen-Loève. Ensuite, on applique la méthode au cas du mouvement brownien fractionnaire. Dans ce cas, afin de tester la validité de la méthode, on réalise une « reconstruction » du processus initial, en le représentant comme la mixture de ses lois conditionnelles dans chacune de ses cellules de Voronoi. Ainsi, en utilisant la méthode de simulation détaillée au chapitre 1, on reconstruit théoriquement un mouvement brownien fractionnaire. La vérification consiste simplement à effectuer une estimation par la méthode de Monte-Carlo de la fonction de covariance du processus obtenu et de vérifier qu'on retrouve bien la fonction de covariance du mouvement brownien fractionnaire.

Dans le tableau 1, on reporte les résultats de cette méthode de Monte-Carlo avec 10 millions de tirage.

0.105061	0.138629	0.15846	0.173817	0.186687
0.138629	0.277258	0.330656	0.365844	0.394071
0.15846	0.330656	0.489116	0.557871	0.605929
0.173817	0.365844	0.557871	0.73168	0.813313
0.186687	0.394071	0.605929	0.813313	1

0.105141	0.138748	0.158596	0.173959	0.186824
0.138748	0.277417	0.330885	0.366075	0.394372
0.158596	0.330885	0.489454	0.558177	0.606266
0.173959	0.366075	0.558177	0.731923	0.813579
0.186824	0.394372	0.606266	0.813579	1.0003

TABLE 1 – Covariance théorique (à gauche) et estimée (à droite) $\mathbb{E}[X_{t_i}X_{t_j}]$ du mouvement brownien fractionnaire reconstruit, avec pour coefficient de Hurst $H = 0.7$. Le nombre de trajectoires utilisées pour cette simulation de Monte-Carlo est 1×10^7.

Nous avons donc maintenant une méthode fiable pour calculer les bases de Karhunen-Loève de processus gaussiens plus généraux, nous permettant de calculer leur quantification optimale. La figure 3 représente un quantifieur quadratique optimal du mouvement brownien fractionnaire avec $H = 0.25$.

0.3 Principaux résultats du chapitre 3

Le chapitre 3 apporte de nouveaux résultats théoriques sur la quantification fonctionnelle et la stratification. Tout le chapitre repose sur la notion de pont généralisé.

0.3.1 Les ponts généralisés

Soit X une semimartingale gaussienne centrée partant de 0 sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ de filtration naturelle $\mathcal{F}^X$ sur $[0, T]$. Le théorème de Fernique garantit que $\int_0^T \mathbb{E}\left[X_t^2\right] dt < +\infty$. Le

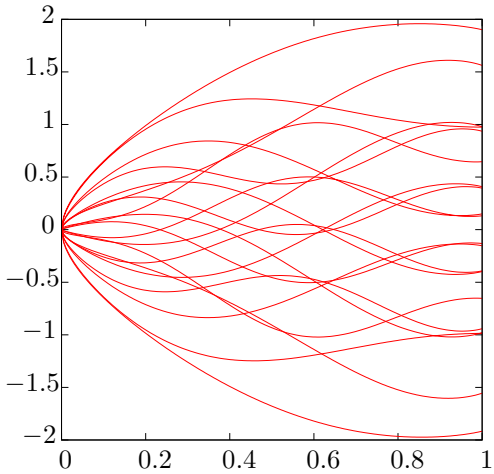

FIGURE 3 – Quantifieur N-optimal quadratique du mouvement brownien fractionnaire sur $[0,1]$ de coefficient de Hurst $H = 0.25$ avec $N = 20$.

but est de calculer le conditionnement par rapport à une famille finie $\overline{Z}_T := (Z_T^i)_{i\in I}$ de variables aléatoires gaussiennes, mesurables par rapport à $\sigma(X_t, t \in [0,T])$, où $I \subset \mathbb{N}$ est une partie finie de $\mathbb{N}^*$. Comme dans [1], on se restreint au cas où les variables aléatoires $(Z_T^i)_{i\in I}$ sont les valeurs finales de processus de la forme $Z_t^i = \int_0^t f_i(s)dX_s$, $i \in I$, pour une certaine famille finie $\overline{f} = (f_i)_{i\in I}$ de fonctions $L^2_{loc}([0,T])$. Le pont généralisé de $(X_t)_{t\in[0,T]}$ associé à $\overline{f}$ de valeur finale $\overline{z} = (z_i)_{i\in I}$ est le processus $\left(X^{\overline{f},\overline{z}}\right)_{t\in[0,T]}$ ayant pour distribution

$$X^{\overline{f},\overline{z}} \overset{\mathcal{L}}{\sim} \mathcal{L}\left(X \middle| Z_T^i = z_i, i \in I\right). \tag{6}$$

Le cas du pont brownien sur $[0,T]$ peut être obtenu en prenant X un mouvement brownien standard, $|I| = 1$, $\overline{f} = \{f\}$ et $f \equiv 1$.

En termes d'espaces de Hilbert gaussiens, si H est l'espace gaussien engendré par $(X_s)_{s\in[0,T]}$ et $H_{\overline{Z}_T}$ est le sous-espace fermé de H engendré par $(Z_T^i)_{i\in I}$, on note $H^{\perp}_{\overline{Z}_T}$ son complémentaire orthogonal dans H. Toute variable aléatoire (gaussienne) $G \in H$ se décompose en $G = \mathrm{Proj}_{\overline{Z}_T}(G) \overset{\perp\!\!\!\perp}{+} \mathrm{Proj}^{\perp}_{\overline{Z}_T}(G)$, où $\mathrm{Proj}_{\overline{Z}_T}$ et $\mathrm{Proj}^{\perp}_{\overline{Z}_T}$ sont les projections orthogonales sur $H_{\overline{Z}_T}$ et $H^{\perp}_{\overline{Z}_T}$. Avec ces notations, on a $\mathbb{E}\left[G \middle| (Z_T^i)_{i\in I}\right] = \mathrm{Proj}_{\overline{Z}_T}(G)$.

En fait, on va considérer des ponts browniens généralisés correspondant à certaines familles $\overline{f}$ particulières. Comme X est un processus gaussien continu, sa fonction de covariance est continue (voir [7, VIII.3]). On note alors $(e_i^X, \lambda_i^X)_{i\geq 1}$ ses fonctions propres et valeurs propres de Karhunen-Loève. Alors, si on définit la fonction f_i^X comme la primitive de $-e_i^X$ s'annulant en $t = T$, *i.e.* $f_i^X(t) = \int_t^T e_i^X(s)ds$, une intégration par partie donne

$$\int_0^T X_s e_i^X(s)ds = \int_0^T f_i^X(s)dX_s. \tag{7}$$

Pour une partie finie $I \subset \mathbb{N}^*$, on note $X^{I,\overline{y}}$ et on appelle *pont généralisé de Karhunen-Loève* le pont généralisé associé avec les fonctions $(f_i^X)_{i\in I}$ et ayant pour point final $\overline{y} = (y_i)_{i\in I}$. Ce processus a pour distribution $\mathcal{L}(X|Y_i = y_i, i \in I)$, où Y_i est la i-ème coordonnée de Karhunen-Loève.

0.3.2 Les ponts généralisés de Karhunen-Loève comme semimartingales

Le théorème de Jirina assure l'existence d'un noyau de transition

$$\nu_{\overline{Z}_T|\left((X_t)_{t\in[0,s]}\right)} : \mathcal{B}(\mathbb{R}^I) \times C^0\left([0,s],\mathbb{R}\right) \to \mathbb{R}_+,$$

correspondant à la loi conditionnelle $\mathcal{L}\left(\overline{Z}_t \middle| \left((X_t)_{t\in[0,s]}\right)\right)$.
On fait maintenant l'hypothèse supplémentaire $(\mathcal{H})$ que, pour tout $s \in [0,T)$ et toute fonction $(x_u)_{u\in[0,s]} \in C^0\left([0,s],\mathbb{R}\right)$, la loi de probabilité $\nu_{\overline{Z}_T|\left((X_t)_{t\in[0,s]}\right)}\left(d\overline{y},(x_u)_{u\in[0,s]}\right)$ est absolument continue par rapport à la mesure de Lebesgue. On note $\Pi_{(x_u)_{u\in[0,s]},T}$ sa densité. La matrice de covariance de cette distribution gaussienne sur $\mathbb{R}^I$ s'écrit

$$Q(s,T) := \mathbb{E}\left[\left(\overline{Z}_T - \mathbb{E}\left[\overline{Z}_T \middle| (X_u)_{u\in[0,s]}\right]\right)\left(\overline{Z}_T - \mathbb{E}\left[\overline{Z}_T \middle| (X_u)_{u\in[0,s]}\right]\right)^* \middle| (X_u)_{u\in[0,s]}\right].$$

Si X est une martingale, on a $Q(s,T) = \left(\left(\int_s^T f_i(u)f_j(u)d\langle X\rangle_u\right)\right)_{(i,j)\in I^2}$. Rappelons qu'une semimartingale continue X est gaussienne si et seulement si $\langle X\rangle$ est déterministe (voir par exemple [17]). Donc, cette hypothèse supplémentaire équivaut à supposer que

$$Q(s,T) \quad \text{est inversible pour tout } s \in [0,T). \tag{$\mathcal{H}$}$$

Le théorème suivant résulte d'une approche similaire à celle développée dans l'article [1] pour le cas du mouvement brownien, qui est ici étendue au cas plus général d'une semimartingale gaussienne continue centrée partant de 0. Les démonstrations font appel à des outils de grossissement de filtration et les preuves sont détaillées au chapitre 3.

Théorème 0.3.1. *Sous l'hypothèse $(\mathcal{H})$, pour tout $s \in [0,T)$, et pour $\mathbb{P}_{\overline{Z}_T}$-presque sûrement $\overline{y} \in \mathbb{R}^I$, $\mathbb{P}\left[\cdot \middle| \overline{Z}_T = \overline{y}\right]$ est équivalente à $\mathbb{P}$ sur $\mathcal{F}_s^X$ et sa dérivée de Radon-Nikodym est donnée par*

$$\frac{d\mathbb{P}\left[\cdot \middle| \overline{Z}_T = \overline{y}\right]}{d\mathbb{P}}_{|\mathcal{F}_s^X} = \frac{\Pi_{(X_u)_{u\in[0,s]},T}(\overline{y})}{\Pi_{0,T}(\overline{y})}.$$

Proposition 0.3.2 (Les ponts généralisés comme semimartingales)**.** *Définissons la filtration $\mathcal{G}^X$ par $\mathcal{G}_t^X = \sigma\left(\overline{Z}_T, \mathcal{F}_t^X\right)$, le grossissement de $\mathcal{F}^X$ correspondant au conditionnement précédent. On considère le processus stochastique $D_s^{\overline{y}} := \frac{d\mathbb{P}\left[\cdot \middle| \overline{Z}_T\right]}{d\mathbb{P}}_{|\mathcal{F}_s^X} = \frac{\Pi_{(X_t)_{t\in[0,s]},T}(\overline{y})}{\Pi_{0,T}(\overline{y})}$ pour $s \in [0,T)$.*

Sous l'hypothèse $(\mathcal{H})$, si $D^{\overline{y}}$ est continue, alors X est une $\mathcal{G}^X$-semimartingale continue sur $[0,T)$.

Remarque (Modification continue)**.** *Dans la proposition 0.3.2, si on suppose seulement que $D^{\overline{y}}$ a une modification continue $\mathcal{D}^{\overline{y}}$, alors à chacune de ses modifications continues est associée une $\mathcal{G}^X$-semimartingale continue sur $[0,T)$ et toutes ces semimartingales sont des modifications les unes des autres.*

Proposition 0.3.3 (Continuité de $D^{\overline{y}}$)**.** *Si $\mathcal{F}^X$ est une filtration brownienne standard, alors $D^{\overline{y}}$ a une modification continue.*

On prouve dans le chapitre 3 que l'hypothèse $(\mathcal{H})$ est bien vérifiée dans le cas des ponts généralisés de Karhunen-Loève du mouvement brownien standard, du pont brownien et des processus d'Ornstein-Uhlenbeck. Le cas des processus d'Ornstein-Uhlenbeck fait appel à quelques compléments sur les propriétés d'injectivité de l'intégrale de Wiener, développés en annexe.

- La première et principale conséquence de ce nouveau résultat est qu'en fait, la loi conditionnelle d'une semimartingale gaussienne dans une cellule de Voronoi de son quantifieur optimal (ou quantifieur produit optimal) reste une semimartingale (non gaussienne). Cette propriété nous permet de définir précisément la stratification fonctionnelle de la solution d'une équation différentielle stochastique, comme nous l'avons mentionné dans la présentation du chapitre 1.

- De plus, ces résultats suggèrent une nouvelle approche pour la quantification fonctionnelle de solutions d'équations différentielles stochastiques, la *quantification partielle*.

0.3.3 Sur la quantification partielle

Sous les mêmes hypothèses, considérons le développement de Karhunen-Loève de X sur $[0,T]$,

$$X = \sum_{i\in I} Y_i e_i^X \overset{\perp\!\!\!\perp}{+} \sum_{i\in\mathbb{N}^*\setminus I} \sqrt{\lambda_i^X}\xi_i e_i^X, \tag{8}$$

où $(Y_i)_{i\in I} \overset{\mathcal{L}}{\sim} \mathcal{N}(0, \mathrm{diag}(\lambda_i)_{i\in I})$. Soit maintenant $\widehat{Y}^\Gamma$ un quantifieur stationnaire de taille N de Y. $\widehat{Y}^\Gamma$ peut s'écrire comme la projection au plus proche voisin sur un nuage $\Gamma = (\gamma_1, \cdots, \gamma_N)$.

$$\widehat{Y}^\Gamma = \mathrm{Proj}_\Gamma(Y), \qquad \text{où } \mathrm{Proj}_\Gamma \text{ est une projection au plus proche voisin sur } \Gamma.$$

On définit maintenant le processus $\widetilde{X}^{I,\Gamma}$ en remplaçant Y par $\widehat{Y}^\Gamma$ dans la décomposition (8),

$$\widetilde{X}^{I,\Gamma} = \sum_{i\in I} \widehat{Y}_i^\Gamma e_i^X \overset{\perp\!\!\!\perp}{+} \sum_{i\in\mathbb{N}^*\setminus I} \sqrt{\lambda_i^X}\xi_i e_i^X.$$

La loi conditionnelle de $\widetilde{X}^{I,\Gamma}$ sachant que Y tombe dans la cellule de Voronoi de γ_k est la loi d'un pont généralisé de Karhunen-Loève de point final γ_k. En d'autres termes, on a quantifié les coordonnées de Karhunen-Loève de X correspondant à $i \in I$ et pas les autres. Ce processus ainsi défini $\widetilde{X}^{I,\Gamma}$ est appelée quantification partielle de X.

Considérons l'équation différentielle stochastique

$$dS_t = b(t,S_t)dt + \sigma(t,S_t)dX_t, \qquad S_0 = x \in \mathbb{R}, \quad t \in [0,T], \tag{9}$$

où $b(t,x)$ et $\sigma(t,x)$ sont des fonctions boréliennes, lipschitziennes par rapport à x, ce uniformément en t et où σ et $b(\cdot,0)$ sont bornées. Cette équation différentielle stochastique admet une unique solution forte S. La loi conditionnelle de S sachant que $Y_i = y_i$ pour $i \in I$ est celle de la solution forte de l'équation différentielle stochastique $dS_t = b(t,S_t)dt + \sigma(t,S_t)dX_t^{I,\overline{y}}$, avec $S_0 = x$ et où $X_t^{I,\overline{y}}$ est le pont généralisé de Karhunen-Loève associé.

En conséquence, on définit la quantification partielle de S à partir de la quantification partielle $\widetilde{X}^{I,\Gamma}$ de X en remplaçant X par $\widetilde{X}^{I,\Gamma}$ dans l'équation différentielle stochastique (9). La *quantification partielle* $\widetilde{S}^{I,\Gamma}$ de S est le processus dont la loi conditionnelle sachant que Y tombe dans la cellule de Voronoi de γ_k est la solution forte de la même équation différentielle stochastique où X est remplacé par le pont généralisé de Karhunen-Loève de point final γ_k,

$$d\widetilde{S}_t^{I,\Gamma} = b\left(t, \widetilde{S}_t^{I,\Gamma}\right) dt + \sigma\left(t, \widetilde{S}_t^{I,\Gamma}\right) d\widetilde{X}_t^{I,\Gamma}.$$

Le chapitre se termine par deux résultats de convergence (L^p et presque sûre) de ce schéma de quantification partielle vers la solution de l'EDS (9).

0.4 Principaux résultats du chapitre 4

0.4.1 Quantification fonctionnelle de solutions d'équations différentielles stochastiques

Une application de la quantification fonctionnelle de processus gaussiens X sur un intervalle $[0,T]$ est la quantification d'une équation différentielle stochastique dirigée par X, dès qu'on peut définir

l'intégrale stochastique correspondante. Dans le cas présent, on supposera que X est une semi-martingale gaussienne continue centrée partant de 0. Des exemples typiques de tels processus sont le mouvement brownien standard, le pont brownien et les processus d'Ornstein-Uhlenbeck centrés partant de 0. Comme cela a déjà été mentionné précédemment, le théorème de Fernique garantit que $\int_0^T \mathbb{E}\left[X_t^2\right] dt < +\infty$. De plus, la continuité trajectorielle du processus X implique la continuité de sa fonction de covariance de Γ^X sur $[0,T]^2$. On peut obtenir un quantifieur stationnaire de la solution d'une EDS en remplaçant X par un quantifieur stationnaire $\widehat{X}$ dans l'EDS écrite au sens de Stratonovich. Une première étude de cette question a été faite dans le cas unidimensionnel dans l'article [10]. Le cas de diffusions multidimensionnelles plus générales est traité dans l'article [16], en utilisant des techniques issues de la théorie des trajectoires rugueuses.

Formellement, considérons σ le processus stochastique défini comme la solution forte de l'équation différentielle stochastique

$$d\sigma_t = b(t,\sigma_t)dt + \theta(t,\sigma_t)dX_t, \quad \sigma_0 \in \mathbb{R}, \tag{10}$$

où $b(t,x)$ et $\theta(t,x)$ sont des fonctions boréliennes, lipschitziennes par rapport à x uniformément en t et $|b(\cdot,0)| + |\theta(\cdot,0)|$ est bornée sur $[0,T]$. Sous ces conditions, il existe une unique solution forte de l'EDS (10) sur l'intervalle $[0,T]$. On rappelle que si M et H sont des semimartingales continues, l'intégrale de Stratonovich $H \circ M$ est définie par $H \circ M := H \cdot M + \frac{1}{2}\langle H, M\rangle$, où $H \cdot M$ désigne l'intégrale d'Itô de H par rapport à M. Si on suppose que $\theta(t,x)$ est dérivable par rapport à x, on peut réécrire l'équation différentielle stochastique (10) en termes d'intégrale de Stratonovich $d\sigma_t = b(t,\sigma_t)dt - \frac{1}{2}d\langle\theta(\cdot,\sigma), X\rangle_t + \theta(t,\sigma_t) \circ dX_t, \quad \sigma_0 \in \mathbb{R}$. En utilisant que $d\langle\theta(\cdot,\sigma), X\rangle_t = \theta'_x(t,\sigma_t)\theta(t,\sigma_t)d\langle X\rangle_t$, on obtient

$$d\sigma_t = b(t,\sigma_t)dt - \frac{1}{2}\theta'_x(t,\sigma_t)\theta(t,\sigma_t)d\langle X\rangle_t + \theta(t,\sigma_t) \circ dX_t.$$

Rappelons qu'une semimartingale continue centrée est gaussienne si et seulement si $\langle X\rangle$ est une fonction déterministe du temps, voir par exemple [17]. La variation quadratique $\langle X\rangle$ est explicite dans les cas précédemment cités du mouvement brownien standard, du pont brownien et des processus d'Ornstein-Uhlenbeck.

Dans cette équation, on remplace X par un quantifieur stationnaire de X. Ce faisant, on obtient un ensemble d'équations différentielles ordinaires définissant un quantifieur stationnaire de σ. Soit donc $\chi := (\chi^i)_{1\le i\le N}$ les trajectoires d'un quantifieur stationnaire de X. Les trajectoires $(\widehat{\sigma}^i)_{1\le i\le N}$ du quantifieur $\widehat{\sigma}$ sont les solutions des équations différentielles ordinaires

$$d\widehat{\sigma}_t^i = b(t,\widehat{\sigma}_t^i)dt - \frac{1}{2}\theta'_x(t,\widehat{\sigma}_t^i)\theta(t,\widehat{\sigma}_t^i)d\langle X\rangle_t + \theta(t,\widehat{\sigma}_t^i)\left(\chi^i\right)'(t)dt, \quad \widehat{\sigma}_0^i = \sigma_0 > 0. \tag{11}$$

Dans certains cas particuliers, ces équations différentielles peuvent avoir des solutions explicites, comme dans le cas lognormal. Si on considère le cas où $b(t,x) = x\mu(t)$ et $\theta(t,x) = x\gamma(t)$, l'équation (11) devient

$$d\widehat{\sigma}_t^i = \widehat{\sigma}_t^i\mu(t)dt - \widehat{\sigma}_t^i\frac{\gamma(t)^2}{2}d\langle X\rangle_t + \widehat{\sigma}_t^i\gamma(t)\left(\chi^i\right)'(t)dt, \quad \widehat{\sigma}_0^i = \sigma_0 > 0,$$

ce qui donne

$$\widehat{\sigma}_t^i = \sigma_0 \exp\left(\int_0^t \mu(s)ds + \int_0^t \gamma(s)\left(\chi^i\right)'(s)ds - \frac{1}{2}\int_0^t \gamma^2(s)d\langle X\rangle_s\right). \tag{12}$$

Dans le cas général, on peut utiliser des méthodes numériques de résolutions d'équations différentielles comme les méthodes de Runge-Kutta, ou le schéma de Bulirsh-Stoer, qui est particulièrement adapté au cas de solutions d'équations différentielles très régulières.
Sur la figure 4, nous représentons un quantifieur produit du processus défini par l'équation (12) quand X est un processus d'Ornstein-Uhlenbeck sur $[0,3]$ issu de 0 avec des paramètres de retour à la moyenne et volatilité tous deux égaux à 1, avec $\gamma \equiv 1$, $\mu \equiv 0$ et $\sigma_0 = 100$ et où χ est un $5 \times 2 \times 2$-quantifieur produit de X.

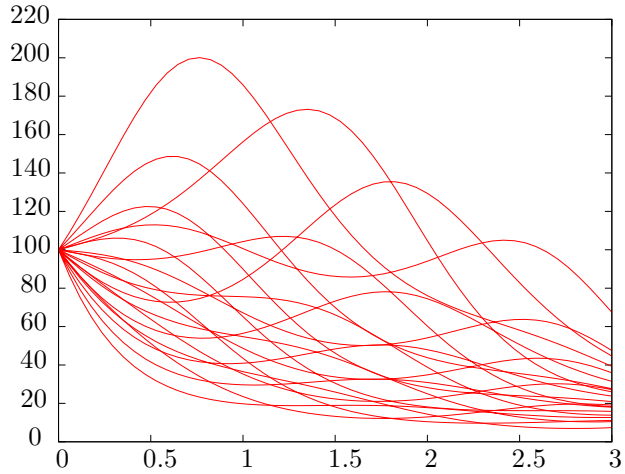

FIGURE 4 – Quantifieur fonctionnel quadratique produit $5 \times 2 \times 2$ de la solution de l'EDS (11) sur $[0, 3]$ quand X est un processus d'Ornstein-Uhlenbeck partant de 0 avec paramètres de retour à la moyenne et volatilité de 1. Les paramètres de l'équation différentielle stochastique sont $\gamma \equiv 1$, $\mu \equiv 0$ et $\sigma_0 = 100$.

0.4.2 Application à la valorisation d'options vanilles dans les modèles à volatilité stochastique

Considérons maintenant un modèle de volatilité stochastique sous une probabilité risque neutre de la forme

$$\begin{cases} dF_t = F_t \sigma_t dW_t, \quad F_0 > 0, \\ d\sigma_t = b(t, \sigma_t)dt + \theta(t, \sigma_t)dW_t^\sigma, \quad \sigma_0 > 0, \\ d\langle W, W^\sigma \rangle_t = \rho dt. \end{cases} \tag{13}$$

Le mouvement brownien W se décompose en la somme de W^σ et d'un mouvement brownien standard W^F indépendant de W^σ.

$$dW_s = \rho dW_s^\sigma + \sqrt{1-\rho^2} dW_s^F \qquad \text{où } W^\sigma \perp\!\!\!\perp W^F.$$

On note $(\mathcal{F}_t^\sigma)_{t\geq 0}$ et $(\mathcal{F}_t^F)_{t\geq 0}$ les filtrations naturelles des mouvements browniens W^σ et W^F. La solution de l'EDS (13), $F_t = F_0 \exp\left(\int_0^t \sigma_s dW_s - \frac{1}{2}\int_0^t \sigma_s^2 ds\right)$ peut s'écrire sous la forme d'un produit

$$F_t = \underbrace{F_0 \exp\left(\rho \int_0^t \sigma_s dW_s^\sigma - \frac{\rho^2}{2}\int_0^t \sigma_s^2 ds\right)}_{:=A_t} \underbrace{\exp\left(\sqrt{1-\rho^2}\int_0^t \sigma_s dW_s^F - \frac{1-\rho^2}{2}\int_0^t \sigma_s^2 ds\right)}_{:=B_t}, \tag{14}$$

et le processus $(A_t)_{t\in[0,T]}$ ainsi défini est adapté à la filtration $\mathcal{F}^\sigma$. Dans la suite, la fonction Payoff(x,K) désigne ou bien la fonction $(x-K)_+$ ou bien $(K-x)_+$, le Payoff d'un Call ou d'un Put de prix d'exercice K. Un préconditionnement donne l'expression

$$\begin{aligned} \mathbb{E}\left[\text{Payoff}(F_T, K)\right] &= \mathbb{E}\left[\mathbb{E}\left[\text{Payoff}(F_T, K)|\mathcal{F}_T^\sigma\right]\right] \\ &= \mathbb{E}\left[\text{PrimeBS}\left(A_T, \left((1-\rho^2)\int_0^T \sigma_s^2 ds\right)^{\frac{1}{2}}, T, K\right)\right], \end{aligned}$$

où A_T est la valeur finale du processus défini dans l'équation (14) et où $\text{PrimeBS}(F, \sigma, T, K)$ est la formule fermée pour le prix d'un Call ou d'un Put dans le modèle de Black et Scholes, sans taux d'intérêt ni dividende, avec un Forward F, une volatilité σ, une maturité T et un prix d'exercice K.

À ce stade, on est donc confronté à un problème de cubature par rapport à la distribution du processus de volatilité σ. Cette cubature est effectuée en utilisant le quantifieur fonctionnel $(\widehat{\sigma}^i)_{1\leq i\leq N}$ de σ.

$$\mathbb{E}[\text{Payoff}(F_T,K)] \approx \sum_{i=1}^{N} p_i \text{PrimeBS}\Big(A_T^i, \big((1-\rho^2)\int_0^T \widehat{\sigma}^i(s)^2 ds\big)^{\frac{1}{2}}, T, K\Big), \tag{15}$$

où $(A_T^i)_{1\leq i\leq N}$ désigne le quantifieur de A_T déduit de $(\widehat{\sigma}^i)_{1\leq i\leq N}$.

- Dans cette équation, $(\alpha_i)_{1\leq i\leq N}$ et $(p_i)_{1\leq i\leq N}$ sont respectivement les trajectoires d'un quantifieur fonctionnel de W^σ et les poids associés. Les fonctions $(\widehat{\sigma}^i)_{1\leq i\leq N}$ sont les trajectoires du quantifieur de σ obtenues à partir de $(\alpha_i)_{1\leq i\leq N}$ en résolvant les EDO (11).

- Les valeurs correspondantes de $\left(\int_0^T \widehat{\sigma}^i(s)^2 ds\right)_{1\leq i\leq N}$ utilisées dans la formule (15) sont déduites de cette quantification.

- Pour calculer les termes A_T^i, on doit évaluer la version quantifiée de l'intégrale stochastique $\int_0^T \sigma_s dW_s^\sigma = \int_0^T \sigma_s \circ dW_s^\sigma - \frac{1}{2}\int_0^T d\langle\sigma, W^\sigma\rangle_t = \int_0^T \sigma_s \circ dW_s^\sigma - \frac{1}{2}\int_0^T \theta^2(t,\sigma_t)dt$. Cela conduit au quantifieur

$$A_T^i = F_0 \exp\left(\rho\int_0^T \widehat{\sigma}^i(t)\alpha_i'(t)dt - \frac{\rho}{2}\int_0^T \theta^2(t,\widehat{\sigma}_t^i)dt - \frac{\rho^2}{2}\int_0^T \widehat{\sigma}^i(t)^2 dt\right), \quad 1\leq i\leq N.$$

Dans le chapitre 4 on rappelle que l'erreur de cubature par quantification fonctionnelle stationnaire décroît logarithmiquement vers 0, ce qui est très lent. Cependant, on peut considérablement améliorer les performances pratiques de la méthode en utilisant une extrapolation de Richardson-Romberg de l'erreur de cubature. Nous détaillons ces questions dans la section 4.2.2.

Les résultats numériques peuvent encore être améliorés en utilisant une sorte de méthode de réduction de variance pour la cubature par quantification, détaillée à la section 4.3.1. L'idée principale est d'utiliser le Forward estimé par cubature au lieu du Forward théorique pour le calcul de volatilité implicite. Ce faisant, on obtient un Smile de volatilité plus régulier, et plus proche de sa valeur théorique quand on teste la méthode dans le cas particulier du modèle SABR.

La figure 5 représente le Smile de volatilité implicite estimé par cubature par quantification (et extrapolation) et la valeur de référence donnée par la formule de Hagan de développement asymptotique de faible maturité. On constate que la précision obtenue est suffisante pour une utilisation en pratique de notre méthode.

0.4.3 La quantification normale

Dans la suite du chapitre 4, nous proposons un nouveau schéma de quantification de solutions d'équations différentielles stochastiques, basé sur la notion de quantification partielle introduite au chapitre 3. Ce schéma permet d'approcher la solution d'une équation différentielle stochastique par une mixture de processus gaussiens. Pour cette raison, nous appelons cette méthode d'approximation la « quantification normale ».

Définition 0.4.1 (Quantification normale)**.** *Soit X une martingale continue gaussienne centrée partant de 0 sur $[0,T]$. Soit I une partie finie de $\mathbb{N}^*$. Considérons la décomposition*

$$X = \sum_{i\in I} Y_i e_i^X \overset{\perp}{+} \sum_{i\in\mathbb{N}^*\setminus I} \sqrt{\lambda_i^X}\xi_i e_i^X.$$

Soit Γ un quantifieur stationnaire de $Y=(Y_i)_{i\in I}$ et $\widehat{Y}^\Gamma$ la projection de Y correspondante. On note $\widehat{X}^{I,\Gamma}$ et $\widetilde{X}^{I,\Gamma}$ les quantifications fonctionnelles et partielles de X correspondant à Γ et I.

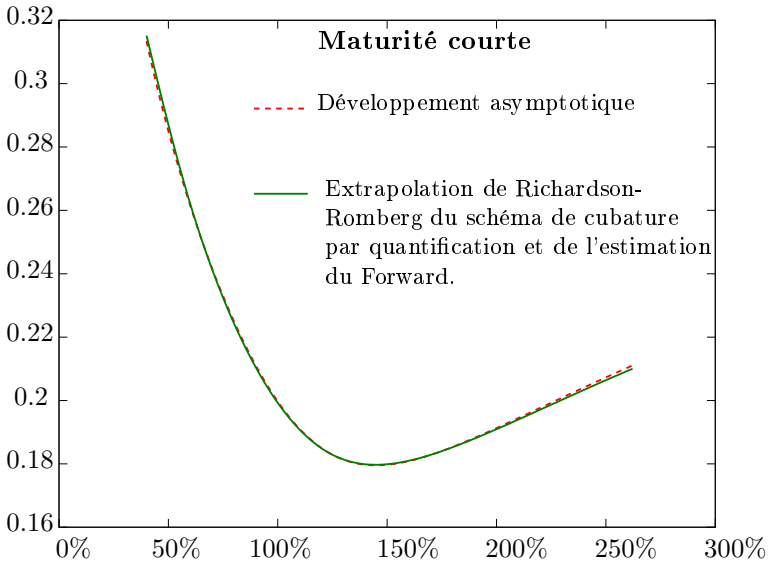

FIGURE 5 – Smile de volatilité implicite dans le modèle SABR avec $\beta = 1$, $\gamma = 0.3$, $\sigma_0 = 0.2$, $T = 1$ et $\rho = -0.5$. La courbe continue correspond à une extrapolation de Richardson-Romberg pour le couple (208-54) de la formule de cubature basée sur la quantification fonctionnelle.

Soit S la solution forte de l'EDS

$$dS_t = b(t, S_t)dt + \sigma(t, S_t)dX_t, \quad S_0 = x, \tag{16}$$

où $b(t,x)$ et $\sigma(t,x)$ sont des fonctions boréliennes, lipschitziennes par rapport à x uniformément en t et où σ et $|b(\cdot, 0)|$ sont bornées. Alors on note

- *$\widehat{S}^{I,\Gamma}$ la quantification fonctionnelle de S, obtenue en remplaçant X par $\widehat{X}^{I,\Gamma}$ dans l'EDS (16) écrite au sens de Stratonovich.*
- *$\widetilde{S}^{I,\Gamma}$ la quantification partielle de S correspondante, obtenue en remplaçant X par $\widetilde{X}^{I,\Gamma}$ dans l'EDS (16) comme au chapitre 3.*

Le processus $\widetilde{X}_t^{I,\Gamma}$ se décompose en $\widetilde{X}_t^{I,\Gamma} = \widehat{X}_t^{I,\Gamma} \overset{\perp}{+} X^{I,\overline{0}}$, où $X^{I,\overline{0}}$ est le pont généralisé de Karhunen-Loève de point final $\overline{0}$ associé à I. On obtient

$$d\widetilde{S}_t^{I,\Gamma} = b\left(t, \widetilde{S}_t^{I,\Gamma}\right) dt + \sigma\left(t, \widetilde{S}_t^{I,\Gamma}\right) d\widehat{X}_t^{I,\Gamma} + \sigma\left(t, \widetilde{S}_t^{I,\Gamma}\right) dX_t^{I,\overline{0}}, \quad \widetilde{S}_0^{I,\Gamma} = x. \tag{17}$$

Nous définissons la quantification normale de cette EDS comme la solution $\widehat{S}^{I,\Gamma}$ de l'EDS

$$d\widehat{S}_t^{I,\Gamma} = b\left(t, \widehat{S}_t^{I,\Gamma}\right) dt + \sigma\left(t, \widehat{S}_t^{I,\Gamma}\right) d\widehat{X}_t^{I,\Gamma} + \sigma\left(t, \widehat{S}_t^{I,\Gamma}\right) dX_t^{I,\overline{0}}, \quad \widetilde{S}_0^{I,\Gamma} = x.$$

C'est la même EDS que (17), dans laquelle $\widetilde{S}_t^{I,\Gamma}$ est remplacé par $\widehat{S}_t^{I,\Gamma}$ dans les termes de volatilité et de dérive.

Théorème 0.4.1 (Erreur de quantification quadratique de la quantification normale d'EDS)**.** *Avec les mêmes notations et sous les hypothèses de la définition 4.4.4, pour tout $t \in [0, T)$ nous avons*

$$\mathbb{E}\left[\sup_{u\in[0,t]} \left|\widetilde{S}_u - \widehat{S}_u\right|^2 \middle| \widehat{Y}^\Gamma = \gamma_k\right] = O\left(\left\|\widetilde{S}^{I,\Gamma} - \widehat{S}^{I,\Gamma}\right\|_2^2\right). \tag{18}$$

Remarque. *Ces résultats peuvent être aisément étendus au cas plus général de semimartingales gaussiennes, dès qu'il existe une mesure localement finie ν sur $[0,T]$ telle que pour presque tout $\omega \in \Omega$, le terme de variation finie $dV(\omega)$ dans la décomposition canonique de X soit absolument continu par rapport à ν.*

À partir de la quantification normale d'une EDS, on peut facilement définir une notion de quantification lognormale d'EDS en considérant la quantification normale de l'exponentielle de la solution de l'EDS considérée (voir définition 4.4.5), et obtenir le même type de contrôle d'erreur. Comme c'est le cas pour la quantification fonctionnelle ordinaire, nous déduisons aisément des méthodes de cubature associées à ces quantifications normale et lognormale, consistant dans le cas d'options vanilles à approcher le prix d'une option dans un modèle à volatilité et dérive locale par une somme pondérée de prix dans des modèles respectivement normaux et lognormaux. Nous faisons ensuite la conjecture (appuyée par nos résultats numériques) que l'erreur de cubature de la quantification normale se comporte comme l'erreur de cubature de la quantification fonctionnelle, en d'autres termes,

- la stationnarité du quantifieur fonctionnel associé permet de gagner un ordre de convergence pour l'erreur de cubature associée,
- l'erreur de cubature associée admet un développement asymptotique par rapport à la taille du quantifieur N de la forme $K\mathcal{E}_N^2 + o(\mathcal{E}_N^2)$ quand N tend vers $+\infty$, où $\mathcal{E}_N$ est la distorsion quadratique de la quantification fonctionnelle associée.

Ainsi, nous avons une méthode de cubature associée à la distribution d'équations différentielles stochastiques de la forme

$$dS_t = b(t,S_t)dt + \sigma(t,S_t)dX_t,$$

i.e. comportant un terme de volatilité locale et un terme de dérive locale.

0.4.4 Application à la valorisation dans les modèles à volatilité locale stochastique

On se place maintenant dans un cadre plus général. On suppose que sous une probabilité risque-neutre, le Forward a pour dynamique

$$\begin{cases} dF_t = \sigma_t F_t g(t,F_t)dW_t, \\ d\sigma_t = b(t,\sigma_t)dt + \theta(t,\sigma_t)dW_t^\sigma, \end{cases} \tag{19}$$

où W et W^σ sont des mouvements browniens standards. On suppose que W se décompose en $\rho dW_t^\sigma + \sqrt{1-\rho^2}dW_t^F$, où W^F est indépendant W^σ. On note respectivement $\mathcal{F}^F$ et $\mathcal{F}^\sigma$ les filtrations naturelles des mouvements browniens W^σ et W^F. De plus, on suppose que $b(t,x)$ et $\theta(t,x)$ sont boréliennes, lipschitziennes par rapport à x uniformément en $t \in [0,T]$ et que $\theta(t,\cdot)$ est de classe C^1 pour tout $t \in [0,T]$. On suppose que $g(t,x)$ est une fonction borélienne bornée et que $g(t,\cdot)$ est de classe C^1.

Cette situation correspond à de nombreux modèles à volatilité classique, comme SABR et Heston. Nous détaillons les motivations de l'introduction de ce genre de modèles, impliquant un terme de volatilité locale en plus de la volatilité stochastique dans l'introduction du chapitre 4. Nous écrivons l'EDS (19) en termes d'intégrale de Stratonovich, ce qui donne

$$\begin{cases} dF_t &= \sigma_t g(t,F_t)F_t\sqrt{1-\rho^2}dW_t^F + \sigma_t g(t,F_t)F_t\rho \circ dW_t^\sigma \\ &\quad -\frac{\rho^2}{2}\sigma_t^2 g(t,F_t)g_x'(t,F_t)F_t^2dt - \frac{\rho^2}{2}\sigma_t^2 g(t,F_t)^2F_tdt - \frac{\rho}{2}F_t g(t,F_t)\theta(t,\sigma_t)dt. \end{cases} \tag{20}$$

Considérons maintenant un quantifieur produit stationnaire α de W^σ, correspondant à la décomposition $N_1 \times \cdots \times N_n$. La trajectoire de α correspondant au multi-indice $\underline{i} := \{i_1, \cdots, i_n, \cdots\}$ est de la forme

$$\alpha^{\underline{i}} = \sum_{n\geq 1} \sqrt{\lambda_n^W} x_{i_n}^{N_n} e_n^W.$$

La trajectoire de σ correspondant au multi-indice $\underline{i}$ est définie comme la solution de l'EDO obtenue en remplaçant W^σ par $\alpha^{\underline{i}}$ dans l'équation différentielle stochastique écrite en termes d'intégrale de Stratonovich (20). On remplace le mouvement brownien W^σ par son quantifieur $\alpha^{\underline{i}}$ et σ par $\sigma^{\underline{i}}$ dans l'équation (20). On obtient

$$\begin{aligned} dF_t^{\underline{i}} &= \sigma_t^{\underline{i}} g\left(t, F_t^{\underline{i}}\right) F_t^{\underline{i}} \sqrt{1-\rho^2} dW_t^F \\ &\quad \underbrace{+ \sigma_t^{\underline{i}} g\left(t, F_t^{\underline{i}}\right) F_t^{\underline{i}} \rho \left(\alpha^{\underline{i}}\right)'(t) dt - \frac{\left(\rho \sigma_t^{\underline{i}}\right)^2}{2} g\left(t, F_t^{\underline{i}}\right) g_x'\left(t, F_t^{\underline{i}}\right) \left(F_t^{\underline{i}}\right)^2 dt \\ &\qquad - \frac{\left(\rho^2 \sigma_t^{\underline{i}}\right)^2}{2} g\left(t, F_t^{\underline{i}}\right)^2 F_t^{\underline{i}} dt - \frac{\rho}{2} F_t^{\underline{i}} g\left(t, F_t^{\underline{i}}\right) \theta\left(t, \sigma_t^{\underline{i}}\right) dt}_{:= F_t^{\underline{i}} \mu_{\underline{i}}\left(t, F_t^{\underline{i}}\right) dt} \\ &= F_t^{\underline{i}} \mu_{\underline{i}}\left(t, F_t^{\underline{i}}\right) dt + \sigma_t^{\underline{i}} g\left(t, F_t^{\underline{i}}\right) F_t^{\underline{i}} \sqrt{1-\rho^2} dW_t^F. \end{aligned} \tag{21}$$

En d'autres termes, $F_t^{\underline{i}}$ a une volatilité locale et une dérive locale. En préconditionnant par le mouvement brownien W^σ, on obtient comme dans le cas précédent

$$\mathbb{E}\left[(F_t - K)_+\right] = \mathbb{E}\left[\underbrace{\mathbb{E}\left[(F_t - K)_+ | \mathcal{F}_T^\sigma\right]}_{=\phi((W^\sigma)_{t\in[0,T]})}\right]. \tag{22}$$

Cette espérance est ensuite calculée par la méthode de cubature issue de la quantification fonctionnelle :

$$\mathbb{E}[(F_t - K)_+] \approx \sum_{i \in I} p_i \phi_i.$$

La fonction ϕ apparaissant dans l'équation (22) correspond à la valorisation d'un Call ou d'un Put dans un modèle à volatilité locale et dérive locale, comme dans l'équation (21). En conséquence et contrairement au cas où nous n'avions pas de terme de volatilité locale, nous ne disposons pas de formule fermée. Nous proposons donc d'utiliser la méthode de cubature par quantification normale pour traiter ce problème. Nous nous reportons au chapitre 4 pour les résultats obtenus par cette méthode.

0.5 Principaux résultats du chapitre 5

Considérons $\Gamma = \{\gamma_1, \cdots, \gamma_N\} \subset \mathbb{R}^d$ un ensemble de N points de $\mathbb{R}^d$. Le problème de la recherche rapide de plus proche voisin consiste à déterminer pour un nouveau point $x \in \mathbb{R}^d$ quel est l'élément de Γ le plus proche de x.

Quand on doit effectuer un très grand nombre de recherches dans le même ensemble Γ, un prétraitement sur Γ sera profitable s'il permet de créer une structure de données rendant le temps de recherche moyen plus court. Ce problème a été résolu de manière quasi-optimale dans le cas des petites dimensions. La plupart des algorithmes ont une complexité asymptotique logarithmique en le nombre de points et un temps de prétraitement de $O(n \log n)$. En pratique, les méthodes diffèrent par leur efficacité selon les ensembles de points auxquels ils sont confrontés et la distribution des points dont on cherche le plus proche voisin.

Le problème de la recherche rapide de plus proche voisin est particulièrement critique dans le cas de la quantification vectorielle, d'une part parce que la plupart des algorithmes d'optimisation de grilles de quantification, comme la méthode de Lloyd, nécessitent justement de faire un grand nombre de recherches dans le quantifieur à optimiser, d'autre part parce que le calcul par la méthode de Monte-Carlo des poids et distorsions locales nécessite lui aussi un grand nombre de recherches. Enfin, c'est un problème critique pour les méthodes de réduction de variance utilisant un quantifieur Voronoi comme variable de contrôle, comme cela a été signalé dans l'article [8].

Beaucoup d'algorithmes de recherche de plus proche voisin reposent sur un partitionnement récursif de l'ensemble Γ résultant en une structure de recherche par arbre. La méthode la plus

populaire est l'algorithme Kd-tree [4]. Cet algorithme a été améliorée dans l'article [11] par McNames qui tire avantage de la géométrie de l'ensemble Γ en utilisant une analyse en composantes principales de l'ensemble de points considéré.

Dans le chapitre 5, on propose un nouvel algorithme de recherche de plus proche voisin par arbre de recherche. Comme les deux méthodes citées précédemment, l'ensemble Γ est récursivement partitionné jusqu'à ce qu'on ait un nombre suffisamment petit de points dans chaque nœud terminal de l'arbre. L'algorithme proposé tire profit de la géométrie du nuage de points en utilisant des méthodes de partitionnement par quantification. En effet, la quantification vectorielle quadratique est le pendant continu de critères de classification automatique (et l'algorithme K-means consiste simplement à appliquer l'algorithme de Lloyd à une distribution empirique).

Les tests effectués dans le chapitre 5 sur cette nouvelle décomposition du nuage de points montrent qu'elle serait plus adaptée à la recherche de plus proches voisins que la décomposition issue de l'arbre d'axe principal proposé dans [11]. Nous proposons donc un algorithme basé sur cette nouvelle décomposition spatiale, similaire aux méthodes usuelles de recherche par arbre. Les premiers tests effectués avec cette méthode montrent que s'il semble mieux résister à l'augmentation de la dimension que les autres méthodes citées précédemment, il se comporte moins bien que l'arbre d'axe principal en dimensions 2, 3 et 4. Nous proposons alors une optimisation de l'algorithme permettant d'améliorer ses performances en petites dimensions. Cette optimisation nécessite cependant un exposé plus approfondi de notions relatives aux maillages de Delaunay et diagrammes de Voronoi.

0.A Principaux résultats de l'appendice A

Si X est une variable aléatoire L^2 sur $\mathbb{R}^d$ et $\Gamma = \{\gamma_1, \cdots, \gamma_N\}$ est un ensemble de N points distincts de $\mathbb{R}^d$, la meilleure façon d'approcher X par une variable aléatoire prenant ses valeurs sur Γ est d'utiliser une projection au plus proche voisin de X sur Γ, $\widehat{X}^\Gamma = \mathrm{Proj}_\Gamma(X)$. Le problème d'optimisation (1) se traduit donc en un problème de minimisation plus simple, portant sur l'ensemble Γ :

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

Nous nous référons à l'article [13] pour une revue des méthodes disponibles permettant de résoudre ce problème numériquement. Mentionnons par exemple l'algorithme CLVQ (Competitive Learning Vector Quantization) qui est une sorte de méthode de gradient stochastique appliquée à l'erreur de quantification, dont le gradient a une représentation intégrale. L'algorithme CLVQ converge presque sûrement vers un quantifieur optimal (ce résultat de convergence n'a en fait été démontré que dans le cas des distributions à support compact). Un autre algorithme couramment utilisé pour la quantification est l'algorithme de Lloyd, qui converge vers un quantifieur stationnaire non nécessairement optimal. La limite peut ne même pas être un minimum local de l'erreur de quantification. La convergence vers un quantifieur optimal n'est à ce jour garantie que dans le cas des distributions unidimensionnelles strictement log-concaves, comme les lois gaussiennes en dimension 1. Dans ce cas, nous avons de plus unicité du quantifieur optimal. L'algorithme de Lloyd, écrit de façon formelle, implique le calcul d'espérances, qui en pratique peuvent être évaluées par des méthodes de Monte-Carlo. Cependant, dans le cas unidimensionnel, quand la densité et la fonction de répartition sont connues, on dispose de formules fermées pour les espérances impliquées dans l'algorithme.

Dans l'appendice A, nous nous consacrons à l'étude des méthodes déterministes d'optimisation de grilles de quantification pour le cas gaussien unidimensionnel. Ces méthodes déterministes permettent d'obtenir des grilles très rapidement avec une précision difficilement atteignable avec des algorithmes stochastiques. Cela rend leur calcul « à la volée » possible pour les multiples applications utilisant les grilles de quantification unidimensionnelles, comme la quantification produit de processus gaussiens. De plus, sur le site web `www.quantize.maths-fi.com` [15], nous mettons à disposition une grande base de données de grilles de quantification gaussiennes. Pour produire ces grilles de référence, nous utilisons ces algorithmes stochastiques et une librairie de calcul en

précision arbitraire [2], qui nous ont permis de produire des grilles optimales unidimensionnelles avec une précision relative de 10^{-32} de la taille $N = 1$ à $N = 10^4$. En d'autres termes, les grilles unidimensionnelles proposées sur le site peuvent être considérées comme *exactes* pour les nombres flottants de simple, double et quadruple précision.
Dans le tableau ci-dessous, nous reportons les valeurs des points et des poids correspondant d'un quantifieur quadratique N-optimal de la loi gaussienne centrée réduite sur $\mathbb{R}$ avec $N = 9$. Ces valeurs numériques ont une précision relative de 10^{-32}.

Points	Poids
$-2.25466363591241546397232903003063 82$	$3.1053737504986977564788528825468893 \times 10^{-2}$
$-1.4763917385976070721619675715970733$	$8.4483855789973427268858217803130418 \times 10^{-2}$
$-0.91879588388282995755991264252455596$	$1.3232941900133077367386905238926516 \times 10^{-1}$
$-0.44363864762697592079813433780655785$	$1.6436025567507831709996174755021895 \times 10^{-1}$
0.0	$1.7554546405726100878504490686383316 \times 10^{-1}$
$0.44363864762697592079813433780655785$	$1.6436025567507831709996174755021895 \times 10^{-1}$
$0.91879588388282995755991264252455596$	$1.3232941900133077367386905238926516 \times 10^{-1}$
$1.4763917385976070721619675715970733$	$8.4483855789973427268858217803130418 \times 10^{-2}$
$2.25466363591241546397232903003063 82$	$3.1053737504986977564788528825468893 \times 10^{-2}$

Bibliographie

[1] Larbi Alili. Canonical decompositions of certain generalized Brownian bridges. *Electronic communications in probability*, 7 :27–35, 2002.

[2] David H. Bailey, Yozo Hida, Xiaoye S. Li, Brandon Thompson, Karthik Jeyabalan, and Alex Kaiser. The ARPREC libraries, 2010.

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

[4] Jon Louis Bentley. Multidimensional binary search trees used for associative searching. *Commun. ACM*, 18(9) :509–517, 1975.

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

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

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

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

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

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

[11] James McNames. A fast nearest-neighbor algorithm based on a principal axis search tree. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(9) :964–976, 2001.

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

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

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

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

[16] Gilles Pagès and Afef Sellami. Convergence of multi-dimensional quantized SDE's. In Catherine Donati-Martin, Antoine Lejay, and Alain Rouault, editors, *Séminaire de Probabilités XLIII*, pages 269–308. Springer, Berlin, 2010.

[17] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion.* Springer, 3rd edition, 2005.

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

[19] Benedikt Wilbertz. *Construction of optimal quantizers for Gaussian measures on Banach spaces.* PhD thesis, Universität Trier, 2008.

Chapter 1

Functional quantization-based stratified sampling methods

Abstract

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

Theoretical aspects of stratification lead to a strong link between the problem of optimal L^2-quantization of a random variable and the variance reduction that can be achieved. We first put the emphasis on the consistency of quantization for designing strata in stratified sampling methods in both finite-dimensional and infinite-dimensional frameworks. We show that this strata design has a uniform efficiency among the class of Lipschitz continuous functionals.

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

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

Joint work with Gilles Pagès.

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

Introduction

The quantization of a random variable X consists in its approximation by a random variable Y taking finitely many values. This problem has been initially investigated for its applications to signal transmission and for compression issues. (See [9].) In this context, quantization was a method of signal discretization. The point of interest was to design the random variable Y in order to minimize the resulting error for a fixed quantization level N. This led to the concept of optimal quantization.

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

Since the early 2000's, the infinite-dimensional setting has been extensively investigated from both theoretical and numerical viewpoints with a special attention paid to functional quantization [20, 25]. Bi-measurable stochastic processes are viewed as random variables taking values in their path spaces such as $L^2_T := L^2([0,T], dt)$.

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

Variance reduction methods can be used to reduce dramatically the computation time of a Monte-Carlo simulation, or to increase its accuracy. Main variance reduction methods are (adaptive) control variate, pre-conditioning, importance sampling and stratification [10, 19]. The problem is that these methods may strongly depend on the payoff or the model and imply specific changes in the practical implementation of the Monte-Carlo simulation. Thus, most institutions do not implement the most advanced methods in practice except for marginal cases.

In this chapter, we point out theoretical aspects of quantization that lead to a strong link between the problem of optimal L^2-quantization of a random variable and the variance reduction that can be achieved by stratification. We emphasize the consistency of quantization for designing strata in stratified sampling methods in both finite-dimensional and infinite-dimensional frameworks. Then we devise a stratified sampling algorithm based on product functional quantization for path-dependent functionals of multi-factor Brownian diffusions. We show that this strata design has a uniform efficiency among the class of Lipschitz continuous functionals of the Brownian motion. The simulation cost of the conditional path is $O(n)$ where n is the number of discretization dates, as for naive Monte-Carlo simulations. In this context, this stratification-based variance reduction method can be considered as a guided Monte-Carlo simulation. (See Figure 1.5.) The method extends to any Gaussian process as soon as its Karhunen-Loève decomposition is explicitly known. This is the case for the Brownian bridge or the Ornstein-Uhlenbeck process. The special case of the Ornstein-Uhlenbeck process is derived in Appendix 1.A.

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

The chapter is organized as follows. Section 1.1 presents the main results about optimal quantization that are required bellow. The emphasis is on the functional quantization of Gaussian processes. Section 1.2 presents the first historic quantization-based variance reduction method:

using quantization as a control variate variable, as proposed in [25, 18]. Then Section 1.3 outlines the links between quantization and stratification. The emphasis is on the Gaussian case. The method is specified in the functional case for Gaussian processes in Section 1.4. We present a simulation method for the Brownian motion and other examples of Gaussian processes (such as the Ornstein-Uhlenbeck process and the Brownian bridge) that preserves the $O(n)$ simulation complexity where n is the number of time steps. In Section 1.5, we provide numerical experiments of the method with option pricing problems arising in mathematical finance. Appendix 1.A presents the computation of the Karhunen-Loève decomposition of Ornstein-Uhlenbeck processes, and the related numerical methods. A procedure for the computation of the eigenvalues is provided. Appendix 1.B provides closed-form expressions of a regression matrix needed for the functional stratification fast simulation algorithm, in the cases of the standard Brownian motion, the standard Brownian bridge and Ornstein-Uhlenbeck processes.

1.1 Optimal quantization, the abstract framework

1.1.1 Introduction to quantization of random variables

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

The principle of the quantization of a random variable X taking its values in E is to approximate X by a random variable Y taking a finite number N of values in E. The discrete random variable Y is a quantizer of X.

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

$$\min\left\{\|X - Y\|_p, Y : \Omega \to E \text{ measurable, } \operatorname{card}(Y(\Omega)) \leq N\right\}. \tag{1.1}$$

Definition 1.1.1 (Voronoi partition)**.** *Consider $N \in \mathbb{N}^*$, $\Gamma = \{\gamma_1, \cdots, \gamma_N\} \subset E$ and let $C = \{C_1, \cdots, C_N\}$ be a Borel partition of E. C is a Voronoi partition associated with Γ if $\forall i \in \{1, \cdots, N\}$, $C_i \subset \{\xi \in E, |\xi - \gamma_i| = \min_{j \in \{1, \cdots, N\}} |\xi - \gamma_j|\}$.*

If $C = \{C_1, \cdots, C_N\}$ is a Voronoi partition associated with $\Gamma = \{\gamma_1, \cdots, \gamma_N\}$, it is clear that $\forall i \in \{1, \cdots, N\}, \gamma_i \in C_i$. C_i is called Voronoi slab associated with γ_i in C and γ_i is the centre of the slab C_i.
One denotes $C_i = \operatorname{slab}_C(\gamma_i)$, and for every $a \in \Gamma$, $W(a|\Gamma)$ is the closed subset of E defined by $W(a|\Gamma) = \left\{y \in E, |y - a| = \min_{b \in \Gamma} |y - b|\right\}$.

Definition 1.1.2 (Nearest neighbour projection)**.** *Let us consider the fixed point set $\Gamma = \{\gamma_1, \cdots, \gamma_N\} \subset E$ and $C = \{C_1, \cdots, C_N\}$ the associated Voronoi partition. The nearest neighbour projection onto Γ is the application $\operatorname{Proj}_\Gamma := \sum_{i=1}^{N} \gamma_i \mathbf{1}_{C_i}$.*

Proposition 1.1.1. *Let X be an E-valued L^p random variable, and Y taking its values in the fixed point set $\Gamma = \{\gamma_1, \cdots, \gamma_N\} \subset E$ where $N \in \mathbb{N}$. Set $\widehat{X}^\Gamma$ the random variable defined by $\widehat{X}^\Gamma := \operatorname{Proj}_\Gamma(X)$ where $\operatorname{Proj}_\Gamma$ is a nearest neighbour projection onto Γ, called a Voronoi Γ-quantizer of X.*
Then we clearly have $\left|X - \widehat{X}^\Gamma\right| \leq |X - Y|$ a.s.. Hence $\left\|X - \widehat{X}^\Gamma\right\|_p \leq \|X - Y\|_p$.

As a consequence of the previous proposition, solving the minimization problem (1.1) amounts to solving the simpler minimization problem

$$\min\left\{\|X - \operatorname{Proj}_\Gamma(X)\|_p, \ \Gamma \subset E, \ \operatorname{card}(\Gamma) \leq N\right\}. \tag{1.2}$$

The quantity $\|X - \mathrm{Proj}_\Gamma(X)\|_p$ is called the mean L^p-quantization error. When this minimum is reached, one refers to optimal quantization.
The problem of the existence of a minimum has been investigated for decades on its numerical and theoretical aspects in the finite-dimensional case [23, 11].

- For every $N \geq 1$, the L^p-quantization error is Lipschitz continuous and reaches a minimum. An N-tuple that achieves the minimum has pairwise distinct components, as soon as $\mathrm{card}(\mathrm{supp}(\mathbb{P}_X)) \geq N$. This result stands in the general abstract case of a random variable valued in a reflexive separable Banach space. (This has been prooved in [20].)

- If $\mathrm{card}(X(\Omega))$ is infinite, this minimum strictly decreases to 0 as N goes to infinity. The rate of convergence is ruled by Theorem 1.1.2 in the finite-dimensional case.

Theorem 1.1.2 (Zador).
- (Sharp rate) *(See [11]) Let $r > 0$ and $X \in L^{p+\eta}(\mathbb{P})$ for some $\eta > 0$. Let $\mathbb{P}_X(d\xi) = \phi(\xi)d\xi + \mu(d\xi)$ be the canonical decomposition of the distribution of X (μ and the Lebesgue measure are singular). Then, (if $\phi \not\equiv 0$), the L^r quantization error of level N, $\mathcal{E}_{N,r}$ satisfies*

$$\mathcal{E}_{N,r}(X, \mathbb{R}^d) \underset{N\to\infty}{\sim} \widetilde{J}_{r,d} \times \left(\int_{\mathbb{R}^d} \phi^{\frac{d}{d+r}}(u)du \right)^{\frac{1}{d}+\frac{1}{r}} \times N^{-\frac{1}{d}}, \tag{1.3}$$

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

- (Non-asymptotic upper bound) *(See [22]) Let $d \geq 1$. There exists $C_{d,r,\eta} \in (0, \infty)$ such that, for every $\mathbb{R}^d$-valued random vector X,*

$$\forall N \geq 1, \qquad \mathcal{E}_{N,r}(X, \mathbb{R}^d) \leq C_{d,r,\eta} \|X\|_{r+\eta} N^{-\frac{1}{d}}. \tag{1.4}$$

This mainly says us that $\min \left\{ \left\| X - \widehat{X} \right\|_p, \ \mathrm{card}(\Gamma) \leq N \right\} \underset{N\to\infty}{\sim} C_{\mathbb{P}_X,p,d} N^{-\frac{1}{d}}$. The first statement of the theorem was first proved for distributions with compact supports by Zador in [32]. Then a first extension to general probability distributions on $\mathbb{R}^d$ is developed in [5]. The first mathematically rigorous proof can be found in [11]. The non-asymptotic error bound of the second statement is proved in [22].

In Figure 1.1, the Voronoi partition of a random N-quantizer and an L^2-optimized N quantizer of the $\mathcal{N}(0, I_2)$ distribution are given.

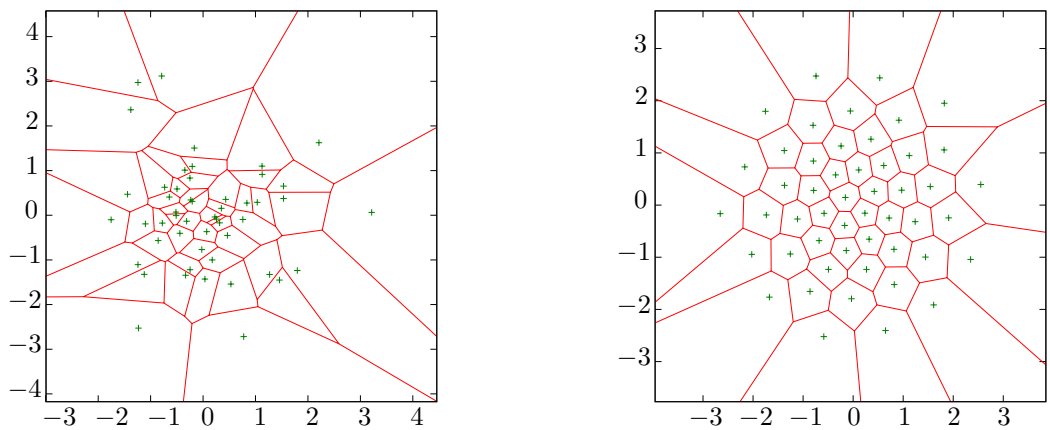

Figure 1.1: Voronoi partition of a random quantizer and a L^2-optimized N-quantizer of the $\mathcal{N}(0, I_2)$ distribution in $\mathbb{R}^2$. ($N = 48$).

1.1.2 Stationarity and centroidal Voronoi tessellations

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

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

Definition 1.1.3 (Stationarity). *A quantizer Y of X is stationary (or self-consistent) if*

$$Y = \mathbb{E}[X|Y]. \tag{1.5}$$

Proposition 1.1.3 (Stationarity of L^2-optimal quantizers). *A (quadratic) optimal quantizer is stationary.*

We refer to [11] for a detailed proof in the finite-dimensional setting and to [20] in the separable Hilbert setting.
The stationarity is a particularity of the quadratic case ($p = 2$). In other L^p cases, a similar property involving the notion of p-centre occurs. A proof of is available in [12].
A consequence is, if $Y = \mathrm{Proj}_\Gamma(X)$ is an L^2-optimal quantizer, and $C = \{C_1, \cdots, C_n\}$ is the associated Voronoi partition, one has $\forall y \in \Gamma, y = \mathbb{E}[X|X \in \mathrm{slab}_C(y)]$.

Proposition 1.1.4. *Let X be an H-valued L^2 random variable. Let us denote by D_N^X the squared quadratic quantization error associated with a codebook of size N with respect to X.*

$$\begin{array}{rccc} D_N^X : & H^N & \to & \mathbb{R}_+ \\ & \Gamma = (\gamma_1, \cdots, \gamma_N) & \mapsto & \mathbb{E}\left[\min\limits_{1 \le i \le N} |X - \gamma_i|_H^2\right]. \end{array}$$

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

$$\nabla D_N^X(\Gamma) = 2\Big(\int_{C_i(\Gamma)} (\gamma_i - \xi)\mathbb{P}_X(d\xi)\Big)_{1 \le i \le N} = 2\left(\mathbb{E}\left[\left(\widehat{X}^\Gamma - X\right)\mathbf{1}_{\left\{\widehat{X}^\Gamma = \gamma_i\right\}}\right]\right)_{1 \le i \le N}. \tag{1.6}$$

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

We refer to [27] for a detailed proof.

Definition 1.1.4 (Centroidal projection). *Let $C = \{C_1, \cdots, C_N\}$ be a Borel partition of H. Let us define for $1 \le i \le N$, $G_i = \begin{cases} \mathbb{E}[X|X \in C_i] & \text{if } \mathbb{P}[X \in C_i] \neq 0, \\ 0 & \text{in the other case,} \end{cases}$ the centroids associated with X and C.*

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

Lemma 1.1.5 (Huyghens, variance decomposition). *Let X be a H-valued L^2 random variable, $N \in \mathbb{N}^*$ and $C = (C_i)_{1 \le i \le N}$ a Borel partition of H. Consider $\mathrm{Proj}_{C,X} = \sum\limits_{i=1}^N G_i \mathbf{1}_{C_i}$ the associated centroidal projection. Then one has,*

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

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

Proof:

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

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

1.1.3 Optimal quantization and principal component analysis

Reduction of dimension

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

Proposition 1.1.6. *Let U be a finite-dimensional linear subspace of H. Then*

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

In other words, the quadratic quantization error with respect to $\Gamma \subset U$ consists of the projection error and the quantization error of the projected random variable. We refer to [20] for a detailed proof.

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

It follows from Proposition 1.1.6 that

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

Covariance operator of a random variable

Definition 1.1.5. *Let X be a centered H-valued L^2 random variable.*

The covariance operator $C_X : H \to H$ of X is defined by $C_X y = \mathbb{E}[\langle y, X\rangle X]$.

1. In the finite-dimensional case, the matrix of C_X in the canonical basis is the covariance matrix of X.

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

$$C_X y = \int_{[0,T]} y(s)\Gamma_X(s,\cdot)ds, \quad y \in L^2([0,T], dt). \tag{1.7}$$

In [20], it is proved that linear subspaces U of H spanned by n-stationary codebooks of Gaussian measures correspond to principal components of X. In other words, they are spanned by eigenvectors of C_X corresponding to the m largest eigenvalues. Thus these subspaces correspond to the first m principal components of X.

Theorem 1.1.7. *Let Γ be an optimal codebook for the Gaussian random variable X, $U = \mathrm{span}(\Gamma)$ and $m = \dim U$. Then $C_X(U) = U$ and $\mathbb{E}\left[|X - \Pi_U(X)|^2\right] = \sum_{j \geq m+1} \lambda_j^X$, where $\lambda_1^X \geq \lambda_2^X \geq \cdots > 0$ are the ordered non-zero eigenvalues of C_X (written as many times as their multiplicity).*

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

We now deduce the final representation of $\mathcal{E}_N(X)$.

$$\mathcal{E}_N(X)^2 = \sum_{j \geq m+1} \lambda_j^X + \mathcal{E}_N\left(\bigotimes_{j=1}^m \mathcal{N}\left(0, \lambda_j^X\right)\right)^2 \quad \text{for } m \geq d_N(X), \tag{1.8}$$

$$\mathcal{E}_N(X)^2 < \sum_{j \geq m+1} \lambda_j^X + \mathcal{E}_N\left(\bigotimes_{j=1}^m \mathcal{N}\left(0, \lambda_j^X\right)\right)^2 \quad \text{for } 1 \leq m < d_N(X). \tag{1.9}$$

A detailed proof of this result is available in [20]. Equations (1.8) and (1.9) show that for the quantization of a Gaussian process X, as soon as we know its Karhunen-Loève basis $(e_n^X)_{n \in \mathbb{N}^*}$ and its eigenvalues $(\lambda_n^X)_{n \in \mathbb{N}^*}$, the problem of optimal L^2-quantization comes to the problem of the quantization of a Gaussian vector of dimension d_N.

1.1.4 Product quantization

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

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

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

Proposition 1.1.8 (Case of independent marginals)**.** *With the same notations, if one assumes that the marginals of X, $(\langle X, e_1 \rangle, \langle X, e_2 \rangle, \cdots)$ are independent, then one can choose for each $k \in I$ the values $\Gamma^k = \left\{x_1^k, \cdots, x_{N_k}^k\right\}$ such that $Y^k = \mathrm{Proj}_{\Gamma^k}(\langle X, e_k \rangle)$ is a stationary quantizer of $\langle X, e_k \rangle$. Then $Y = \mathrm{Proj}_\Gamma(X)$ is a stationary quantizer of X.*

This method yields a stationary quantizer with a simple projection rule. A drawback of product quantization is that one needs to restrict to the case of independent marginals in order to preserve stationarity.

1.1.5 Numerical optimal quantization

Various numerical algorithms have been developed to numerically obtain an optimal N-grid with a minimal quadratic quantization error in the finite-dimensional setting. A review of these methods is available in [27]. Let us mention Lloyd's algorithm for the quadratic case, which is the natural probabilistic counterpart of a classification algorithm due to Forgy [8].

Another algorithm is a stochastic gradient method which is suggested by the fact that the L^2-quantization distortion function is differentiable at any N-tuple having pairwise distinct components and a $\mathbb{P}_X$-negligible Voronoi tessellation boundary and has an integral representation. The algorithm is deeply investigated in [24].

Equation (1.6) shows that any Voronoi quantizer associated with a critical point of D_N^X is a stationary quantizer. In the case of one-dimensional distributions, such as the Gaussian distribution, the Hessian of the distortion is known and can be represented by a tridiagonal matrix. Hence, it is easy to invert and a Newton-Raphson method can be implemented. It is completely detailed in [24] in the Gaussian case. It remains the fastest way to compute L^2-optimal quantizers of one-dimensional Gaussian variables.

1.1.6 Quantization of Gaussian processes

Quantization

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

We have seen in Section 1.1.3 that in this context, as soon as one knows the Karhunen-Loève system $(e_n^X, \lambda_n^X)_{n\in\mathbb{N}^*}$ of the covariance operator of X, the problem of the L^2-optimal quantization of the process X comes to the quantization of a finite-dimensional Gaussian vector $\bigotimes_{j=1}^m \mathcal{N}(0, \lambda_j^X)$, for some positive integer m, the quantization dimension. The companion parameters of the functional quantizer are easily deduced from the quantizer of $\bigotimes_{j=1}^m \mathcal{N}(0, \lambda_j^X)$ that is used.

All this is valid for any Gaussian process X with a continuous covariance function, as soon as one knows its Karhunen-Loève basis. Several usual Gaussian processes have explicit Karhunen-Loève expansions, such as the Brownian motion and the Brownian bridge. The Ornstein-Uhlenbeck process admits a semi-closed-form for its Karhunen-Loève expansion. (The formula is derived for normalized parameters in the stationary case in [13, p.195].) In Section 1.A, the computation of Karhunen-Loève decomposition of the Ornstein-Uhlenbeck process is detailed in the general Gaussian case ($r_0 \overset{\mathcal{L}}{\sim} \mathcal{N}(m_0, \sigma_0^2)$). As far as we know, the K-L expansion of the fractional Brownian motion is not known.

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

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

$$e_n^W(t) := \sqrt{\frac{2}{T}} \sin\left(\pi(n-1/2)\frac{t}{T}\right), \qquad \lambda_n^W := \left(\frac{T}{\pi(n-1/2)}\right)^2, \quad n \geq 1. \tag{1.10}$$

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

$$e_n^B(t) := \sqrt{\frac{2}{T}} \sin\left(\pi n\frac{t}{T}\right), \qquad \lambda_n^B := \left(\frac{T}{\pi n}\right)^2, \quad n \geq 1. \tag{1.11}$$

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

$$dr_t = -\theta r_t dt + \sigma dW_t, \tag{1.12}$$

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

$$e_n^{OU}(t) := \left(\frac{1}{\sqrt{\frac{T}{2} - \frac{\sin(2\omega_n T)}{4\omega_n}}}\right) \sin(\omega_n t), \qquad \lambda_n^{OU} := \frac{\sigma^2}{\omega_n^2 + \theta^2}, \quad n \geq 1, \tag{1.13}$$

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

$$\theta \sin(\omega_n T) + \omega_n \cos(\omega_n T) = 0,$$

sorted in an increasing order. (Based on results from Section 1.A.)

4. The stationary Ornstein-Uhlenbeck process on $[0,T]$. (See Section 1.A.)

In Figure 1.2, one can see an N-optimal L^2-quantizer of the standard Brownian motion.

Product quantization

Thanks to Equations (1.8) and (1.9), product quantization of the finite-dimensional Gaussian vector $\xi \overset{\mathcal{L}}{\sim} \bigotimes_{j=1}^m \mathcal{N}(0, \lambda_j^X)$ yields a stationary quantizer of the process X. In this context, let us introduce the following notations:

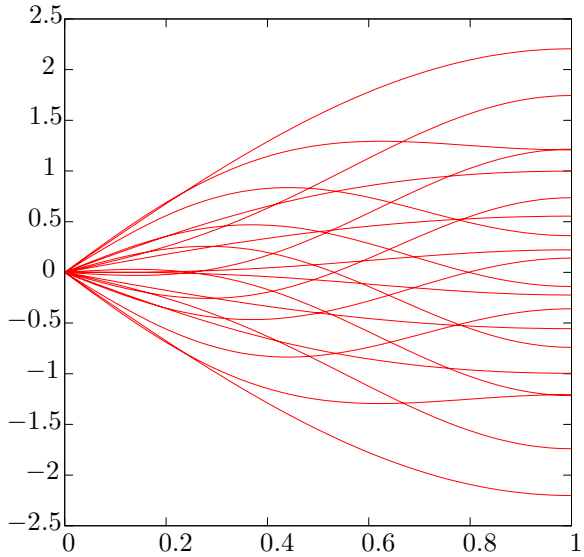

Figure 1.2: Optimal quantizer of a standard Brownian motion on $[0,1]$.

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

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

$$\chi_{\underline{i}} = \sum_{n\geq 1} \sqrt{\lambda_n^X} x_{i_n}^{(N_n)} e_n^X. \tag{1.14}$$

A quantizer χ defined by Equation (1.14) is called a K-L product quantizer. Furthermore, one denotes by $\mathcal{O}_{pq}(X,N)$ the set of the K-L product quantizers of size at most N of X.

In the case of a product quantization, the counterpart of Equation (1.8) is

$$\begin{aligned}\mathbb{E}\left[\min_{\underline{i}} \left|X - \chi_{\underline{i}}\right|^2\right] &= \sum_{n=1}^{m} \lambda_n^X \mathbb{E}\left[\min_{1\leq i_n\leq N_n} \left|\xi_n - x_{i_n}^{(N_n)}\right|\right] + \sum_{n\geq m+1} \lambda_n^X \\ &= \sum_{n=1}^{m} \lambda_n^X \mathbb{E}\left[\min_{1\leq i_n\leq N_n} \left|\xi_n - x_{i_n}^{(N_n)}\right|\right] + \mathbb{E}\left[|X|_{L_T^2}^2\right] - \sum_{n=1}^{m} \lambda_n^X,\end{aligned} \tag{1.15}$$

where m is the quantization dimension.

Product decomposition blind optimization

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

$$\min\Big\{e(\chi),\ \chi \in \mathcal{O}_{pq}(X,N)\Big\}, \tag{1.16}$$

that is, thanks to Equation (1.15)

$$\min\Big\{\sum_{n=1}^{d} \lambda_n^X \min_{\mathbb{R}^{N_n}} \left\|\xi - \widehat{\xi}_n\right\|_2^2 + \sum_{n\geq d+1} \lambda_n^X,\ N_1 \times \cdots \times N_n \leq N,\ d \geq 1\Big\}. \tag{1.17}$$

A solution of (1.16) is called an optimal K-L product quantizer.

The blind optimization procedure consists in computing the criterion for every possible decomposition $N_1 \times \cdots \times N_n \leq N$. For a given Gaussian process X, results can be kept off-line for a future use. Optimal decompositions for a wide range of values of N for both Brownian bridge and Brownian motion are available on the web site `www.quantize.maths-fi.com` [26] for download. The blind optimization procedure is more thoroughly described in [25]. Let us remind that the optimal decomposition depends on the parameters of the Ornstein-Uhlenbeck process (σ and θ in Equation (1.12)) and the maturity.

Some values of optimal decompositions for the stationary Ornstein-Uhlenbeck process are given in Table 1.1.

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

Table 1.1: Record of optimal product decomposition values of the stationary centered Ornstein-Uhlenbeck process given by $dr_t = -\theta r_t dt + \sigma dW_t$ on $[0, T]$ with $\theta = 1$, $\sigma = 1$ and $T = 3$.

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

In Figure 1.3, one can see examples of optimal product quantizers of the Brownian motion and the Brownian bridge on $[0, 1]$. In Figure 1.4, one can see optimal product quantizers of the centered Ornstein-Uhlenbeck process starting from $r_0 = 0$ and a stationary Ornstein-Uhlenbeck on $[0, 3]$.

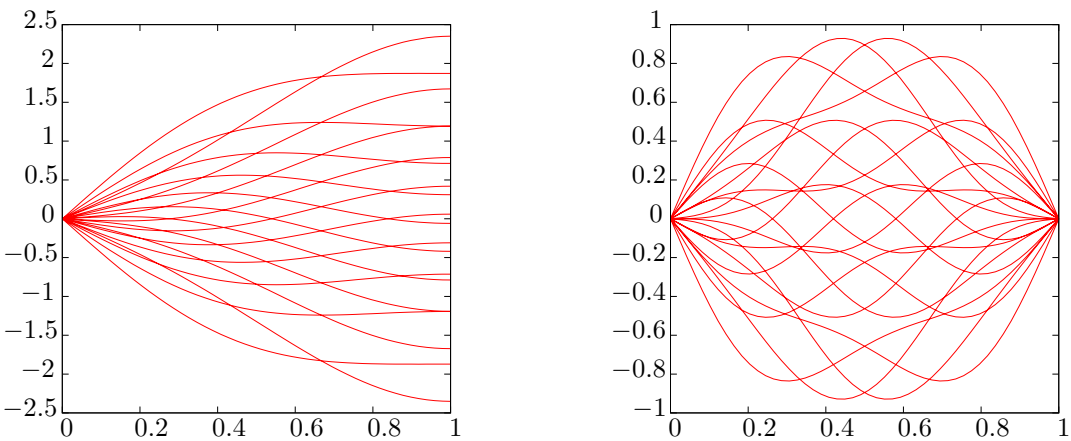

Figure 1.3: Optimal product quantizer of a standard Brownian motion (left) and a standard Brownian bridge (right) on $[0, 1]$.

Rate of decay for the quantization error

In [20], a precise link between the rate problem and Shannon-Kolmogorov's entropy of X is established. This allowed them to compute the exact rate of convergence of the minimal L^2-quantization error under rather general conditions on the eigenvalues of the covariance operator. Typical rates are $O(\log(n)^{-a})$, $a > 0$. This conditions are fulfilled by a large class of processes, such as the Ornstein-Uhlenbeck process and the Brownian motion.

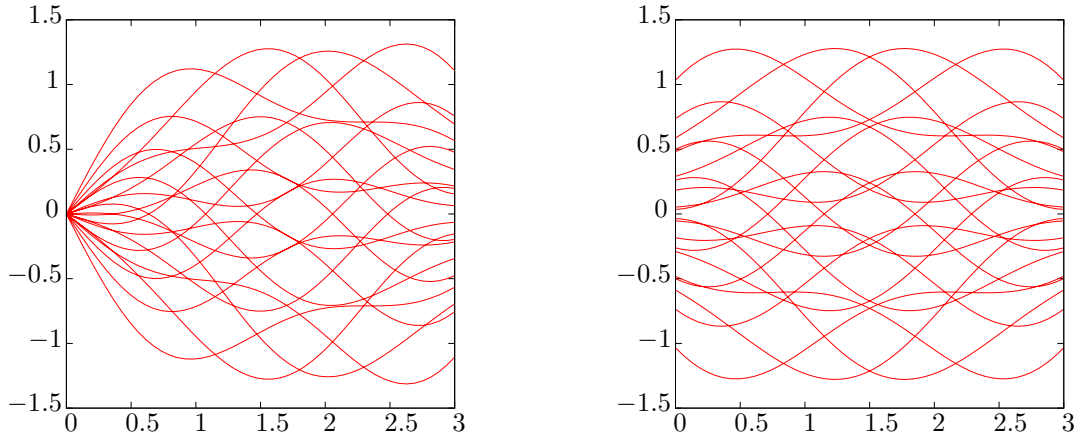

Figure 1.4: Optimal product quantizer of a centered Ornstein-Uhlenbeck process, starting from $r_0 = 0$ (left) and stationary (right) given by $dr_t = -r_t dt + dW_t$, on $[0, 3]$.

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

This method has been initially proposed in [25].

1.2.1 Quantization as a control variate variable

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

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

$$\begin{aligned}\mathbb{E}[F(X)] &= \mathbb{E}\left[F(\mathrm{Proj}(X))\right] + \mathbb{E}\left[F(X) - F(\mathrm{Proj}(X))\right] \\ &= \underbrace{\mathbb{E}\left[F(\mathrm{Proj}(X))\right]}_{(a)} + \underbrace{\frac{1}{M}\sum_{m=1}^{M} F\left(X^{(m)}\right) - F\left(\mathrm{Proj}\left(X^{(m)}\right)\right)}_{(b)} + R_{N,M},\end{aligned} \tag{1.18}$$

where $X^{(m)}, 1 \leq m \leq M$ are M independent copies of X, and $R_{N,M}$ is a remainder term defined by Equation (1.18).

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

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

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

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

$$\left\| F(X) - F\left(\mathrm{Proj}^N(X)\right) \right\|_2 \leq [F]_{\mathrm{Lip}} \frac{C_X}{N^{1/d}}$$

and

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

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

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

and

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

1.2.2 Practical implementation: the problem of fast nearest neighbour search

• **The complexity of the projection:** Concerning practical implementation, one notices in Equation (1.18) that for every step of the Monte-Carlo simulation, one has to compute the projection $\mathrm{Proj}(X^{(m)})$. This is the critical part of the algorithm when dealing with optimal quantization. Hence, the efficiency of the quantization as a control variate variable is conditioned by the efficiency of the projection procedure. When dealing with Voronoi quantization, this is the nearest neighbour projection.

The problem of nearest neighbour projection, also known as the post-office problem [17], has been widely investigated in the area of computational geometry. It is encountered for many applications, such as pattern recognition and information retrieval.

The problem has been solved near optimally for the case of low dimensions. Algorithms differ on their practical efficiency on real data sets. For large dimensions, most solutions have a complexity that is exponential with the dimension, or require a longer query time than the obvious brute force algorithm. In fact for dimension $d > \log N$, a brute force algorithm is usually the best choice. This effect is known as the curse of dimensionality. Still, even in low dimension, fast nearest neighbour search is a critical part of the algorithm. We refer to [31] for a review about fast nearest neighbour search algorithms. Let us also mention [6] for a fast nearest neighbor search algorithm based on vector quantization.

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

• **The functional case:** Another drawback of the method, when dealing with the functional case is that one does not simulate the whole trajectory of the stochastic process but only its marginals at discrete dates. Hence it is not possible to compute its projection. This problem finds its solution in the simulation scheme for Gaussian processes derived in Section 1.4.2 for the functional stratification.

A variance reduction technique using a functional quantizer of the Brownian motion as a control variate has been proposed in [18].

1.3 Application of quantization to stratification

1.3.1 A short background on stratification

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

- Let $(A_i)_{i\in I}$ be a finite $\mathcal{E}$-measurable partition of E. The sets A_i are called *strata*. Assume that the weights $p_i = \mathbb{P}(X \in A_i)$ are known for $i \in I$ and strictly positive.

- Let us define the collection of independent random variables $(X_i)_{i\in I}$ with distribution $\mathcal{L}(X|X \in A_i)$.

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

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

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

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

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

$$\overline{F(X)}^I_M := \sum_{i\in I} p_i \frac{1}{M_i} \sum_{k=1}^{M_i} F\left(X_i^k\right), \tag{1.19}$$

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

Proposition 1.3.1. *With the same notations:*

$$\mathrm{Var}\left(\overline{F(X)}^I_M\right) = \frac{1}{M} \sum_{i\in I} \frac{p_i^2}{q_i} \sigma^2_{F,i}, \tag{1.20}$$

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

Proof: Let us denote $Z_i = \frac{1}{M_i} \sum_{k=1}^{M_i} F(X_i^k)$. The random variables $(Z_i)_{i\in I}$ are independent. We have $\overline{F(X)}^I_M = \sum_{i\in I} p_i Z_i$. Hence, by independence,

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

□

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

$$\min_{(q_i)\in\mathcal{P}_I} \sum_{i\in I} \frac{p_i^2}{q_i} \sigma^2_{F,i} \quad \text{where } \mathcal{P}_I = \left\{ (q_i)_{i\in I} \in \mathbb{R}_+^I \,\middle|\, \sum_{i\in I} q_i = 1 \right\}. \tag{1.21}$$

Sub-optimal choice

The first natural choice is to set

$$q_i = p_i, \qquad i \in I. \tag{1.22}$$

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

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

Optimal choice

The optimal choice is the solution of the constrained minimization problem (1.21). Schwarz's inequality yields

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

As a consequence, the solution of the minimization problem corresponds to the equality case in Schwarz's inequality. Hence the solution of the minimization problem is given by

$$q_i^* = \frac{p_i\sigma_{F,i}}{\sum\limits_{j\in I} p_j\sigma_{F,j}}, \quad i \in I \tag{1.23}$$

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

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

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

so that

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

In [29], Étoré and Jourdain proposed an algorithm for adaptively modifying the proportion of further drawings in each stratum, that converges to the optimal allocation. This can be used in a general framework.

In Section 1.3.2, we will see that the problem of designing good strata, in term of variance reduction is linked with the problem of optimal quantization. Moreover, the case of quantization-based strata have two other advantages:

- The weights p_i are already known, which saves us from evaluating their values during the Monte-Carlo evaluation.
- As concerns the optimal choice for the allocation parameters q_i, one shows in Theorem 1.3.2 that weights can be chosen such that stratification has a uniform efficiency among the class of Lipschitz continuous functionals. This weights have a closed-form expression in the case of quantization-based stratification.

1.3.2 Stratification and quantization

The main drawback induced by using quantization as a control variate variable is that it requires repeated computations of projections onto the quantizer. (Nearest neighbour search in the case of a Voronoi quantizer.) The point when dealing with stratification is that *one does not have to use a projection procedure.* The critical point now is the cost of the simulation of conditional distributions $\mathcal{L}(X|X \in A_i)$, $i \in I$.

Theorem 1.3.2 brings together previous results about stratification and highlights the relationships with the notions of local inertia and intraclass inertia. It stresses the fact that stratification has a uniform efficiency among the class of Lipschitz continuous functionals.

Theorem 1.3.2 (Universal stratification)**.** *Let $A = (A_i)_{i\in I}$ be a partition (stratification) of E. (Keep in mind the notation* $\mathrm{Proj}_{A,Z}$ *for the centroidal projection associated with the random variable Z and the partition A, defined in Definition 1.1.4).*

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

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

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

$$\forall i \in I, \quad \sigma_{F,i} \leq [F]_{\mathrm{Lip}}\sigma_i \quad \text{so that} \quad \sup_{[F]_{\mathrm{Lip}}\leq 1} \sigma_{F,i} \leq \sigma_i. \tag{1.24}$$

2. *In the case of the sub-optimal choice (see Section 1.3.1),*

$$\begin{aligned} \sup_{[F]_{\mathrm{Lip}}\leq 1}\Big(\sum_{i\in I} p_i\sigma_{F,i}^2\Big) \leq \sum_{i\in I} p_i\sigma_i^2 &= \Big\|X - \mathbb{E}[X|\sigma(\{X \in A_i\},\ i \in I)]\Big\|_2^2 \\ &= \Big\|X - \mathrm{Proj}_{A,X}(X)\Big\|_2^2. \end{aligned} \tag{1.25}$$

3. *In the case of the optimal choice (see Section 1.3.1),*

$$\sup_{[F]_{\mathrm{Lip}}\leq 1}\Big(\sum_{i\in I} p_i\sigma_{F,i}\Big)^2 \leq \Big(\sum_{i\in I} p_i\sigma_i\Big)^2, \tag{1.26}$$

and

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

4. *If one considers vector-valued Lipschitz continuous functions $F : E \to E$, then inequalities (1.24), (1.25) and (1.26) hold as equalities.*

Proof: One has

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

Now using that F is Lipschitz, it follows that

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

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

The idea now is to set $I = \{1, \cdots, N\}$ and use the partition $\{A_1, \cdots, A_N\}$ and the N-codebook $\Gamma = \{\gamma_1, \cdots, \gamma_N\}$ associated with the projection $\mathrm{Proj} = \sum_{i=1}^N \gamma_i \mathbf{1}_{A_i}$. In the case of a Voronoi quantization, we have $A_i = \mathrm{slab}_A(\gamma_i)$.
Then for every $i \in I$, there exists a Borel function $\phi(\gamma_i, \cdot) : [0,1]^q \to E$ such that $\phi(\gamma_i, U) \overset{\mathcal{L}}{\sim} \mathcal{L}(X|X \in A_i) = \frac{\mathbf{1}_{A_i}\mathbb{P}_X(d\xi)}{\mathbb{P}[X\in A_i]}$, where $U \overset{\mathcal{L}}{\sim} \mathcal{U}([0,1]^q)$.

Theorem 1.3.2 suggests, in the case of Lipschitz continuous functional to set

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

so that we have a uniform efficiency among the class of Lipschitz continuous functionals. This budget allocation method will be further mentionned as the "universal stratification" weights.

Remark. *Note that the dimension $q \in \{1, 2, \cdots\}$ is arbitrary: one may always assume that $q = 1$ by the fundamental theorem of simulation, but in order to obtain some closed-form expression for $\phi(\gamma_i, \cdot)$, we are led to consider situations where $q \geq 2$ or even infinite when considering a Von Neumann acceptance-rejection method.*

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

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

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

Remark. *Although we focus here on the universality of quantization for designing strata when dealing with Lipschitz continuous functionals, let us mention the adaptative strata design methods recently proposed in [30] and [16] for more general functionals.*

1.3.3 Simulability for hyper-rectangles strata in the independent Gaussian case

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

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

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

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

Then $p_{\underline{i}} = \mathbb{P}(A_{\underline{i}}) = \prod_{k=1}^{d} \left(\mathcal{N}(x_{i_k}^k) - \mathcal{N}(x_{i_k-1}^k)\right)$ and for $-\infty \leq a \leq b \leq \infty$,

$$\mathcal{L}\left(Z | Z \in [a,b]\right) = \mathcal{N}^{-1}\left(\left(\mathcal{N}(b) - \mathcal{N}(a)\right) U + \mathcal{N}(a)\right), \qquad U \overset{\mathcal{L}}{\sim} \mathcal{U}([0,1]). \tag{1.27}$$

1.4 Functional stratification of a Gaussian process

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

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

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

What is done in this section can be easily generalized to multidimensional processes in the case where their coordinates are independent. (For example, when dealing with multi-factor Brownian

diffusions, it does not matter how the Brownian motions are being correlated afterward.) Still we restrict ourselves to the one-dimensional setting for clarity.

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

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

with the same notations as in Section 1.1.6.

We now need to be able to simulate the conditional distribution

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

where $A_{\underline{i}}$ is the slab associated with $\chi_{\underline{i}}$ in the codebook.

To simulate the conditional distribution $\mathcal{L}(X|X \in A_{\underline{i}})$, one will :

- First, simulate the first K-L coordinates of X, using (1.27).
- Then simulate the conditional distribution of the marginals of the Gaussian process, its first coordinates being fixed.

Remark. *We have chosen to use K-L optimal product quantizers instead of optimal quantizers because in this case, the Voronoi cells in this are hyper-rectangles, which allows us to simulate the first K-L coordinates more easily than in the general case. Moreover, the rate of decay of the quantization error is rate-optimal under some conditions on the Karhunen-Loève eigenvalues which are verified in the considered examples [20], and the actual value of the quadratic distorsion remains very close to the optimal value in practice.*

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

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

$$\mathcal{L}\Big(X_{t_0}, \cdots, X_{t_n} \Big| \int_0^T X_s e_1^X(s)ds, \int_0^T X_s e_2^X(s)ds, \cdots, \int_0^T X_s e_d^X(s)ds\Big) \tag{1.28}$$

where $(X_t)_{t\in[0,T]}$ is a L^2 $\mathbb{R}$-valued Gaussian process, and $(e_k^X, \lambda_k^X)_{k\in\mathbb{N}^*}$ is the Karhunen-Loève system associated with the process X.

As X is a Gaussian process, $\Big(X_{t_0}, \cdots, X_{t_n}, \int_0^T X_s e_1^X(s)ds, \cdots, \int_0^T X_s e_d^X(s)ds\Big)$ is a Gaussian vector. Hence, if we denote $Y := \begin{pmatrix} \int_0^T X_s e_1^X(s)ds \\ \vdots \\ \int_0^T X_s e_d^X(s)ds \end{pmatrix}$ and $V := \begin{pmatrix} X_{t_0} \\ \vdots \\ X_{t_n} \end{pmatrix}$, the conditional distribution (1.28) is given by the transition kernel $\nu(y, A) = \mathcal{N}\left(Af_{V|Y}(y), \mathrm{cov}(V - \mathbb{E}[V|Y])\right)$, where $Af_{V|Y} : \mathbb{R}^d \to \mathbb{R}^n$ is an affine function corresponding to the linear regression of V on Y, $Af_{V|Y}(Y) := \mathbb{E}[V|Y]$.

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

$$R_{V|Y} = \left(\left(e_j^X(t_i)\right)\right)_{0\leq i\leq n, 1\leq j\leq d}. \tag{1.29}$$

- The covariance matrix is

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

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

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

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

1.4.2 Faster simulation of conditional paths - Bayesian simulation

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

- Yet, in the context of this chapter, d is the quantization dimension of the process. It is close to $\log(N)$ if N is the number of strata, and n, the number of time steps, is usually very large compared to d.
- Moreover, we make the assumption that the cost of the simulation of $(X_{t_0}, \cdots, X_{t_n})$ is $O(n)$. (So is the case for the Brownian motion, the Ornstein-Uhlenbeck process or the Brownian bridge for example.)
- The idea here is that the conditional distribution $\mathcal{L}(V|Y)$ is determined through the Bayes lemma, by the conditional distribution $\mathcal{L}(Y|V)$ and the two marginal distributions $\mathcal{L}(V)$ and $\mathcal{L}(Y)$.

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

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

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

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

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

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

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

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

Finally, the algorithm writes:

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

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

Let us remind the fact that the affine function $Af_{V|Y}$ is trivially defined in Equation (1.29), because coordinates of Y are independent. Other matrices implied in this algorithm are computed prior to any Monte-Carlo simulation.

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

Still, in the case of the standard Brownian motion, the standard Brownian bridge and Ornstein-Uhlenbeck processes, there are closed-form expressions for the matrix $R_{Y|V}$, available in Appendix 1.B. In these cases, the numerical least-square regression can be avoided.
In the case of the standard Brownian motion, if $t_j = \frac{jT}{n} = jh$, $0 \leq j \leq n$, this yields $R_{Y|V} = ((\alpha_{ij}))_{1\leq i\leq d, 0\leq j\leq n}$, with

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

Now, we have a very fast and easy way to simulate the conditional distribution (1.28) at our disposal.

In Figures 1.5 and 1.6, we plot a few paths of the conditional distribution of various Gaussian processes knowing that they belong to a given L^2 Voronoi cell. The appearance of the drawing suggests to consider the method as a "guided Monte-Carlo simulation".

1.4.3 Blind optimization procedures

We have seen in Section 1.3.2 that the quantity $d(\chi) = \Big(\sum_{\chi_{\underline{i}} \in \Gamma} p_{\underline{i}} \sigma_{\underline{i}} \Big)^2$ is an upper bound of the variance of the estimator, given in Equation (1.19) in the case where the functional is 1-Lipschitz continuous. Hence one may want to minimize this criterion instead of the L^2-quantization error. This yields the minimization problem

$$\min \Big\{ d(\chi),\ \chi \in \mathcal{O}_{pq}(X, N) \Big\} \tag{1.30}$$

instead of the minimization problem (1.16).

The same kind of blind optimization procedure as in Section 1.1.6 can be performed. Some values of the optimal decomposition for the standard Brownian motion are given in Table 1.2.

Optimal product decompositions for both Brownian bridge and Brownian motion and for a wide range of values of N are available on the web site `www.quantize.maths-fi.com` [26] for download. When comparing all the decompositions obtained for a quantizer size smaller than 11000, one notices that in the case of the Brownian motion, the optimal decompositions for both criteria are "almost" always the same. The only values where decompositions differ are the ranges $270 - 271$ and $3328 - 3359$. The two criteria do not have very different values for the two decompositions. Therefore, in practice, one can use the same decomposition database for the two applications.

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

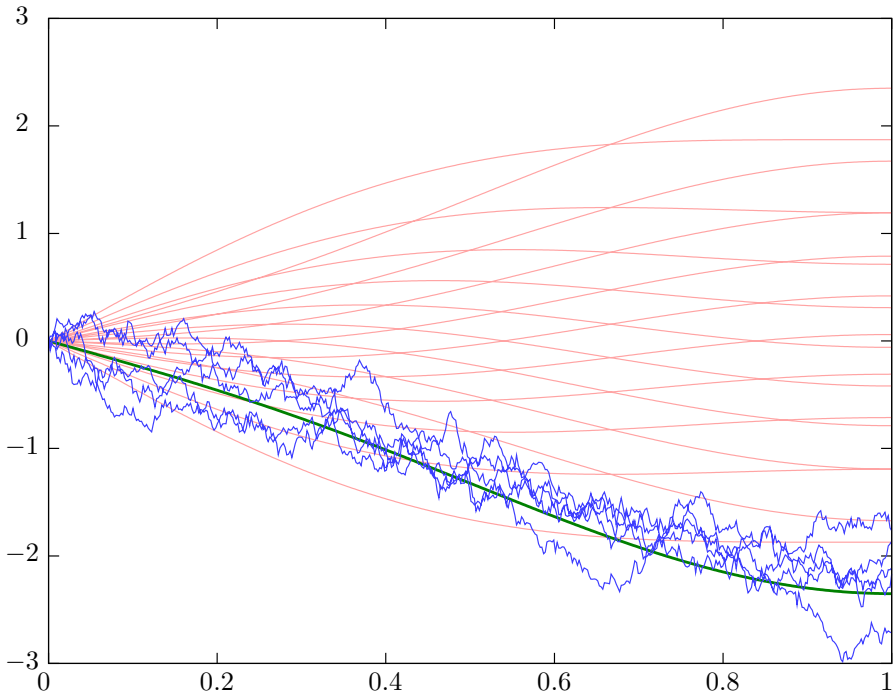

Figure 1.5: Plot of a few paths of the conditional distribution of the Brownian motion, knowing that its path belong to the L^2 Voronoi cell of the highlighted curve in the quantizer.

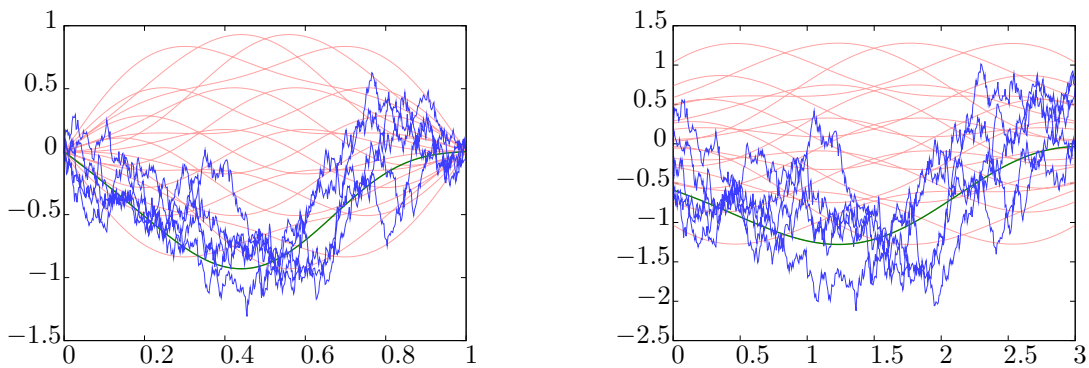

Figure 1.6: Plot of a few paths of the conditional distribution of the Brownian bridge (left) and the stationary Ornstein-Uhlenbeck process (right), knowing that its path belong to the L^2 Voronoi cell of the highlighted curve in the quantizer.

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

Table 1.2: Record of optimal product decomposition record values of the standard Brownian motion with respect to the criterion (1.30).

1.4.4 Functional stratification of solutions of stochastic differential equations

Let us consider the stochastic differential equation

$$dF_t = b(t, F_t)dt + \sigma(t, F_t)dX_t, \quad t \in [0, T], \qquad F_0 = f_0 \tag{1.31}$$

where X is a centered continuous Gaussian semimartingale starting from 0 and where b and σ are Lipschitz continuous in x uniformly in $t \in [0, T]$ and $|b(\cdot, 0)| + |\sigma(\cdot, 0)|$ is bounded over $[0, T]$. In this situation, the SDE (1.31) admits a unique strong solution and $\sup_{t\in[0,T]} |X_t|$ has r-moments for every $r \in (0, \infty)$.

Remark. *In this case, the continuity assumption on the Gaussian process ensures that $\int_0^T \mathbb{E}[X_s^2]ds < \infty$, thanks to Fernique's theorem, and also ensures the continuity of the covariance function, (See [15, VIII.3]).*

The most common approach to perform a Monte-Carlo simulation with the solution of such a stochastic differential equation, is to use a discretization scheme such as the Euler scheme [10]. In this setting, we propose to simply replace Gaussian process X by a stratified version of X in the Euler scheme. This approach is justified in many aspects:

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

2. In the one-dimensional setting, if we make the additional hypothesis on σ and b that

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

   as soon as the drift of the Lamperti transform of the SDE (1.31) is Lipschitz continuous, it is proved in [21] that the unique strong solution of (1.31), seen as a functional of the underlying Gaussian process X is $\|\cdot\|_p$-Lipschitz continuous.

   Hence one stands in the case of a Lipschitz continuous functional where one can use the results of Section 1.3.2 about universal stratification.

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

In the next section, we perform numerical experiments for which a stratified Brownian motion is plugged in the Euler scheme of the considered SDE. Another example with an Ornstein-Uhlenbeck process is also provided.

1.5 Application to option pricing

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

The special case of the Brownian motion allows us to use functional stratification as a generic variance reduction method for the case of functionals of Brownian diffusions. Even in the multidimensional case, no matter how the independent Brownian motions are correlated or used

afterwards; no matter if it is used for diffusing the underlying stock, a stochastic volatility process or an actualization factor. It can be used as a variance reduction method.

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

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

1.5.1 Benchmark with an Up-In Call option pricing in the Black and Scholes model

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

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

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

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

The numerical results are reported in Table 1.3 when using the method with 20 stratas and Table 1.4 when using the method with 100 stratas. In this tables, the first column correspond to Broadie and Glasserman's closed-form expression proxy. The second one corresponds to a simple Monte-Carlo estimator. The last three columns correspond to a stratified sampling estimator with different simulation allocation for each strata.

The "sub-optimal weights" column stands for the allocation budget of Equation (1.22). The "Lip.-optimal weights" column stand for the "universal stratification" budget allocation proposed in Section 1.3.2. Both of these two cases have explicit allocation rules. Last column, "optimal weights" corresponds to an estimation of the optimal budget allocation given in expression (1.23).

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

Table 1.3: Numerical results for the Up-In Call option, with 20 stratas.

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

Table 1.4: Numerical results for the Up-In Call option, with 100 stratas.

1.5.2 Test with an Auto-Call pricing in the CEV model

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

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

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

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

Let us remind the fact that there are closed-form expressions for vanilla option pricing in this model that can be expressed as a function of the noncentral chi-square distribution [14]. A first test of consistency for the method was to check that we could find the same price when performing such a Monte-Carlo simulation. The tested path-dependent payoff that we consider here is the so-called "Auto-Call" payoff.

Description of the Auto-Call payoff:

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

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

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

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

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

Table 1.5: Numerical results for the Auto-Call option in the CEV model, with 20 and 50 stratas.

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

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

$$dS_t = \theta(\alpha - \ln S_t) S_t dt + \sigma S_t dW_t, \tag{1.32}$$

under the risk-neutral probability.

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

$$dX_t = \theta(\mu - X_t)dt + \sigma dW_t \qquad \text{with } \mu = \alpha - \frac{\sigma^2}{2\theta}. \tag{1.33}$$

This model, proposed by Schwartz in [28] is an example of stochastic behavior of commodity prices that takes into account mean reversion. Such exponentials of Ornstein-Uhlenbeck processes

are very common in commodity derivatives models. One particularity in these markets is that the spot is not directly observed. Derivatives mostly rely on futures of the considered commodity. Still, one takes this one factor "toy" model as a simple case study for our variance reduction method.

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

One uses the stratified estimator with the Ornstein-Uhlenbeck process. Optimal product decompositions for the criterion (1.30) are used and available in Table 1.6 where the numerical results are reported.

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

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

Table 1.6: Numerical results for the Asian option in Schwartz's model, with 20, 50 and 100 stratas.

To perform this computation, one had to use a non-centered Ornstein-Uhlenbeck quantizer. Building such a quantizer is a straightforward extension of the centered case. As showed in Section 1.A, if r is an Ornstein-Uhlenbeck process on $[0, T]$ following the dynamic $dr_t = \theta(\mu - r_t)dt + \sigma dW_t$, $r_0 \overset{\mathcal{L}}{\sim} \mathcal{N}(m_0, \sigma_0^2)$, with nonzero values of μ and m_0, one has

$$X_t = \underbrace{m_0 e^{-\theta t} + \mu(1 - e^{-\theta t})}_{(1)=\text{non-stochastic path}} + \left(\begin{array}{c}\text{centered Ornstein-Uhlenbeck process}\\ \text{corresponding to } m_0 = \mu = 0\end{array}\right). \tag{1.34}$$

Hence, one only needs to add the expectation (1) to the centered optimal (product) quantizer to get an optimal (product) quantizer for the non-centered case. An example of such a non-centered Ornstein-Uhlenbeck product quantizer is available in Figure 1.7.

1.5.4 Commentaries on the numerical results

In every tested case, one notices that the quantization-based stratified sampling method reduces noticeably the variance of the Monte-Carlo estimator. The "universal stratification" allocation proposed in Theorem 1.3.2 overcomes the sub-optimal weight allocation. Still in the case of the Auto-Call, its advantage is not very perceptible.

Moreover, the "optimal allocation" estimation yields a very good variance reduction factor. This suggests to implement either a simple prior rough estimation of the optimal allocation or a more sophisticated algorithm such as the one proposed in [29] by Étoré and Jourdain.

1.A Computation of the Karhunen-Loève expansion of the Ornstein-Uhlenbeck process

In this section, one details the Karhunen-Loève decomposition of the Ornstein-Uhlenbeck process. Proposition 1.A.3 brings the results together. Section 1.A.3 presents the numerical method for computing this decomposition.

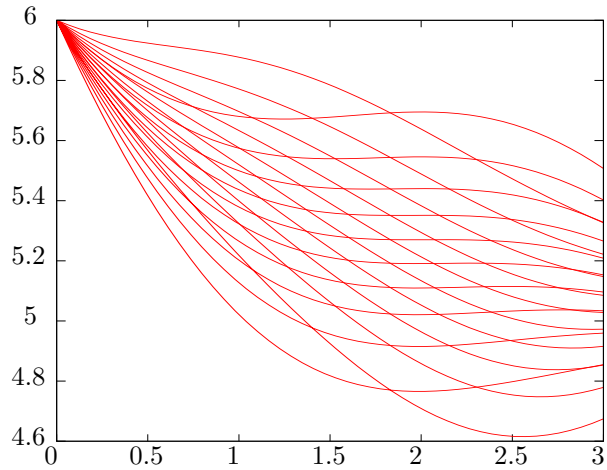

Figure 1.7: Functional 10×2-product quantizer of an Ornstein-Uhlenbeck process starting from $r_0 = 6$ defined by the diffusion $dr_t = \theta(\mu - r_t)dt + \sigma dW_t$ with $\mu = 5$, $\sigma = 0.3$ and $\theta = 0.8$ on $[0, 3]$.

1.A.1 The Ornstein-Uhlenbeck process

The Ornstein-Uhlenbeck process is defined by the SDE

$$dr_t = \theta(\mu - r_t)dt + \sigma dW_t, \qquad \text{with } \sigma \geq 0 \text{ and } \theta > 0. \tag{1.35}$$

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

$$r_t = r_0 e^{-\theta t} + \mu(1 - e^{-\theta t}) + \int_0^t \sigma e^{\theta(s-t)} dW_s. \tag{1.36}$$

If one assumes that r_0 is Gaussian ($r_0 \overset{\mathcal{L}}{\sim} \mathcal{N}(m_0, \sigma_0^2)$) and is independent from W, the process $(r_t)_{t>0}$ is Gaussian. One has $\mathbb{E}[r_t] = m_0 e^{-\theta t} + \mu(1 - e^{-\theta t})$ and $\mathrm{cov}(r_s, r_t) = \frac{\sigma^2}{2\theta} e^{-\theta(s+t)} \left(e^{2\theta \min(s,t)} - 1\right) + \sigma_0^2 e^{-\theta(s+t)}$.
Moreover $\lim_{t\to\infty} \mathrm{Var}(r_t) = \frac{\sigma^2}{2\theta}$ (the long-term variance). If the initial variance σ_0^2 is equal to long-term variance $\frac{\sigma^2}{2\theta}$, the process is stationary and the covariance writes $\mathrm{cov}(r_s, r_t) = \frac{\sigma^2}{2\theta} e^{-\theta|s-t|}$.
The total variance of the process on $[0, T]$ is

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

1.A.2 The Ornstein-Uhlenbeck covariance operator

The Ornstein-Uhlenbeck covariance operator is given by

$$T^{OU} f(t) = \int_0^T \frac{\sigma^2}{2\theta} e^{-\theta(s+t)} \left(e^{2\theta \min(s,t)} - 1\right) f(s)ds + \int_0^T \sigma_0^2 e^{-\theta(s+t)} f(s)ds. \tag{1.37}$$

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

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

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

$$g'' - \theta^2 g = -\sigma^2 f, \tag{1.38}$$

with

$$\sigma_0^2 g'(0) = \left(\sigma^2 - \theta\sigma_0^2\right) g(0) \qquad \text{and} \qquad g'(T) = -\theta g(T). \tag{1.39}$$

Proof:

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

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

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

One gets $g''(t) = \theta^2 g(t) - \sigma^2 f(t)$. Moreover, Equation (1.39) comes when identifying expressions with $t = 0$ and $t = T$. □

Proposition 1.A.2. *Conversely, if $g \in C^2([0,T])$ and if functions f and g satisfy Equations (1.38) and (1.39) then $g = T^{OU} f$.*

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

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

An integration by parts yields

$$\begin{aligned} T^{OU} g'' &= -\sigma_0^2 g'(0)e^{-\theta t} - \sigma^2 g(t) + \tfrac{\sigma^2}{2} g(0)e^{-\theta t} - \left(\theta\sigma_0^2 - \tfrac{\sigma^2}{2}\right) g(0)e^{-\theta t} + \theta^2 T^{OU} g(t) \\ &= -\sigma^2 g(t) + \theta^2 T^{OU} g(t) \quad \text{thanks to (1.39).} \end{aligned}$$

□

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

$$\lambda g'' + (\sigma^2 - \lambda\theta^2) g = 0. \tag{1.40}$$

Hence the solution of the ordinary differential equation (1.40) on $[0,T]$ has the form $g(t) = A\cos(\omega t) + B\sin(\omega t)$, with $\omega = \sqrt{\frac{\sigma^2 - \lambda\theta^2}{\lambda}} \Leftrightarrow \lambda = \frac{\sigma^2}{\omega^2 + \theta^2}$.
Equation (1.39) yields $\omega B\sigma_0^2 = (\sigma^2 - \theta\sigma_0^2)A$. Hence, function $g(x)$ writes

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

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

$$\omega\sigma^2\cos(\omega T) + \left(-\omega^2\sigma_0^2 + \theta\sigma^2 - \theta^2\sigma_0^2\right)\sin(\omega T) = 0. \tag{1.41}$$

Conversely, by the same computation, one sees that $\lambda_n \in]0, \|T^{OU}\|_2]$ is an eigenvalue of T^{OU} if and only if Equation (1.41) is fulfilled.

Proposition 1.A.3. *Finally, if one knows the sorted increasing sequence (ω_n) of the strictly positive solutions of Equation (1.41), the Karhunen-Loève eigensystem $\left(e_n^{OU}, \lambda_n^{OU}\right)$ of the Ornstein-Uhlenbeck covariance operator T^{OU} are given by:*

- $\lambda_n^{OU} = \frac{\sigma^2}{\omega_n^2 + \theta^2}$, *and*
- $e_n^{OU}(t) = K_n\left(\omega_n\sigma_0^2\cos(\omega_n t) + (\sigma^2 - \theta\sigma_0^2)\sin(\omega_n t)\right)$, *where K_n is the normalization constant. If $(\sigma, \sigma_0) \neq (0,0)$, K_n is given by*

$$\begin{aligned} 1/K_n^2 = \frac{1}{2\omega_n}\sigma_0^2(\sigma^2 - \theta\sigma_0^2)\left(1 - \cos(2\omega_n T)\right) + \frac{1}{2}\sigma_0^4\omega_n^2\left(T + \frac{1}{2\omega_n}\sin(2\omega_n T)\right) \\ + \frac{1}{2}(\sigma^2 - \theta\sigma_0^2)^2\left(T - \frac{1}{2\omega_n}\sin(2\omega_n T)\right). \end{aligned} \tag{1.42}$$

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

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

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

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

where C_n is the normalization constant. C_n is given by

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

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

As we have seen in the previous section, everything comes to evaluate numerically the strictly positive solutions of Equation (1.41).

Deterministic start point

In this case, ($\sigma_0 = 0$), one can check that elements of $\left\{\frac{\pi}{2T} + k\frac{\pi}{T},\ k \in \mathbb{N}\right\}$ are not solutions of Equation (1.41), thus the equation comes to

$$\theta \tan(\omega T) = -\omega. \tag{1.43}$$

The case where $\theta = 0$ comes to the case of the Brownian motion, hence one assumes that $\theta \neq 0$. Solutions of this equation are illustrated in Figure 1.8. One can easily show that a unique solution ω_n lies in each interval $\left(\frac{n\pi}{T} - \frac{\pi}{2T}, \frac{n\pi}{T}\right)$, for $n \in \mathbb{N}^*$.

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

Non-deterministic start point

Let us assume now that $\sigma_0 \neq 0$ and consider Equation (1.41) again. The term $-\omega^2\sigma_0^2 + \theta\sigma^2 - \theta^2\sigma_0^2$ never vanishes on $(0, +\infty)$ if $\theta^2\sigma_0^2 - \theta\sigma^2 \geq 0$.
First case: $\theta^2\sigma_0^2 - \theta\sigma^2 \geq 0$**.**
Here, everything comes to the equation

$$\tan(\omega T) = \frac{\omega\sigma^2}{\omega^2\sigma_0^2 + \theta^2\sigma_0^2 - \theta\sigma^2}. \tag{1.44}$$

Solutions of this equation are illustrated in Figure 1.9.

We can easily show that $\forall n \in \mathbb{N}^*, \exists! \omega \in \left(\frac{n\pi}{T}, \frac{n\pi}{T} + \frac{\pi}{2T}\right)$ that is solution of Equation (1.41). Moreover a solution lies in $\left(0, \frac{\pi}{2T}\right)$ if and only if $(\theta^2\sigma_0^2 - \theta\sigma^2)T - \sigma^2 < 0$.
Second case: $\theta^2\sigma_0^2 - \theta\sigma^2 < 0$**.**
Here, the term $-\omega^2\sigma_0^2 + \theta\sigma^2 - \theta^2\sigma_0^2$ vanishes for $\omega = V := \sqrt{\theta\frac{\sigma^2}{\sigma_0^2} - \theta^2}$. If V is not a solution of Equation (1.41), (*i.e.* if V does not belong to $\left\{\frac{\pi}{2T} + k\frac{\pi}{T} | k \in \mathbb{N}\right\}$), no other value of this set is a solution, and everything comes again to the same Equation (1.44). Solutions of this equation are illustrated in Figure 1.10. We can then easily show that $\forall n \in \mathbb{N}^* \cap]V, +\infty[, \exists! \omega \in \left(\frac{n\pi}{T}, \frac{n\pi}{T} + \frac{\pi}{2T}\right)$, ω is solution of Equation (1.41). Moreover, in every nonempty interval $\left(\frac{k\pi}{T} - \frac{\pi}{2T}, \frac{k\pi}{T} + \frac{\pi}{2T}\right) \cap (0, V)$, $k \in \mathbb{N}^*$ there is another solution of the equation.

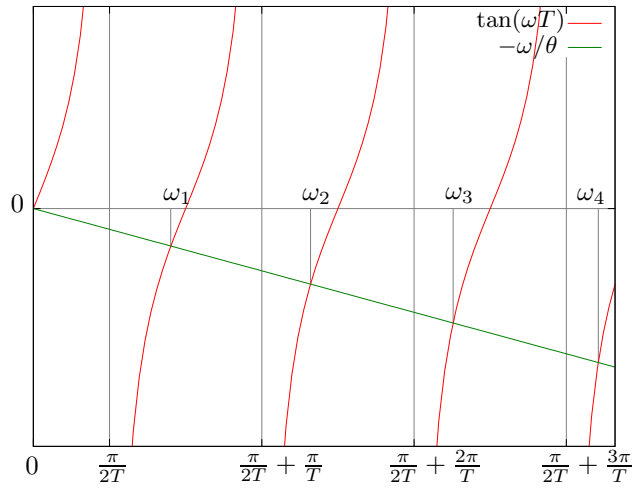

Figure 1.8: **(Deterministic start point)**. Solutions of Equation (1.43). (Ornstein-Uhlenbeck process starting from a determined point r_0, $\sigma_0 = 0$.) Numerical values for this figure are $T = 3$, $\sigma = 1$ and $\theta = 3$.

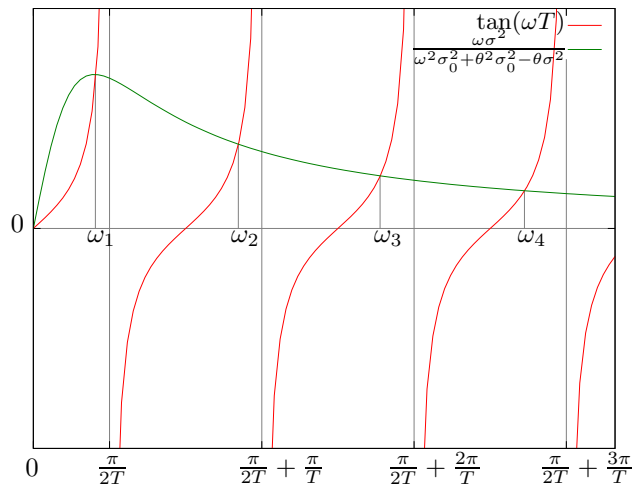

Figure 1.9: **(Non-deterministic start point, $\theta^2\sigma_0^2 - \theta\sigma^2 \geq 0$)**. Solutions of Equation (1.43). (Ornstein-Uhlenbeck process starting from $r_0 \overset{\mathcal{L}}{\sim} \mathcal{N}(0, \sigma_0^2)$, $\sigma_0 \neq 0$.) Numerical values for this figure are $T = 3$, $\sigma = 1$, $\theta = 3$ and $\sigma_0^2 = 0.4$.

A procedure for computing Ornstein-Uhlenbeck eigenvalues

A procedure for the computation of the n-th eigenvalue of the Ornstein-Uhlenbeck covariance operator is given in Algorithm 1.A.1. In this algorithm, the function **search**$(a, \text{left}, \text{right})$ stands for a root finding method. It fills argument a with the root of Equation (1.41) that is bracketed by $[\text{left}, \text{right}]$.

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

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

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

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

{There is a unique solution ω_n of (1.41) in the interval $\left(\frac{n\pi}{T} - \frac{\pi}{2T}, \frac{n\pi}{T}\right)$. }

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

else

{Here $\sigma_0 > 0$.}

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

{The vertical asymptote of the right-hand side of Equation 1.44 lies on the left of 0. }

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

{There is a unique solution ω_n of (1.41) in the interval $\left(0, \frac{\pi}{2T}\right)$. }

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

else

{The smallest strictly positive solution ω_1 of Equation (1.41) lies in the interval $\left(\frac{\pi}{2T}, \frac{\pi}{T}\right)$.}

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

end if

else

{The vertical asymptote of the right-hand side of Equation 1.44 lies on the right of 0. }

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

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

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

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

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

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

else

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

end if

end if

end if

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

Return λ_n.

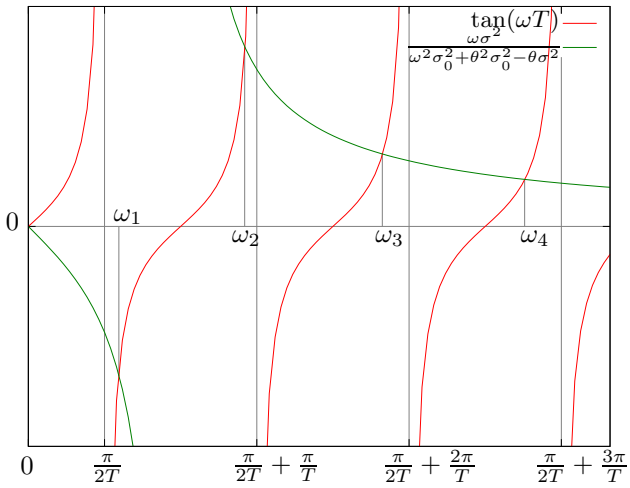

Figure 1.10: **(Non-deterministic start point, $\theta^2\sigma_0^2 - \theta\sigma^2 < 0$).** Solutions of Equation (1.43). (Ornstein-Uhlenbeck process starting from $r_0 \overset{\mathcal{L}}{\sim} \mathcal{N}(0, \sigma_0^2)$, $\sigma_0 \neq 0$.) Numerical values for this figure are $T = 3$, $\sigma = 1$, $\theta = 3$ and $\sigma_0^2 = 0.3$.

A numerical guess for the value of ω_n.

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

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

Now, in the case of the Ornstein-Uhlenbeck eigenvalue computation, for a deterministic start point, Equation (1.43) can be approximated by

$$\theta\psi(\omega_n T + n\pi) = -\omega_n \quad n \geq 1. \tag{1.45}$$

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

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

In this section, we use the same notations as in Section 1.4.2. We give the closed-form expression of the matrix $R_{Y|V} := ((\alpha_{ij}))_{1\leq i\leq d, 0\leq j\leq n} \in M_{d,n}(\mathbb{R})$ which corresponds to the affine function $Af_{Y|V}$ defined by $\mathbb{E}[Y|V] = Af_{Y|V}(V)$, in the cases of the standard Brownian motion the standard Brownian bridge and Ornstein-Uhlenbeck processes on $[0, T]$.

In the general case, this linear least-square minimization can be performed numerically, but this preliminary step can become time-consuming when the number of simulation dates grows. Thus, these closed-form expressions become important for this purpose.

Consider $t_0 = 0 \leq t_1 \leq \cdots \leq t_n = T$ a subdivision of $[0, T]$, and let us assume that X is one of

the above cited processes. Then X is a Markov process, and

$$\mathbb{E}\left[\int_0^T X_s e_i^X(s)ds \middle| X_{t_0}, \cdots, X_{t_n}\right] = \sum_{j=0}^{n-1} \underbrace{\mathbb{E}\left[\int_{t_j}^{t_{j+1}} X_s e_i^X(s)ds \middle| X_{t_j}, X_{t_{j+1}}\right]}_{=f_j^i(X_{t_j}, X_{t_{j+1}})}, \tag{1.46}$$

where f_j^i is an affine function.

1.B.1 The case of the standard Brownian motion

Now, assuming that $X = W$ is a standard Brownian motion on $[0,T]$, using Equation (1.46), we have, if $t_j \neq t_{j+1}$, $f_j^i(x,y) = \mathbb{E}\left[\int_{t_j}^{t_{j+1}} \left(x + \frac{s-t_j}{t_{j+1}-t_j}(y-x) + Y_{s-t_j}^{B,t_{j+1}-t_j}\right) e_i^W(s)ds\right]$, where $Y_{s-t_j}^{B,t_{j+1}-t_j}$ is a standard Brownian bridge on $[t_j, t_{j+1}]$. Hence,

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

Simple computations lead to

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

and

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

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

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

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

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

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

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

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

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

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

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

1.B.2 The case of the standard Brownian bridge

If $X = B$ is a standard Brownian bridge on $[0,T]$, Equation (1.46) yields, if $t_j \neq t_{j+1}$, $f_j^i(x,y) = \mathbb{E}\left[\int_{t_j}^{t_{j+1}} \left(x + \frac{s-t_j}{t_{j+1}-t_j}(y-x) + \left(Y_{s-t_j}^{B,t_{j+1}-t_j}\right)\right) e_i^B(s)ds\right]$, where $Y_{s-t_j}^{B,t_{j+1}-t_j}$ is a standard Brownian bridge on $[t_j, t_{j+1}]$. Hence, very similarly to the case of the standard Brownian motion,

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

Simple computations lead to:

$$\int_{t_j}^{t_{j+1}} e_i^B(s)ds = \sqrt{\frac{2}{T}}\frac{T}{\pi i}\left(\cos\left(\pi i \frac{t_j}{T}\right) - \cos\left(\pi i \frac{t_{j+1}}{T}\right)\right), \tag{1.47}$$

and

$$\int_{t_j}^{t_{j+1}} s e_i^B(s)ds = \sqrt{\frac{2}{T}}\frac{T}{\pi i}\left(t_j \cos\left(\pi i \frac{t_j}{T}\right) - t_{j+1}\cos\left(\pi i \frac{t_{j+1}}{T}\right)\right) + \sqrt{\frac{2}{T}}\left(\frac{T}{\pi i}\right)^2\left(\sin\left(\pi i \frac{t_{j+1}}{T}\right) - \sin\left(\pi i \frac{t_j}{T}\right)\right). \tag{1.48}$$

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

$$A_j^i = \lambda_i^B\left(\left(e_i^B\right)'(t_j) - \frac{e_i^B(t_{j+1}) - e_i^B(t_j)}{t_{j+1}-t_j}\right), \tag{1.49}$$

$$B_j^i = \lambda_i^B\left(\frac{e_i^B(t_{j+1}) - e_i^B(t_j)}{t_{j+1}-t_j} - \left(e_i^B\right)'(t_{j+1})\right). \tag{1.50}$$

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

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

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

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

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

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

- $\alpha_{i0} = A_0^i = \begin{cases} \lambda_i^B\left(\left(e_i^B\right)'(t_0) - \frac{e_i^B(t_1)-e_i^B(t_0)}{t_1-t_0}\right) & \text{if } t_1 \neq t_0, \\ 0 \text{ in the other case.} \end{cases}$

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

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

Remark. *A noticeable fact is that we obtain exactly the same expression as for the Brownian motion, where (e_n^W, λ_n^W) is replaced by (e_n^B, λ_n^B).*

1.B.3 The case of centered Ornstein-Uhlenbeck processes

We assume here that $X = r$ is an Ornstein-Uhlenbeck process defined on $[0,T]$ by the SDE $dr_t = -\theta r_t dt + \sigma dW_t$, with $\theta > 0$, $\sigma \geq 0$, W a standard Brownian motion on $[0,T]$ and $r_0 \overset{\mathcal{L}}{\sim} \mathcal{N}(0,\sigma_0^2)$ independent of W. Consider $t_0 = 0 \leq t_1 \leq \cdots \leq t_n = T$ a subdivision of $[0,T]$. Considering Equation (1.46) and using the conditional Fubini theorem, we obtain

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

Now, we easily prove that

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

Hence

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

where $\left(e_n^{OU}\right)_{n\geq 1}$ are the Karhunen-Loève eigenfunctions of the considered Ornstein-Uhlenbeck process.

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

As previously, the equality cases will be handled. It is useful in numerical applications, when dealing with small time steps that make the numerical evaluation of $e_i^{OU}(t_{j+1}) - e_i^{OU}(t_j)$ inaccurate.

The Ornstein-Uhlenbeck process starting from 0

In this case ($\sigma_0^2 = 0$), as proved in Appendix 1.A, the Karhunen-Loève eigensystem is given by

$$e_n^{OU}(t) := \left(\frac{1}{\sqrt{\frac{T}{2} - \frac{\sin(2\omega_n T)}{4\omega_n}}}\right) \sin(\omega_n t), \qquad \lambda_n^{OU} := \frac{\sigma^2}{\omega_n^2 + \theta^2}, \quad n \geq 1, \tag{1.51}$$

where ω_n are the (sorted) strictly positive solutions of the equation $\theta \sin(\omega_n T) + \omega_n \cos(\omega_n T) = 0$. For $K \in \mathbb{R}$, $\omega \in \mathbb{R}^*$ and $(t_a, t_b) \in \mathbb{R}^2$, we have

$$\int_{t_a}^{t_b} \exp(Ks)\sin(\omega s)ds = \frac{K}{K^2+\omega^2}\left(e^{Kt_b}\sin(\omega t_b) - e^{Kt_a}\sin(\omega t_a)\right) - \frac{\omega}{K^2+\omega^2}\left(e^{Kt_b}\cos(\omega t_b) - e^{Kt_a}\cos(\omega t_a)\right).$$

Using this formula with $\omega = \omega_i$, and multiplying by $\frac{1}{\sqrt{\frac{T}{2} - \frac{\sin(2\omega_i T)}{4\omega_i}}}$, we obtain

$$\int_{t_a}^{t_b} \exp(Ks)e_i^{OU}(s)ds = \frac{K}{K^2+\omega^2}\left(e^{Kt_b}e_i^{OU}(t_b) - e^{Kt_a}e_i^{OU}(t_a)\right) - \frac{1}{K^2+\omega^2}\left(e^{Kt_b}\left(e_i^{OU}\right)'(t_b) - e^{Kt_a}\left(e_i^{OU}\right)'(t_a)\right).$$

This yields

$$\begin{aligned}\int_{t_j}^{t_{j+1}} e^{\theta(t_{j+1}-s)}e_i^{OU}(s)ds = \frac{-\theta}{\theta^2+\omega_i^2}&\left(e_i^{OU}(t_{j+1}) - e^{\theta(t_{j+1}-t_j)}e_i^{OU}(t_j)\right)\\ &- \frac{1}{\theta^2+\omega_i^2}\left(\left(e_i^{OU}\right)'(t_{j+1}) - e^{\theta(t_{j+1}-t_j)}\left(e_i^{OU}\right)'(t_j)\right).\end{aligned} \tag{1.52}$$

$$\int_{t_j}^{t_{j+1}} e^{-\theta(t_{j+1}-s)} e_i^{OU}(s)ds = \frac{\theta}{\theta^2+\omega_i^2}\left(e_i^{OU}(t_{j+1}) - e^{-\theta(t_{j+1}-t_j)} e_i^{OU}(t_j)\right) - \frac{1}{\theta^2+\omega_i^2}\left(\left(e_i^{OU}\right)'(t_{j+1}) - e^{-\theta(t_{j+1}-t_j)}\left(e_i^{OU}\right)'(t_j)\right). \quad (1.53)$$

$$\int_{t_j}^{t_{j+1}} e^{\theta(s-t_j)} e_i^{OU}(s)ds = \frac{\theta}{\theta^2+\omega_i^2}\left(e^{\theta(t_{j+1}-t_j)} e_i^{OU}(t_{j+1}) - e_i^{OU}(t_j)\right) - \frac{1}{\theta^2+\omega_i^2}\left(e^{\theta(t_{j+1}-t_j)}\left(e_i^{OU}\right)'(t_{j+1}) - \left(e_i^{OU}\right)'(t_j)\right). \quad (1.54)$$

$$\int_{t_j}^{t_{j+1}} e^{-\theta(s-t_j)} e_i^{OU}(s)ds = \frac{-\theta}{\theta^2+\omega_i^2}\left(e^{-\theta(t_{j+1}-t_j)} e_i^{OU}(t_{j+1}) - e_i^{OU}(t_j)\right) - \frac{1}{\theta^2+\omega_i^2}\left(e^{-\theta(t_{j+1}-t_j)}\left(e_i^{OU}\right)'(t_{j+1}) - \left(e_i^{OU}\right)'(t_j)\right). \quad (1.55)$$

Finally, we have

$$A_j^i = \frac{-\theta}{\theta^2+\omega_i^2}\left(\frac{2e_i^{OU}(t_{j+1})}{e^{\theta(t_{j+1}-t_j)} - e^{-\theta(t_{j+1}-t_j)}} - e_i^{OU}(t_j)\frac{e^{\theta(t_{j+1}-t_j)} + e^{-\theta(t_{j+1}-t_j)}}{e^{\theta(t_{j+1}-t_j)} - e^{-\theta(t_{j+1}-t_j)}}\right) + \frac{1}{\theta^2+\omega_i^2}\left(e_i^{OU}\right)'(t_j), \quad (1.56)$$

and

$$B_j^i = \frac{\theta}{\theta^2+\omega_i^2}\left(e_i^{OU}(t_{j+1})\frac{e^{\theta(t_{j+1}-t_j)} + e^{-\theta(t_{j+1}-t_j)}}{e^{\theta(t_{j+1}-t_j)} - e^{-\theta(t_{j+1}-t_j)}} - \frac{2e_i^{OU}(t_j)}{e^{\theta(t_{j+1}-t_j)} - e^{-\theta(t_{j+1}-t_j)}}\right) - \frac{1}{\theta^2+\omega_i^2}\left(e_i^{OU}\right)'(t_{j+1}). \quad (1.57)$$

We recompose the coefficients $((\alpha_{ij}))_{1\le i\le d, 0\le j\le n}$ of the regression matrix.

- For every $1 \le j < n$, $\alpha_{ij} = A_j^i + B_{j-1}^i$. The terms involving $\left(e_i^{OU}\right)'$ vanishes.
- $\alpha_{i0} = A_0^i$ and $\alpha_{in} = B_{n-1}^i$.
- What equality cases concerns, we can easily prove that $\lim_{t_{j+1}\to t_j} A_j^i = 0$ and $\lim_{t_{j-1}\to t_j} B_{j-1}^i = 0$ and deduce the corresponding formula when some dates in the schedule are equal.

The general Ornstein-Uhlenbeck process

In this case ($r_0 \overset{\mathcal{L}}{\sim} \mathcal{N}(0,\sigma_0^2)$ with $\sigma_0^2 > 0$), as proved in Appendix 1.A, the Karhunen-Loève eigensystem is given by

$$e_n^{OU}(t) := K_n\left(\omega_n\sigma_0^2\cos(\omega_n t) + \left(\sigma^2 - \theta\sigma_0^2\right)\sin(\omega_n t)\right), \qquad \lambda_n^{OU} := \frac{\sigma^2}{\omega_n^2+\theta^2}, \quad n \ge 1, \quad (1.58)$$

where ω_n are the (sorted) strictly positive solutions of the equation

$$\omega_n\sigma^2\cos(\omega_n T) + (\theta\sigma^2 - \theta^2\sigma_0^2 - \omega_n^2\sigma_0^2)\sin(\omega_n T) = 0,$$

and

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

We can factorize Equation (1.58) and write

$$e_n^{OU}(t) := K_n\sqrt{\omega_n^2\sigma_0^4 + (\sigma^2 - \theta\sigma_0^2)^2}\sin(\omega_n t + \phi_n), \quad \text{with } \phi_n = \arccos\left(\frac{\sigma^2 - \theta\sigma_0^2}{\sqrt{\omega_n^2\sigma_0^4 + (\sigma^2 - \theta\sigma_0^2)^2}}\right)$$

and $\lambda_n^{OU} := \frac{\sigma^2}{\omega_n^2+\theta^2}$, $n \geq 1$. Using that for $K \in \mathbb{R}$, $\omega \in \mathbb{R}^*$ and $(t_a, t_b) \in \mathbb{R}^2$,

$$\int_{t_a}^{t_b} \exp(Ks)\sin(\omega s + \phi)ds = \frac{K}{K^2+\omega^2}\left(e^{Kt_b}\sin(\omega t_b + \phi) - e^{Kt_a}\sin(\omega t_a + \phi)\right) - \frac{\omega}{K^2+\omega^2}\left(e^{Kt_b}\cos(\omega t_b + \phi) - e^{Kt_a}\cos(\omega t_a + \phi)\right), \quad (1.59)$$

we see that the expressions for $((\alpha_{ij}))_{1\leq i\leq d, 0\leq j\leq n}$ established in Section 1.B.3 remain valid in this case.

Bibliography

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

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

[3] Richard P. Brent. Algorithms for minimization without derivatives. *(Englewood Cliffs, NJ: Prentice-Hall); reprinted 2002 (New York: Dover), Chapters 3, 4.[1]*, 1973.

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

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

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

[7] Sylvain Corlay. Partial functional quantization and generalized bridges. *Preprint*, 2011.

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

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

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

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

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

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

[14] Ying-Lin Hsu, Tsung-I Lin, and Cheng-Few Lee. Constant Elasticity of Variance (CEV) option pricing model: Integration and detailed derivation. *Mathematics and Computers in Simulation*, 79(1):60 – 71, 2008.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[31] Panayotis Tsaparas. Nearest-neighbor search in multidimensional spaces. *Qualifying Depth Oral Report 319-02, Dept. of Computer Science, University of Toronto, 1999. 2*, 1999.

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

Chapter 2

The Nyström method for functional quantization with an application to the fractional Brownian motion

Abstract

It is recognized that the constructive quadratic optimal functional quantization of a Gaussian process requires the numerical evaluation of its Karhunen-Loève eigensystem (see [15]). Closed-form expressions are available for several processes such as the Brownian motion, the Brownian bridge and Ornstein-Uhlenbeck processes, but not in the general case. For example, the Karhunen-Loève decomposition of the fractional Brownian motion is not known.

In this chapter, the so-called "Nyström method" is tested to compute optimal quantizers of Gaussian processes. In particular, we derive the optimal quantization of the fractional Brownian motion by approximating the first terms of its Karhunen-Loève decomposition.

A numerical test of the "functional stratification" variance reduction algorithm is performed with the fractional Brownian motion.

Keywords: integral equation, Nyström method, functional quantization, vector quantization, Karhunen-Loève, Gaussian process, Brownian motion, Brownian bridge, Ornstein-Uhlenbeck, fractional Brownian motion, numerical integration, optimal quantization, product quantization, variance reduction, stratification.

Introduction

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be probability space, and E a reflexive separable Banach space. The norm on E is denoted by $|\cdot|$.

The quantization of a random variable X, taking its values in E consists in its approximation by a random variable Y taking finitely many values. The resulting error of this discretization is the L^p norm of $|X-Y|$. If we settle on a fixed maximum cardinal N for $Y(\Omega)$, the minimization of the error comes to the following optimization problem:

$$\min\left\{\||X-Y|\|_p\,,\ Y:\Omega\to E \text{ measurable},\ \operatorname{card}(Y(\Omega))\le N\right\}. \tag{2.1}$$

A solution of (2.1) is an optimal quantizer of X. This problem was first investigated for signal transmission and compression issues. More recently, quantization has been introduced in numerical probability to devise quadrature schemes [18], solving multidimensional stochastic control problems [2] and for variance reduction [4]. Since the 2000's, the infinite-dimensional setting has been investigated from both theoretical and numerical viewpoints, especially in the quadratic case [15] but also in other Banach spaces [27]. One elementary property of a L^2 optimal quantizer is the stationarity: $\mathbb{E}[X|Y]=Y$.

We now assume that X is a bi-measurable stochastic process on $[0,T]$ verifying $\int_0^T \mathbb{E}\left[|X_t|^2\right]dt<\infty$ so that it can be considered as a random variable valued in the Hilbert space $H=L^2([0,T])$. We assume that its covariance function Γ^X is continuous. In the seminal article on Gaussian functional quantization [15], it is shown that in the centered Gaussian case, linear subspaces U of H spanned by N-stationary quantizers correspond to principal components of X. In other words, they are spanned by eigenvectors of the covariance operator of X. Thus, the quadratic optimal quantization of Gaussian processes involves its Karhunen-Loève decomposition $\left(e_n^X,\lambda_n^X\right)_{n\ge1}$.
To perform optimal quantization, the Karhunen-Loève expansion is first truncated at a fixed order m and then the $\mathbb{R}^m$-valued Gaussian vector, constituted of the m first coordinates of the process on its Karhunen-Loève decomposition, is quantized. To reach optimal quantization, we have to determine both the optimal rank of truncation $d^X(N)$ (the quantization dimension) and the optimal $d^X(N)$-dimensional Gaussian quantizer corresponding to the first coordinates, $\bigotimes_{j=1}^{d^X(N)}\mathcal{N}\left(0,\lambda_j^X\right)$.
Usual examples of such processes are the standard Brownian motion on $[0,T]$, the standard Brownian bridge on $[0,T]$, the fractional Brownian motion and Ornstein-Uhlenbeck processes.

Another possibility is to use a product quantization of the distribution $\bigotimes_{j=1}^{m}\mathcal{N}\left(0,\lambda_j^X\right)$. The product quantization is the Cartesian product of the optimal quantizers of the standard one-dimensional Gaussian distributions $\mathcal{N}\left(0,\lambda_i^X\right)_{1\le i\le d^X(N)}$. In the case of independent marginals, this yields a stationary quantizer, *i.e.* a quantizer Y of X which satisfies $\mathbb{E}[X|Y]=Y$.This property, shared with optimal quantizers, results in a convergence rate of a higher order for the quantization-based cubature method, as we can see in [21]. One advantage of this setting is that the one-dimensional Gaussian quantization is a fast procedure. In [20], deterministic optimization methods (as Newton-Raphson) are shown to converge rapidly to the unique optimal quantizer of the one-dimensional Gaussian distribution. Moreover, a sharply optimized database of quantizers of standard univariate and multivariate Gaussian distributions is available on the web site `www.quantize.maths-fi.com` [22] for download. Still, we have to determine the quantization level for each dimension to obtain optimal product quantization. In this case, the minimization of the distortion comes to:

$$\min\left\{\sum_{j=1}^{d}\mathcal{E}_{N_j}^2\left(\mathcal{N}\left(0,\lambda_j^X\right)\right)+\sum_{j\ge d+1}\lambda_j^X, N_1\times\cdots\times N_d\le N, d\ge1\right\}. \tag{2.2}$$

A solution of (2.2) is called an optimal K-L product quantizer. This problem can be solved by the "blind optimization procedure", which consists in computing the criterion for every possible

decomposition $N_1 \times \cdots \times N_d$ with $N_1 \geq \cdots \geq N_d$. The result of this procedure can be kept off-line for a future use. Optimal decompositions for a wide range of values of N for both Brownian motion and Brownian bridge are available on the web site `www.quantize.maths-fi.com` [22]. Another fact on quadratic functional product quantization is that it is shown to be rate-optimal.

In [16], the rate of convergence to zero of the quantization error is investigated. A complete solution is provided for the case of Gaussian processes under rather general conditions on the eigenvalues of the covariance operator. Rates of convergence are available for the above cited examples of Gaussian processes. The asymptotics of the quantization dimension $d^X(N)$ are investigated. The following theorem combines these results:

Theorem 2.0.1 (Functional quantization asymptotics)**.** *Let X be a centered bi-measurable Gaussian process on $[0,T]$ with a continuous covariance function. Let us denote by $\left(e_n^X, \lambda_n^X\right)_{n\geq 1}$ its Karhunen-Loève eigensystem. Let $(Y_N)_{N\geq 1}$ be a sequence of quadratic optimal N-quantizers for X. We assume that*

$$\lambda_n^X \sim \frac{\kappa}{n^b} \text{ as } n \to \infty \qquad (b > 1).$$

We have:

- $\operatorname{span}(Y_N(\Omega)) = \operatorname{span}\left\{e_1^X, \cdots, e_{d^X(N)}^X\right\}$ *and* $d^X(N) \gtrsim 2b^{-\frac{b}{b-1}} \log(N)$ *as* $N \to \infty$.
- $\mathcal{E}_N(X) = \|X - Y_N\|_2 \sim \sqrt{\kappa}\sqrt{b^b(b-1)^{-1}}(2\ln N)^{-\frac{b-1}{2}}$ *as* $N \to \infty$.

It is shown in [15] that the Karhunen-Loève eigenvalues of the fractional Brownian motion, $\left(\lambda_n^{B^H}\right)_{n\geq 1}$ verify

$$\lambda_n^{B^H} \sim \frac{\nu_H}{n^{2H+1}} \text{ as } n \to \infty,$$

for some positive constant ν_H. Thus the fractional Brownian motion satisfies the hypothesis of Theorem 2.0.1.

From a constructive viewpoint, the numerical computation of the optimal quantization or the optimal product quantization requires a numerical evaluation of the Karhunen-Loève eigenfunctions and eigenvalues, at least the very first terms. (As seen in Theorem 2.0.1, the quantization dimension of usual Gaussian processes increases asymptotically as the logarithm of the size of the quantizer, so that it is most likely that it is small. For instance, the quantization dimension $d^W(N)$ of the Brownian motion with $N = 10000$ is 9.) The Karhunen-Loève decompositions of usual Gaussian processes have closed-form expressions. It is the case for the standard Brownian motion, the Brownian bridge and Ornstein-Uhlenbeck processes. (The case of Ornstein-Uhlenbeck processes is detailed in [4]).

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

$$e_n^W(t) := \sqrt{\frac{2}{T}} \sin\left(\pi(n-1/2)\frac{t}{T}\right), \qquad \lambda_n^W := \left(\frac{T}{\pi(n-1/2)}\right)^2, \quad n \geq 1. \tag{2.3}$$

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

$$e_n^B(t) := \sqrt{\frac{2}{T}} \sin\left(\pi n\frac{t}{T}\right), \qquad \lambda_n^B := \left(\frac{T}{\pi n}\right)^2, \quad n \geq 1. \tag{2.4}$$

3. The Ornstein-Uhlenbeck process on $[0,T]$, starting from 0, defined by the SDE $dr_t = \theta(mu - r_t)dt + \sigma dW_t$, with $\sigma \geq 0$, $\theta > 0$ and W a standard Brownian motion on $[0,T]$, (see [4]).

$$e_n^{OU}(t) := \left(\frac{1}{\sqrt{\frac{T}{2} - \frac{\sin(2\omega_n T)}{4\omega_n}}}\right) \sin(\omega_n t), \qquad \lambda_n^{OU} := \frac{\sigma^2}{\omega_n^2 + \theta^2}, \quad n \geq 1, \tag{2.5}$$

where ω_n are the (sorted) strictly positive solutions of the equation

$$\theta \sin(\omega_n T) + \omega_n \cos(\omega_n T) = 0.$$

4. The stationary Ornstein-Uhlenbeck process on $[0,T]$, defined by the same SDE with $r_0 \overset{\mathcal{L}}{\sim} \mathcal{N}\left(0, \frac{\sigma^2}{2\theta}\right)$, (see [4]).

$$e_n^{OU}(t) := C_n \left(\omega_n \cos(\omega_n t) + \theta \sin(\omega_n t)\right), \qquad \lambda_n^{OU} := \frac{\sigma^2}{\omega_n^2 + \theta^2}, \quad n \geq 1, \tag{2.6}$$

where ω_n are the (sorted) strictly positive solutions of the equation

$$2\theta\omega_n \cos(\omega_n T) + \left(\theta^2 - \omega_n^2\right) \sin(\omega_n T) = 0,$$

and

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

In [4], the general setting of an arbitrary initial variance σ_0 for the Ornstein-Uhlenbeck process is handled. A procedure for the computation of ω_n is also provided. Other examples of explicit Karhunen-Loève expansions are available in [5] and [25]. In [11], Istas derived a semi closed-form expression for the Karhunen-Loève expansion of the spherical fractional Brownian motion.

In a more general setting, we do not have a closed-form expression for the Karhunen-Loève decomposition. For instance, as far as we know, the K-L expansion of the fractional Brownian motion is not known. Hence, a numerical method to evaluate first Karhunen-Loève eigenfunctions is the "missing link" on the path to the constructive optimal quantization of more Gaussian processes.

However, we can derive rate-optimal quantization of Gaussian processes using other series expansions as proposed by Luschgy and Pagès in [17, 19]. In this setting, the case of the fractional Brownian motion can be derived using a rate-optimal series expansion established by Dzhaparidze and van Zanten in [8, 9] as done by Junglen and Luschgy in [14]. Admissible series expansions for this approach can also be derived with the method proposed by Jaimez and Valderrama in [12]. In their article, they consider the transformation $\phi : (X_t)_{t\in[a,b]} \mapsto (f(t)X_{\tau(t)})_{t\in[\alpha,\beta]}$, where $\tau : [\alpha,\beta] \to \mathbb{R}$ is a strictly increasing continuous function such that $\tau(\alpha) = a$ and $\tau(\beta) = b$ and f is a continuous strictly positive function on $[\alpha,\beta]$. They prove that the Karhunen-Loève expansion of $(X_t)_{t\in[0,T]}$ with respect to Lebesgue's measure is transformed into the Karhunen-Loève expansion of $(f(t)X_{\tau(t)})_{t\in}$ with respect to the measure $f(t)^{-2}d\tau(t)$. Other constructive approaches for functional quantization are proposed by Wilbertz in [27].

Here, we experiment with the so-called "Nyström method" [1, 6, 24] for approximating the solution of the functional eigenvalue problem which defines the Karhunen-Loève decomposition. First, we compare the result of the numerical method with the closed-form expressions available for the Brownian motion, the Brownian bridge and Ornstein-Uhlenbeck processes. Then we handle the special case of the functional quantization of the fractional Brownian motion.

Functional quantization of Gaussian processes have numerous applications in numerical probability. In [4], a variance reduction method based on the functional quantization of a Gaussian process was proposed. This method can be seen as a "guided Monte-Carlo simulation" (see Figure 2.3). Still, it was only applicable with Gaussian processes for which we could have a numerical evaluation of the Karhunen-Loève eigenfunctions. Such a variance reduction method would be of high interest in Monte-Carlo simulations implying the fractional Brownian motion because its simulation schemes have a high complexity.

Subsequently, we test this "functional stratification" variance reduction algorithm with option pricing problems within the context of the counterpart of the classical Black and Scholes model for the fractional Brownian motion. First, the case of a vanilla option is benchmarked with the closed-form expression available in this case. Then the case of discrete barrier options is tested.

2.1 The Nyström method

Let X be a bi-measurable Gaussian stochastic process on $[0,T]$ defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We assume that $\int_{[0,T]} \mathbb{E}[X_s^2]ds < \infty$. Let us denote by Γ^X the covariance function of X

defined by $\Gamma^X(t,s) = \mathrm{cov}(X_t, X_s)$. We assume that Γ^X is a continuous function on $[0,T]\times[0,T]$[1]. The covariance operator C_X of X is defined by $C_X f = \int_{[0,T]} \Gamma^X(\cdot,s)f(s)ds$. It is a symmetric positive trace-class operator on $L^2([0,T])$. The Karhunen-Loève basis associated with X, denoted by $(e_n^X)_{n\geq 1}$ is the Hilbert basis of $L^2([0,T])$ constituted of the eigenvectors of C_X indexed following the decreasing order of eigenvalues. Now, we are led to solve numerically the eigenvalue problem

$$\int_0^T \Gamma^X(\cdot,s)f_k(s)ds = \lambda_k f_k, \qquad k \geq 1, \tag{2.7}$$

where both the eigenvalues and the eigenvectors have to be determined. The Nyström method relies on the choice of a quadrature rule $\int_0^T f(s)ds \approx \sum_{i=1}^n w_j f(s_j)$, where $(w_j)_{1\leq j\leq n}$ is the sequence of weights of the quadrature rule and $(s_j)_{1\leq j\leq n}$ are the abscissas at which f is evaluated. If we plug this quadrature rule in Equation (2.7), we get

$$\sum_{j=1}^n w_j \Gamma^X(t,s_j)f_k(s_j) \approx \lambda_k f_k(t) \qquad t\in[0,T]. \tag{2.8}$$

Evaluating Equation (2.8) at the quadrature points yields the approximate eigenvalue problem

$$\sum_{j=1}^n w_j \Gamma^X(s_i,s_j)f_k(s_j) = \lambda_k f_k(s_i) \qquad 1\leq i\leq n. \tag{2.9}$$

Let f denote the vector $\begin{pmatrix} f_k(s_1) \\ \vdots \\ f_k(s_n)\end{pmatrix}$, let $(K_{ij})_{1\leq i,j\leq n}$ be the matrix $\left(\Gamma^X(s_i,s_j)\right)_{1\leq i,j\leq n}$. We define the diagonal matrices λ and D by $\lambda := (\mathrm{diag}(\lambda_k))_{1\leq k\leq n}$ and $D := \mathrm{diag}(w_k)_{1\leq k\leq n}$. Then Equation (2.9) becomes

$$KDf = \lambda f. \tag{2.10}$$

Therefore, within this approximation, the functional eigenvalue problem turns into a matrix eigenvalue problem. As K is a covariance matrix, it is symmetric. However, since the weights are not equal for most quadrature rules, the matrix KD is not symmetric. As mentioned in [24], numerical methods for matrix orthogonalization are much simpler in the symmetric case. As a consequence, we should restore the symmetry if possible, or favor uniformly weighted quadrature rules. The method proposed in [24] to restore symmetry is the following:

Multiplying Equation (2.10) by $D^{1/2} = \mathrm{diag}(\sqrt{\omega_i})_{1\leq i\leq n}$ on the left, we get

$$\left(D^{1/2}KD^{1/2}\right)h = \lambda h, \quad \text{where} \quad h = D^{1/2}f. \tag{2.11}$$

Equation (2.11) is now in the form of a symmetric eigenvalue problem. In our framework (square-integrable kernels), this provides a good approximation of the n largest eigenvalues and the associated eigenfunctions.

2.1.1 Choice of the quadrature method

Classical numerical methods for real-valued symmetric matrix diagonalization are

- The Jacobi transform for symmetric diagonalization.
- A tridiagonalization (by Givens or Householder reduction) followed by a QL algorithm with implicit shifts.

[1] In the case where X is assumed to be pathwise continuous on $[0,T]$, Fernique's theorem ensures that $\int_{[0,T]} \mathbb{E}[X_s^2]ds < \infty$ (see *e.g.* [13]), and Γ^X is also continuous (see [13, VIII.3]).

All these numerical methods have a $O(n^3)$ complexity. As a consequence, the natural choice for the quadrature method would be the highest order to keep n as small as possible (as a Gaussian quadrature method).

However we will see, that the Nyström method associated with lower order quadrature rules may admit an asymptotic error expansion in even powers of the step sizes as soon as the covariance function is differentiable (or continuous and piecewise differentiable). So is the case for the trapezoidal quadrature rule and the generalized midpoint quadrature rule. As a consequence, instead of using the high order integration rule, we prefer to use a Richardson-Romberg extrapolation on the result of the whole procedure with the trapezoidal quadrature formula or the midpoint quadrature formula.

We could reach an accuracy which approaches the machine roundoff error on the first eigenvalues when we benchmark this method on the Brownian motion, the Brownian bridge or Ornstein-Uhlenbeck processes.

An argument in favor of the midpoint rule is that it is an equiweighted quadrature rule, so that the Nyström method comes to a symmetric matrix eigenvalue problem. Moreover, the quadratic n-optimal codebook for the uniform distribution on a real interval $[a,b]$ is $\left(a+\left(i-\frac{1}{2}\right)\frac{b-a}{n}\right)_{1\leq i\leq n}$ so that the quantization-based quadrature rule coincides with the midpoint rule.

2.1.2 Choice of the interpolation method

The natural choice is to use Equation (2.8) as an interpolation method for evaluating f_k,

$$f_k(t) = \frac{1}{\lambda_k}\sum_{j=1}^{n} w_j \Gamma^X(t,s_j) f_k(s_j). \tag{2.12}$$

The same Richardson-Romberg extrapolation can be performed between values of $\sum_{j=1}^{n} w_j \Gamma^X(t,s_j) f_k(s_j)$ with different orders n to compute this integral. The result is then divided by the extrapolated value of λ_k.

A remark on the interpolation method

One purpose of the quantization of a Gaussian process X, is to perform a quantization of the solution of a SDE driven by X, as soon as the corresponding stochastic integral can be defined. We can obtain a quantizer of the diffusion by inserting the quantizer of the Gaussian process in the diffusion equation written in the Stratonovich sense. The most accomplished study on this subject is [23]. This work is mostly specific to the Brownian motion but main results remain valid for continuous semi-martingales that satisfy the Kolmogorov criterion such as the Brownian bridge and Ornstein-Uhlenbeck processes.

Still, a future work could be to extend these results to solution of SDE driven by the fractional Brownian motion and other related processes. In this case, we may also need a numerical approximation of the time-derivative of the eigenfunction in the Karhunen-Loève decomposition. If Γ^X is (weakly) differentiable, a natural evaluation method for the derivative would be $f_k'(t) = \frac{1}{\lambda_k}\sum_{j=1}^{n} w_j \partial_t \Gamma^X(t,s_j) f_k(s_j)$.

One problem is that this method yields an irregular approximation of the derivative. For example, this yields a piecewise constant derivative in the case of the Brownian motion. This causes numerical instabilities when using Runge-Kutta integration methods for ordinary differential equations, which rely on the regularity of the considered Cauchy problem.

As a consequence, a more regular interpolation method can give more satisfactory results when dealing with diffusions. (Spline or rational interpolation methods for instance.)

2.2 Benchmark on known Karhunen-Loève expansions

In this section, we compare the numerical results obtained with the Nyström methods in cases for which we have a closed-form expressions of the Karhunen-Loève expansion. The multi-steps Richardson-Romberg extrapolation consists in using the asymptotic error estimate of the method

$$V = u_n + \frac{K_1}{n^2} + \frac{K_2}{n^4} + \cdots + O\left(\frac{1}{n^{2p}}\right).$$

Writing this expression for p different values of n allows us to solve a $p \times p$ linear system to nullify the $p-1$ first orders of convergence. The three-steps Richardson-Romberg extrapolation with $n = k$, $n = l$ and $n = m$ gives the following solution:

$$\frac{U_k k^4(m^2 - l^2) + U_l l^4(k^2 - m^2) + U_m m^4(l^2 - k^2)}{(m^2 - l^2)(l^2 m^2 + k^4 - m^2 k^2 - l^2 k^2)}.$$

This result is naturally invariant by any permutation of the coefficients (k, m, l). We observed less accurate results when using higher-order Richardson-Romberg extrapolation, so that we settled on a three-steps extrapolation which seems to be a good compromise.

2.2.1 Eigenvalues accuracy

Tables 2.1 and 2.2 report the Karhunen-Loève eigenvalues of the Brownian motion and of the Brownian bridge on $[0, 1]$. Table 2.3 deals with the stationary Ornstein-Uhlenbeck process on $[0, 1]$ defined by the SDE

$$dr_t = -r_t dt + dW_t, \qquad r_0 \overset{\mathcal{L}}{\sim} \mathcal{N}\left(0, \frac{1}{2}\right). \tag{2.13}$$

The first column gives the theoretical value given by the closed-form expression. Following columns give the value computed with the Nyström method with a regular step size with 25, 50 and 100 points. The last column gives the relative error of a three-steps Richardson-Romberg extrapolation method between $n = 25$, $n = 50$ and $n = 100$.

Closed-form	Mid-point Nyström 25 points	Mid-point Nyström 50 points	Mid-point Nyström 100 points	Mid-point Nyström $25 - 50 - 100$ Richardson-Romberg relative error
0.405284735	0.405418094	0.405318070	0.4052930680	1.5984×10^{-13}
0.0450316372	0.0451652077	0.0450649853	0.0450399714	1.1607×10^{-10}
0.0162113894	0.0163453833	0.0162447639	0.0162197259	2.4950×10^{-9}
0.00827111703	0.00840574996	0.00830453112	0.00827945541	1.8869×10^{-8}
0.00500351524	0.00513900777	0.00503698224	0.00501185691	8.5733×10^{-8}

Table 2.1: Record of the first five eigenvalues of the Karhunen-Loève decomposition of the Brownian motion on $[0, 1]$.

With regard to the above numerical results, the Nyström method gives a satisfactory accuracy for performing functional quantization of these processes.

2.2.2 Eigenfunctions accuracy

We now compare the closed-form expression of the eigenfunction with the approximation obtained by "Richardson-Romberg extrapolated mid-point Nyström method". In Table 2.4, we report the ratio between highest absolute difference between the closed-form expression and the approximation on a 300 points regular mesh of $[0, 1]$ on the one hand and the maximum value of the closed-form expression on the other hand. The tested cases are the Brownian motion, the Brownian bridge and the stationary Ornstein-Uhlenbeck process defined by the SDE (2.13) with $\sigma = 1$ and $\theta = 1$.

Closed-form	Mid-point Nyström 25 points	Mid-point Nyström 50 points	Mid-point Nyström 100 points	Mid-point Nyström $25-50-100$ Richardson-Romberg relative error
0.101321184	0.101454622	0.101354524	0.101329517	1.0181×10^{-11}
0.0253302959	0.0254640514	0.0253636556	0.0253386309	6.5299×10^{-10}
0.0112579093	0.0113921955	0.0112913019	0.0112662463	7.4651×10^{-9}
0.00633257398	0.00646760876	0.00636601285	0.00634091389	4.2158×10^{-8}
0.00405284735	0.00418885438	0.00408634582	0.00406119097	1.6188×10^{-7}

Table 2.2: Record of the first five eigenvalues of the Karhunen-Loève decomposition of the Brownian bridge on $[0,1]$.

Closed-form	Mid-point Nyström 25 points	Mid-point Nyström 50 points	Mid-point Nyström 100 points	Mid-point Nyström $25-50-100$ Richardson-Romberg relative error
0.369405405	0.3696101981	0.3694566011	0.3694182037	2.8612×10^{-13}
0.0690018877	0.06916548001	0.06904275722	0.06901210328	8.5348×10^{-12}
0.0225442436	0.02268929627	0.02258041792	0.02255328167	6.9035×10^{-10}
0.0106644656	0.01080431390	0.01069923895	0.01067314723	7.7028×10^{-9}
0.00613945693	0.006277759263	0.006173702808	0.006147997976	4.2950×10^{-8}

Table 2.3: Record of the first five eigenvalues of the Karhunen-Loève decomposition of the stationary Ornstein-Uhlenbeck process defined on $[0,1]$ by the SDE $dr_t = -r_t dt + dW_t, \quad r_0 \overset{\mathcal{L}}{\sim} \mathcal{N}\left(0, \frac{1}{2}\right)$.

Richardson-Romberg $50-100-200$ relative error	e_1	e_2	e_3	e_4	e_5
Standard Brownian motion on $[0,1]$	2.7414×10^{-6}	2.4685×10^{-5}	6.8433×10^{-5}	1.3473×10^{-4}	2.2315×10^{-4}
Standard Brownian bridge on $[0,1]$	1.0964×10^{-5}	4.3908×10^{-5}	9.8867×10^{-5}	1.7584×10^{-4}	2.7245×10^{-4}
Stationary Ornstein-Uhlenbeck process on $[0,1]$ with $\sigma = 1$ and $\theta = 1$	3.0076×10^{-6}	1.6107×10^{-5}	4.9283×10^{-5}	1.0442×10^{-4}	1.8157×10^{-4}

Table 2.4: Record of the largest relative error on the Karhunen-Loève eigenfunctions approximation by the Richardson-Romberg extrapolated mid-point Nyström method (relative with respect to the maximum of the eigenfunction). The number of time steps used for the 3-steps Richardson-Romberg extrapolation are 50, 100 and 200. We used 300 equally spaced points on $[0,1]$. Each column corresponds to an eigenfunction.

2.3 Quantization of the fractional Brownian motion

The normalized fractional Brownian motion B^H, is a centered Gaussian process on $[0,T]$, which has the following covariance function:

$$\Gamma^{B^H}(t,s) = \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |s-t|^{2H}\right), \tag{2.14}$$

where $H \in (0,1)$ is called the Hurst parameter. If $H = \frac{1}{2}$ then the process is the standard Brownian motion.

A simple application of the Nyström method presented in Section 2.1 produces regularly shaped functional quantizers of the fractional Brownian motion. In Figure 2.1, a $5\times2\times2$-product quantizer of the fractional Brownian motion with 3 different values of the Hurst parameter is plotted.

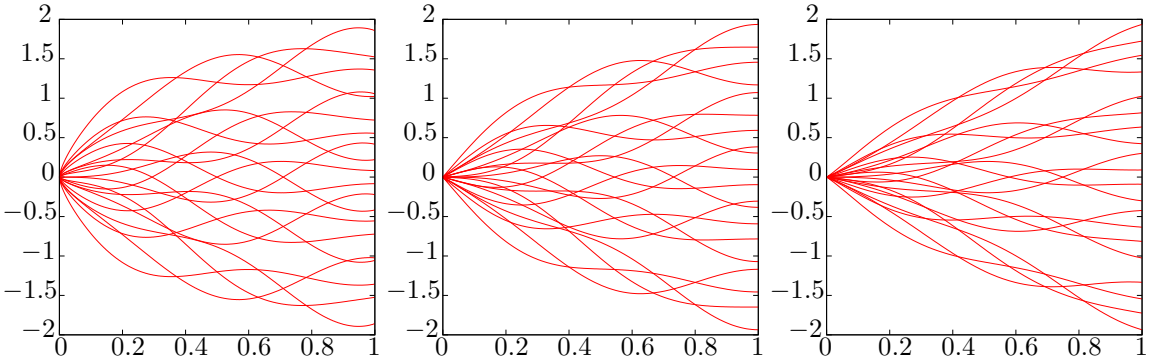

Figure 2.1: $5\times2\times2$-product quantizer of fractional Brownian motions on $[0,1]$ with Hurst exponent $H = 0.3$ (left), $H = 0.5$ (middle) and $H = 0.7$ (right).

Still, for $H < \frac{1}{2}$, the covariance function of the fractional Brownian motion has singularities that break the convergence of the mid-point integration rule in even powers of the step sizes. Indeed, the derivative of $t \mapsto \Gamma^{B^H}(t,s)$ has an infinite limit as $t \to 0^+$ and as ($t \to s^-$ or $t \to s^+$). It breaks also the convergence of the whole associated Nyström method in even powers of the step sizes. In [1, 6, 24], several methods to handle such boundary and diagonal singularities are proposed. We will deal with this in Section 2.3.1

However, so is not the case for $H \geq \frac{1}{2}$, and we can trust in the results of the method in this case. In Table 2.5, we report the first five Karhunen-Loève eigenvalues of the fractional Brownian motion on $[0,1]$ with Hurst exponent $H = 0.7$. The number of time steps are 128, 256 and 512. Last column yields the corresponding three-steps Richardson-Romberg extrapolation. All the computation has been performed with an octuple precision floating point number implementation to increase the accuracy of the 513×513-matrix eigensystem computation. (Let us recall that in the case of the Brownian motion on $[0,1]$, when performing the same computation, we get an relative error smaller than 1×10^{-7} for the first five eigenvalues.)

Mid-point Nyström 128 points	Mid-point Nyström 256 points	Mid-point Nyström 512 points	Mid-point Nyström $128-256-512$ Richardson-Romberg
0.374536638	0.374533535	0.374532774	0.374532521757236
0.0250351543	0.0250343274	0.0250341354	0.0250340726875501
0.00728913038	0.00728860123	0.00728848368	0.0072884458064217
0.00322117252	0.00322075790	0.00322066901	0.0032206406932789
0.00176153269	0.00176116702	0.00176109039	0.00176106615722872

Table 2.5: Record of the first five eigenvalues of the fractional Brownian motion on $[0,1]$ with Hurst exponent $H = 0.7$.

2.3.1 Kernel singularities when $H < \frac{1}{2}$

As pointed out above, the covariance function of the fractional Brownian motion has a boundary singularity as $t \to 0_+$ and a diagonal singularity. In this section, we will use classical methods to handle this kind of singularities. See [1, 6, 24] for a review of these method.

Handling the boundary singularity

Change of variable

The singular behavior covariance function Γ^{B^H} of the fractional Brownian motion, defined in Equation (2.14) can be removed by a change of variable. The change of variable $u = t^{2H}$ and $v = s^{2H}$ in integral (2.7) yields:

$$\int_0^{T^{2H}} \Gamma^{B^H}\left(u^{\frac{1}{2H}}, v^{\frac{1}{2H}}\right) f_k\left(v^{\frac{1}{2H}}\right) \frac{1}{2H} v^{\frac{1}{2H}-1} dv = \lambda_k f_k\left(u^{\frac{1}{2H}}\right). \tag{2.15}$$

(The second change of variable aims at preserving the symmetry of the Kernel.)
This leads to

$$\int_0^{T^{2H}} \frac{1}{2}\left(|u| + |v| - \left|u^{\frac{1}{2H}} - v^{\frac{1}{2H}}\right|^{2H}\right) f_k\left(v^{\frac{1}{2H}}\right) \frac{1}{2H} v^{\frac{1}{2H}-1} dv = \lambda_k f_k\left(u^{\frac{1}{2H}}\right). \tag{2.16}$$

Quadrature rule on a single interval

We now derive a quadrature rule on $[0, T]$ with respect to the weight function $w(v) = \frac{1}{2H} v^{\frac{1}{2H}-1} = \frac{1}{2H} v^{\alpha}$ with $\alpha := \frac{1}{2H} - 1$. The aim is to make the quadrature rule exact with affine functions as for the trapezoidal quadrature rule is, in the case of an integration with a constant weight.

$$\int_l^r \frac{1}{2H} x^{\alpha}(ax + b)dx = w_l(al + b) + w_r(ar + b) \qquad \forall (a, b) \in \mathbb{R}^2.$$

This yields

$$\frac{1}{2H}\left(\frac{a}{\alpha+2}\left(r^{\alpha+2} - l^{\alpha+2}\right) + \frac{b}{\alpha+1}\left(r^{\alpha+1} - l^{\alpha+1}\right)\right) = a(w_l l + w_r r) + b(w_l + w_r) \qquad \forall (a, b) \in \mathbb{R}^2.$$

i.e.

$$\begin{pmatrix} l & r \\ 1 & 1 \end{pmatrix} \begin{pmatrix} w_l \\ w_r \end{pmatrix} = \begin{pmatrix} \frac{1}{2H}\frac{1}{\alpha+2}\left(r^{\alpha+2} - l^{\alpha+1}\right) \\ \frac{1}{2H}\frac{1}{\alpha+1}\left(r^{\alpha+1} - l^{\alpha+1}\right) \end{pmatrix}.$$

The solution of the linear system is

$$w_l = \frac{1}{2H}\frac{(\alpha+1)l^{\alpha+2} + r^{\alpha+2} - (\alpha+2)l^{\alpha+1}r}{(\alpha+1)(\alpha+2)(r-l)}, \qquad w_r = \frac{1}{2H}\frac{(\alpha+1)r^{\alpha+2} + l^{\alpha+2} - (\alpha+2)r^{\alpha+1}l}{(\alpha+1)(\alpha+2)(r-l)}.$$

This is

$$w_l = \frac{l^{\frac{1}{2H}+1} + 2Hr^{\frac{1}{2H}+1} - (2H+1)l^{\frac{1}{2H}}r}{(2H+1)(r-l)}, \qquad w_r = \frac{r^{\frac{1}{2H}+1} + 2Hl^{\frac{1}{2H}+1} - (2H+1)r^{\frac{1}{2H}}l}{(2H+1)(r-l)}.$$

Quadrature rule for equally spaced abscissas

Let us now consider the equally spaced abscissas points $x_i = i\frac{T}{n}$, $i = 0, 1, \cdots, n$. We now use these weights n times to integrate on intervals $\left(x_0^{2H}, x_1^{2H}\right), \left(x_1^{2H}, x_2^{2H}\right), \cdots, \left(x_{n-1}^{2H}, x_n^{2H}\right)$ to obtain the extended rule of quadrature. The convergence rate of this method is the same as for the trapezoidal rule.

Handling the diagonal singularity

We now have to handle the diagonal singularity $|u - v|^{2H}$ in Equation (2.7). One classical method is to use the smoothness of the solution by *subtracting the singularity.*

$$\int_0^T \Gamma^{B^H}(t, s) f(s) ds = \int_0^T \Gamma^{B^H}(t, s)\left(f(s) - f(t)\right) ds + r(t) f(t),$$

where $r(t) = \int_0^T \Gamma^{B^H}(t,s)ds$. The discretized eigenvalue problem is now transformed into

$$\begin{aligned} \lambda_k f_k(t_i) &= \sum_{j=1}^{n} w_j K_{ij} \left(f_k(t_j) - f_k(t_i)\right) + r(t_i) f_k(t_i) \\ &= \sum_{j=1}^{n} w_j K_{ij} f_k(t_j) + \left(r(t_i) - \sum_{j=0}^{n} w_j K_{ij} \right) f_k(t_i). \end{aligned} \tag{2.17}$$

We now define the diagonal matrix $D := \mathrm{diag}(w_i)_{1\le i\le n}$ as in Section 2.1. Moreover, we denote $\Delta := \mathrm{diag}\left(r(t_i) - \sum_{j=0}^{n} w_j K_{ij} \right)_{1\le i\le n}$. Equation (2.17) writes

$$\lambda_k f_k = K D f_k + \Delta f_k.$$

Multiplying by $D^{\frac{1}{2}}$ yields $\lambda h = \left(D^{\frac{1}{2}} K D^{\frac{1}{2}} + \Delta\right) h$, with $h = D^{\frac{1}{2}} f$. As a consequence, we obtain again a symmetric matrix eigenvalue problem. In the case of the fractional Brownian motion, the function $r(t) = \int_0^T \Gamma^{B^H}(t,s)ds$ is derived explicitly:

$$r(t) = \frac{1}{2}\left(\frac{T^{2H+1} - u^{2H+1}}{2H+1} + u^{2H} T - \frac{(T-u)^{2H+1}}{2H+1} \right).$$

Optimal quantization of the fractional Brownian motion

We now use this approximation of the Karhunen-Loève basis to perform an optimal quantization of the fractional Brownian motion with a 50-100-200 three-step Richardson-Romberg extrapolated Nyström method.

In Figure 2.2, we display the quadratic optimal N-quantizer of the fractional Brownian motion on $[0,1]$ with Hurst exponent $H = 0.25$ and $N = 20$. In this case, the quantization dimension is 3.

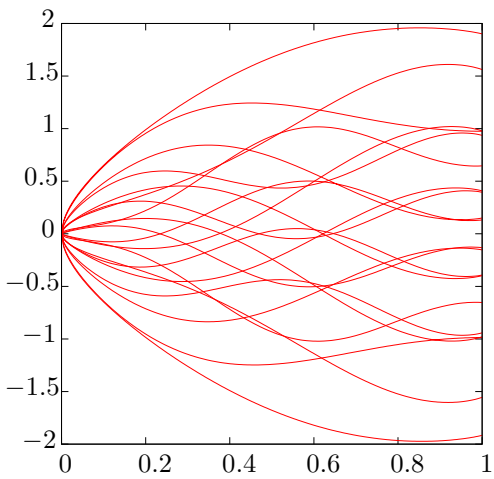

Figure 2.2: Quadratic N-optimal quantizer of the fractional Brownian motion on $[0,1]$ with Hurst parameter $H = 0.25$ and $N = 20$.

2.4 Application to the functional stratification of the fractional Brownian motion

In this section, we experiment with the functional quantization-based stratified sampling algorithm proposed in [4] with the fractional Brownian motion.

2.4.1 Background on stratification

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

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

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

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

Let M be the global budget allocated to the computation of $\mathbb{E}[F(X)]$ and $M_i = q_i M$ the budget allocated to compute $\mathbb{E}[F(X_i)]$ in each stratum. We assume that $\sum_{i\in I} q_i = 1$. This leads to define the (unbiased) estimator of $\mathbb{E}[F(X)]$:

$$\overline{F(X)}_M^I := \sum_{i\in I} p_i \frac{1}{M_i} \sum_{k=1}^{M_i} F(X_i^k), \tag{2.18}$$

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

Proposition 2.4.1. *With the same notations:*

$$\mathrm{Var}\left(\overline{F(X)}_M^I\right) = \frac{1}{M} \sum_{i\in I} \frac{p_i^2}{q_i} \sigma_{F,i}^2, \tag{2.19}$$

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

In [4], it is pointed out that theoretical aspects of quantization lead to a strong link between the problem of optimal L^2-quantization of a random variable and the variance reduction that can be achieved by stratification. Three types of allocation rules for the budgets $(q_i)_{i\in I}$ are proposed:

- The "sub-optimal rule" is to set
$$q_i = p_i, \qquad i \in I. \tag{2.20}$$
Two possible motivations for this choice are the facts that the weights p_i are known and because it always reduces the variance.
- The "optimal rule" is obtained when minimizing the variance in Equation (2.19). The solution of the minimization problem is given by
$$q_i^* = \frac{p_i \sigma_{F,i}}{\sum_{j\in I} p_j \sigma_{F,j}}, \quad i \in I \tag{2.21}$$
and the corresponding minimal variance is $\left(\sum_{i\in I} p_i \sigma_{F,i}\right)^2$.

  A counterpart of this method is that we do not know explicitly the solution $(q_i^*)_{i\in I}$. In [26], Étoré and Jourdain proposed an algorithm for adaptively modifying the proportion of further drawings in each stratum, that converges to the optimal allocation. This can be used in a general framework. Another practical solution is to implement a simple prior rough estimation of the optimal allocation.

- The "Lipschitz optimal" rule. When the partition $(A_i)_{i\in I}$ is a Voronoi partition associated with an optimal quantizer of X, the following setting is considered

$$q_i = \frac{p_i\sigma_i}{\sum_{j\in I} p_j\sigma_j}, \quad i \in I, \tag{2.22}$$

  where σ_i is the local inertia of the random variable X, $\sigma_i^2 = \mathbb{E}\Big[|X - \mathbb{E}[X|X \in A_i]|^2 \Big| X \in A_i\Big]$. It is proved that this setting has a uniform efficiency among the class of Lipschitz continuous functionals. Moreover, local inertia $(\sigma_i)_{i\in I}$ are known. This solution overcomes the "sub-optimal choice" in every test done in [4].

2.4.2 On the functional stratification of Gaussian processes

Let X be a centered bi-measurable Gaussian process on $[0,T]$ with a continuous covariance function on $[0,T]^2$. We are interested by the value of $\mathbb{E}[F(X_{t_0}, X_{t_1}, \cdots, X_{t_n})]$ where $0 = t_0 \le t_1 \le \cdots \le t_n = T$ are $n+1$ dates of interest for the underlying process. Let us assume that $\chi \in \mathcal{O}_{pq}(X,N)$ is a K-L product quantizer of X. The codebook associated with this product quantizer is the set of the paths of the form

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

where $\big(e_n^X, \lambda_n^X\big)$ is the Karhunen-Loève decomposition of the process X on $[0,T]$ and $x_{i_n}^{(N_n)}$ is the i_nth element of an optimal quantizer of size N_n of the standard one-dimensional Gaussian distribution.

We now need to be able to simulate the conditional distribution

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

where $A_{\underline{i}}$ is the slab associated with $\chi_{\underline{i}}$ in the codebook. To simulate this conditional distribution, we will:

- First, simulate the first K-L coordinates of X. A detailed simulation procedure is available in [4]
- Then simulate the conditional distribution of the marginals of the Gaussian process, its first coordinates being fixed.

In this setting, we need to simulate the conditional distribution

$$\mathcal{L}\Big(X_{t_0}, \cdots, X_{t_n} \Big| \int_0^T X_s e_1^X ds, \int_0^T X_s e_2^X(s) ds, \cdots, \int_0^T X_s e_d^X(s) ds\Big). \tag{2.23}$$

Conditional simulation: In [4], two solutions are proposed for the simulation of the conditional distribution (2.23).

- The first one is the naive Cholesky method for Gaussian vector simulation, which has a quadratic complexity in the number of time steps. This first simulation scheme was not competitive for linearly simulatable processes such as the Brownian motion. In the following, we will mention this method as the *brute force method*.
- The other solution, detailed in [4] requires a prior simulation of the unconditional distribution of $(X_{t_0}, \cdots, X_{t_n})$ and has then a linear additional cost. This algorithm will be mentioned in the following as the *Bayesian algorithm*. For Gaussian processes which have a linear simulation scheme in the unconditional case (as Ornstein-Uhlenbeck processes, the Brownian bridge and the Brownian motion), this method is of high interest.

2.4.3 The case of the fractional Brownian motion

Possible methods for simulating the fractional Brownian motion on a schedule $t_0 < t_1 < \cdots < t_n$ are

- the naive Cholesky method which has quadratic complexity,
- and the circulant matrix method which has a $O(n \ln(n))$ complexity [7, 28]. The circulant matrix method is also available for the multifractional Brownian motion [29].

No exact simulation scheme with a linear complexity exists for the fractional Brownian motion. If we choose the Cholesky method, there is no interest in using the Bayesian algorithm proposed in [4]. The brute force Cholesky method is adapted to this situation.

In every other case, if the unconditional simulation method has a smaller complexity, we have interest to use the Bayesian algorithm which has a linear additional cost to the unconditional simulation.

In Figure 2.3, we plot a few paths of the conditional distribution of the fractional Brownian motion with Hurst parameter $H = 0.3$ knowing that they belong to a given L^2 Voronoi cell.

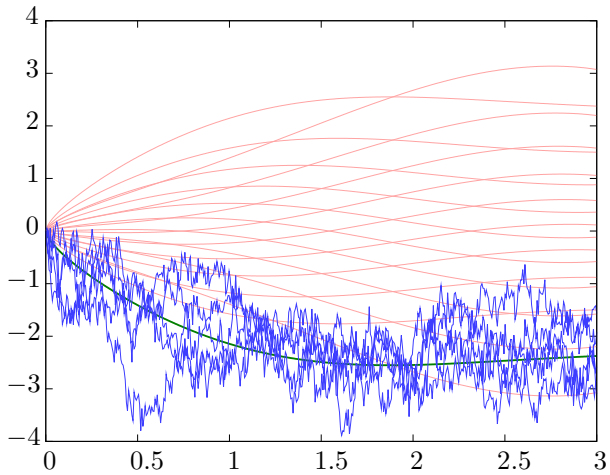

Figure 2.3: Plot of a few paths of the conditional distribution of the fractional Brownian motion with Hurst parameter $H = 0.3$ on $[0, 3]$, knowing that its path belong to the L^2 Voronoi cell of the highlighted curve in the quantizer.

2.4.4 Gaussian process reconstruction

The first numerical test of the functional stratification of the fractional Brownian motion is a method to validate both the eigenfunction computation by the Nyström method and the functional stratification algorithm.

Indeed, one can rebuild the considered Gaussian process from its stratification, by following the steps bellow

- First, simulate the discrete weighted distribution of the strata index $(i, p_i)_{i \in I}$ to select the strata.
- Then simulate the conditional distribution $\mathcal{L}\left(X_{t_0}, \cdots, X_{t_n} \middle| X \in A_i\right)$ of the Gaussian process in the strata by the method described above.

The result should be distributed according to the distribution of the underlying Gaussian process. In Table 2.6, we report the covariance structure $\mathbb{E}\left[X_{t_i} X_{t_j}\right]_{1 \leq i,j \leq n}$ estimated by a Monte-Carlo

simulation when X is a fractional Brownian motion with Hurst parameter $H = 0.7$. The tested schedule is $\left(i\frac{T}{n}\right)_{0\leq i\leq n}$ with $T = 1$ and $n = 5$. The product decomposition of the quantization is $10 \times 5 \times 2$.

0.105061	0.138629	0.15846	0.173817	0.186687
0.138629	0.277258	0.330656	0.365844	0.394071
0.15846	0.330656	0.489116	0.557871	0.605929
0.173817	0.365844	0.557871	0.73168	0.813313
0.186687	0.394071	0.605929	0.813313	1

0.105141	0.138748	0.158596	0.173959	0.186824
0.138748	0.277417	0.330885	0.366075	0.394372
0.158596	0.330885	0.489454	0.558177	0.606266
0.173959	0.366075	0.558177	0.731923	0.813579
0.186824	0.394372	0.606266	0.813579	1.0003

Table 2.6: Theoretical (left) and estimated (right) covariance $\mathbb{E}[X_{t_i}X_{t_j}]$ of the rebuilt fractional Brownian motion with $H = 0.7$. The number of generated paths for this Monte-Carlo simulation was 1×10^7.

In every tested case, when generating Table 2.6, the theoretical value lies in the 95% confidence interval. These confidence intervals were not displayed for briefness. We obtain the same order of accuracy with other values of $H \in (0, 1)$.

2.4.5 Application to option pricing

A stochastic integral with respect to the fractional Brownian motion has been introduced in [10] by Helliot and van der Hoek, and in [3] by Biagini, Øksendal, Sulem and Wallner, using the white noise theory. They proposed a generalization of the Black-Scholes model. As in the classical Black-Scholes market, two assets are available:

- A risk-free asset whose price is given by
$$dS_t^0 = rS_t^0 dt \tag{2.24}$$
- and a risky asset whose price is given by
$$dS_t = \mu S_t dt + \sigma S_t dB_t^H, \tag{2.25}$$
where r, μ and σ are constants and B^H is fractional Brownian motion with Hurst parameter H.

It has been shown that this market presents no arbitrage opportunity and is complete. Moreover, the solution of the stochastic differential Equation (2.25) is given by
$$S_t = S_0 \exp\left(\sigma B_t^H + \mu t - \frac{1}{2}\sigma^2 t^{2H}\right). \tag{2.26}$$

The following theorem, proved in [10] deals with the price of a European Call option.

Theorem 2.4.2 (Fractional Black-Scholes formula)**.** *The price at every time $t \in [0, T]$ of a European Call option with strike price K and maturity T is given by*
$$C(t, S_t) = S_t\mathcal{N}(d_1) - Ke^{-r(T-t)}\mathcal{N}(d_2) \tag{2.27}$$
where
$$d_1 = \frac{\ln\left(\frac{S_t}{K}\right) + r(T-t) + \frac{\sigma^2}{2}(T^{2H} - t^{2H})}{\sigma\sqrt{T^{2H} - t^{2H}}}, \tag{2.28}$$
and
$$d_2 = \frac{\ln\left(\frac{S_t}{K}\right) + r(T-t) - \frac{\sigma^2}{2}(T^{2H} - t^{2H})}{\sigma\sqrt{T^{2H} - t^{2H}}}. \tag{2.29}$$

This closed-form expression is used to benchmark our simulation scheme of the fractional Brownian motion.

Benchmark with a barrier option in a H-fractional Black and Scholes model

Here, we test the numerical method for a barrier option in the fractional Black and Scholes model. For the sake of simplicity, we consider a log-normal Black and Scholes diffusion with no drift (no interest rate and no dividend). The chosen Hurst exponent is $H = 0.3$. The numerical results are reported in Table 2.7.

The results are displayed for different values of the initial spot S, the strike K, the barrier B, the volatility σ, the maturity T and the number of equally spaced fixing dates n.

In this table, the first column corresponds to a simple Monte-Carlo estimator. The last three columns correspond to stratified sampling estimators with different allocation strategies for the Monte-Carlo simulations.

The "sub-optimal weights" column stands for the allocation budget of Equation (2.20). The "Lip.-optimal weights" column stand for the "universal stratification" budget allocation of Equation (2.22). Both these two case have explicit allocation rules. Last column, "Optimal weights" corresponds to an estimation of the optimal budget allocation given in expression (2.21).

Parameters	Simple estimator	Strat. estimator sub-optimal weights	Strat. estimator Lip.-optimal weights	Strat. estimator optimal weights
$S = 100$, $K = 100$ $B = 125$, $\sigma = 0.3$, $T = 1.5$, $n = 11$	12.5947 $[12.4429, 12.7466]$ Var = 600.5711	12.5674 $[12.4732, 12.6615]$ Var = 230.8692	12.5566 $[12.4654, 12.6477]$ Var = 216.3442	12.5890 $[12.5201, 12.6579]$ Var = 123.5426
$S = 100$, $K = 100$ $B = 200$, $\sigma = 0.3$, $T = 1$, $n = 11$	1.3412 $[1.2677, 1.4146]$ Var = 140.5978	1.3826 $[1.3140, 1.4511]$ Var = 122.2808	1.3613 $[1.3002, 1.4224]$ Var = 97.1538	1.3769 $[1.3530, 1.4009]$ Var = 14.9352

Table 2.7: Numerical results for the Up-In Call option, with $100 = 10 \times 5 \times 2$ stratas.

We notice that the quantization-based stratified sampling method noticeably reduces the variance of the Monte-Carlo estimator. The universal stratification allocation rule (2.22) proposed in [4] overcomes the sub-optimal weight allocation. Moreover, the "optimal allocation" estimation yields a better variance reduction factor.

Bibliography

[1] Kendall E. Atkinson. *The numerical solution of integral equations of the second kind.* Cambridge Monographs on Applied and Computational Mathematics, 1999.

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

[3] Francesca Biagini, Bernt Øksendal, Agnès Sulem, and Naomi Wallner. An introduction to white-noise theory and Malliavin calculus for fractional Brownian motion. *Proceedings: Mathematical, Physical and Engineering Sciences*, 460(2041):347–372, 2004.

[4] Sylvain Corlay and Gilles Pagès. Functional quantization-based stratified sampling methods. *Preprint*, 2010.

[5] Paul Deheuvels and Guennadi V. Martynov. A Karhunen-Loève decomposition of a Gaussian process generated by independent pairs of exponential random variables. *Journal of Functional Analysis*, 255(9):2363–2394, 2008.

[6] Leonard Michael Delves and Julie L. Mohamed. *Computational methods for integral equations.* Cambridge University Press, 1985.

[7] Claude R. Dietrich and Garry N. Newsam. Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix. *SIAM Journal Sci. Comput.*, 18:1088–1107, 1997.

[8] Kacha Dzhaparidze and Harry van Zanten. A series expansion of fractional Brownian motion. *Probability theory and related fields*, 130:39–55, 2004.

[9] Kacha Dzhaparidze and Harry van Zanten. Optimality of an explicit series expansion of the fractional Brownian sheet. *Statistics and probability letters*, 71:295–301, 2005.

[10] Robert J. Elliott and John van der Hoek. A general fractional white noise theory and applications to finance. *Mathematical Finance*, 13(2):301–330, 2003.

[11] Jacques Istas. Karhunen-Loève expansion of spherical fractional Brownian motions. *Statistics & Probability Letters*, 76(14):1578–1583, 2006.

[12] Ramón Gutiérrez Jaimez and Mariano J. Valderrama Bonnet. On the Karhunen-Loève expansion for transformed processes. *Trabajos De Estadistica*, 2(2):81–90, 1987.

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

[14] Stefan Junglen and Harald Luschgy. A constructive sharp approach to functional quantization of stochastic processes. *Journal of Applied Mathematics*, 2010.

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

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

[17] Harald Luschgy and Gilles Pagès. High-resolution product quantization for Gaussian processes under sup-norm distortion. *Bernoulli*, 13(3):653–671, 2007.

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

[19] Gilles Pagès and Harald Luschgy. Expansions for Gaussian processes and Parseval frames. *Electronic Journal of Probability*, 14, Paper no. 42:1198–1221, 2010.

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

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

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

[23] Gilles Pagès and Afef Sellami. Convergence of multi-dimensional quantized SDE's. In Catherine Donati-Martin, Antoine Lejay, and Alain Rouault, editors, *Séminaire de Probabilités XLIII*, pages 269–308. Springer, Berlin, 2010.

[24] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical recipes in C++: The art of scientific computing.* Cambridge University Press, February 2002.

[25] Jean-Renaud Pycke. Explicit Karhunen-Loève expansions related to the Green function of the Laplacian. *Banach Center Publ.*, 72:263–270, 2006.

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

[27] Benedikt Wilbertz. *Construction of optimal quantizers for Gaussian measures on Banach spaces.* PhD thesis, Universität Trier, 2008.

[28] Andrew T. A. Wood and Grace Chan. Simulation of stationary Gaussian processes in $[0,1]^d$. *Journal of Comp. and Graphical Statistics*, 3:409–432, 1994.

[29] Andrew T. A. Wood and Grace Chan. Simulation of multifractional Brownian motion. *Proc. Comput. Statist*, pages 233–238, 1998.

Chapter 3

Partial functional quantization and generalized bridges

Abstract

In this chapter, we develop a new approach to functional quantization, which consists in discretizing only the first Karhunen-Loève coordinates of a continuous Gaussian semimartingale X. Using filtration enlargement techniques, we prove that the conditional distribution of X knowing its first Karhunen-Loève coordinates is a Gaussian semimartingale with respect to its natural filtration.

This allows us to define the partial quantization of a solution of a stochastic differential equation with respect to X by simply plugging the partial functional quantization of X in the SDE.

Then we provide an upper bound of the L^p-partial quantization error for the solution of SDE involving the $L^{p+\varepsilon}$-partial quantization error for X, for $\varepsilon > 0$. The *a.s.* convergence is also investigated.

Incidentally, we show that the conditional distribution of a Gaussian semimartingale X, knowing that it stands in some given Voronoi cell of its functional quantization, is a (non-Gaussian) semimartingale. As a consequence, the functional stratification method developed in [7] amounted, in the case of solutions of SDE, to using the Euler scheme of these SDE in each Voronoi cell.

Keywords: Gaussian semimartingale, functional quantization, vector quantization, Karhunen-Loève, Gaussian process, Brownian motion, Brownian bridge, Ornstein-Uhlenbeck, filtration enlargement, stratification, Cameron-Martin space, Wiener integral.

Introduction

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, and E a reflexive separable Banach space. The norm on E is denoted by $|\cdot|$. The quantization of a E-valued random variable X consists in its approximation by a random variable Y taking finitely many values. The resulting error of this discretization is measured by the L^p norm of $|X-Y|$. If we settle on a fixed maximum cardinal for $Y(\Omega)$, the minimization of the error comes to the following minimization problem:

$$\min\left\{ \left\| |X-Y| \right\|_p,\ Y:\Omega \to E \text{ measurable},\ \operatorname{card}(Y(\Omega)) \leq N \right\}. \tag{3.1}$$

A solution of (3.1) is an optimal quantizer of X. This problem, initially investigated as a signal discretization method [10], has then been introduced in numerical probability to devise cubature methods [22] or to solve multidimensional stochastic control problems [4]. Since the early 2000's, the infinite-dimensional setting has been extensively investigated from both constructive numerical and theoretical viewpoints with a special attention paid to functional quantization, especially in the quadratic case [17] but also in some other Banach spaces [28]. Stochastic processes are viewed as random variables taking values in their path spaces such as $L^2_T := L^2([0,T], dt)$.

We now assume that X is a bi-measurable stochastic process on $[0,T]$ verifying $\int_0^T \mathbb{E}\left[|X_t|^2\right] dt < +\infty$, so that this can be viewed as a random variable valued in the separable Hilbert space $L^2([0,T])$. We assume that its covariance function Γ^X is continuous. In the seminal article on Gaussian functional quantization [17], it is shown that in the centered Gaussian case, linear subspaces U of $L^2([0,T])$ spanned by L^2-optimal quantizers correspond to principal components of X. In other words, they are spanned by the first eigenvectors of the covariance operator of X. Thus, the quadratic optimal quantization of Gaussian processes involves its Karhunen-Loève decomposition $(e_n^X, \lambda_n^X)_{n\geq 1}$.

If Y is a quadratic N-optimal quantizer of the Gaussian process X and $d^X(N)$ is the dimension of the subspace of $L^2([0,T])$ spanned by $Y(\Omega)$, the quadratic quantization error $\mathcal{E}_N(X)$ verifies

$$\mathcal{E}_N^2(X) = \sum_{j\geq m+1} \lambda_j^X + \mathcal{E}_N^2\left(\bigotimes_{j=1}^m \mathcal{N}\left(0, \lambda_j^X\right)\right) \text{ for } m \geq d^X(N). \tag{3.2}$$

$$\mathcal{E}_N^2(X) < \sum_{j\geq m+1} \lambda_j^X + \mathcal{E}_N^2\left(\bigotimes_{j=1}^m \mathcal{N}\left(0, \lambda_j^X\right)\right) \text{ for } 1 \leq m < d^X(N). \tag{3.3}$$

To perform optimal quantization, the decomposition is first truncated at a fixed order m and then the $\mathbb{R}^m$-valued Gaussian vector, constituted of the m first coordinates of the process on its Karhunen-Loève decomposition, is quantized. To reach optimality, we have to determine the optimal rank of truncation $d^X(N)$ (the quantization dimension) and the optimal $d^X(N)$-dimensional quantizer corresponding to the first coordinates $\bigotimes_{j=1}^{d^X(N)} \mathcal{N}\left(0, \lambda_j^X\right)$. Usual examples of such processes are the standard Brownian motion on $[0,T]$, the Brownian bridge on $[0,T]$, Ornstein-Uhlenbeck processes and the fractional Brownian motion. In Figure 3.1, we display the quadratic optimal N-quantizer of the fractional Brownian motion on $[0,1]$ with Hurst exponent $H = 0.25$ and $N = 20$.

Another possibility is to use a product quantization of the distribution $\bigotimes_{j=1}^{d^X(N)} \mathcal{N}(0, \lambda_j^X)$. The product quantization is the Cartesian product of the optimal quadratic quantizers of the standard one-dimensional Gaussian distributions $\mathcal{N}(0, \lambda_i^X)_{1\leq i\leq d^X(N)}$. In the case of independent marginals, this yields a stationary quantizer, *i.e.* a quantizer Y of X which satisfies $\mathbb{E}[X|Y] = Y$. This property, shared with optimal quantizers, results in a convergence rate of a higher order by one for the quantization-based cubature method. One advantage of this setting is that the one-dimensional Gaussian quantization is a fast procedure. In [23], deterministic optimization methods (as Newton-Raphson) are shown to converge rapidly to the unique optimal quantizer of the one-dimensional

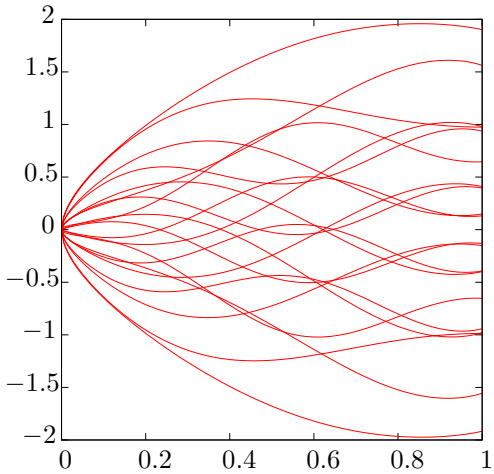

Figure 3.1: Quadratic N-optimal quantizer of the fractional Brownian motion on $[0,1]$ with Hurst parameter $H = 0.25$ and $N = 20$. The quantization dimension is 3.

Gaussian distribution. Moreover, a sharply optimized database of quantizers of standard univariate and multivariate Gaussian distributions is available on the web site `www.quantize.maths-fi.com` [25] for download. Still, we have to determine the quantization size for each direction to obtain optimal product quantization. In this case, the minimization of the distortion (3.2) comes to:

$$\min \left\{ \sum_{j=1}^{d} \mathcal{E}_{N_j}^2 \left(\mathcal{N}\left(0, \lambda_j^X\right)\right) + \sum_{j \geq d+1} \lambda_j^X, \ N_1 \times \cdots \times N_d \leq N, \ d \geq 1 \right\}. \tag{3.4}$$

In [17], the rate of convergence to zero of the quantization error is investigated. A complete solution is provided for the case of Gaussian processes under rather general conditions on the eigenvalues of the covariance operator. Rates of convergence are available for the above cited examples of Gaussian processes. The asymptotics of the quantization dimension $d^X(N)$ is investigated in [18, 20].

From a constructive viewpoint, the numerical computation of the optimal quantization or the optimal product quantization requires a numerical evaluation of the Karhunen-Loève eigenfunctions and eigenvalues, at least the very first terms. (As seen in [17, 18, 20], under rather general conditions on its eigenvalues, the quantization dimension of a Gaussian process increases asymptotically as the logarithm of the size of the quantizer. Hence it is most likely that it is small. For instance, the quantization dimension of the Brownian motion with $N = 10000$ is 9.) The Karhunen-Loève decompositions of several usual Gaussian processes have a closed-form expression. This is the case for the standard Brownian motion, the Brownian bridge and Ornstein-Uhlenbeck processes. (The case of Ornstein-Uhlenbeck processes is derived in [7], in the general setting of an arbitrary initial variance σ_0. A pseudo-algorithm for the computation of ω_λ is also provided.) Another example of explicit Karhunen-Loève expansion is derived in [8] by Deheuvels and Martynov.

In the general case, no closed-form expression of the Karhunen-Loève expansion is available. For instance, the Karhunen-Loève expansion of the fractional Brownian motion is not known. To fulfill the requirement of a numerical evaluation of those functions, it is possible to use numerical methods related to integral equations to solve the eigenvalue problem that defines the Karhunen-Loève expansion. A review of these methods is available in [3]. In [6], the so-called "Nyström method" is used to compute the first terms of the Karhunen-Loève decomposition of the fractional Brownian motion for its optimal functional quantization.

An application of the quantization of a Gaussian process X, is to perform a quantization of the solution of a SDE with respect X, when a stochastic integration with respect to X can

be defined. In the following, we will assume that X is a continuous Gaussian semimartingale on $[0,T]$. The Brownian motion, the Brownian bridge and Ornstein-Uhlenbeck processes are semimartingales, but the fractional Brownian motion with Hurst exponent $H \neq \frac{1}{2}$ is not. We can obtain a stationary quantizer of the diffusion by inserting the quantizer of the Gaussian process in the diffusion equation written in the Stratonovich sense. In [26], Pagès and Sellami proved the a.s. convergence of this quantization when the quantizer size goes to infinity. The rate of convergence is also investigated. This work is mostly specific to the Brownian motion but main results remain valid for continuous semimartingales which satisfy the Kolmogorov criterion such as the Brownian bridge and Ornstein-Uhlenbeck processes.

3.1 Quantization-based cubature and related inequalities

The idea of quantization-based cubature method is to approach the probability distribution of the random variable X by the distribution of a quantizer Y of X. As Y is a discrete random variable, we can write $\mathbb{P}_Y = \sum_{i=1}^{N} p_i \delta_{y_i}$. If $F : E \to \mathbb{R}$ is a Borel functional,

$$\mathbb{E}[F(Y)] = \sum_{i=1}^{N} p_i F(y_i). \tag{3.5}$$

Hence, if we have access to the weighed discrete distribution $(y_i, p_i)_{1 \leq i \leq N}$ of Y, we are able to compute the right-hand side of Equation (3.5). Now, we review some error bounds that can be derived when approaching $\mathbb{E}[F(X))]$ by the quantity (3.5). See [24] for more details on error bounds.

1. *If $X \in L^2$, Y a quantizer of X of size N and F is Lipschitz continuous, then*

$$|\mathbb{E}[F(X)] - \mathbb{E}[F(Y)]| \leq [F]_{\mathrm{Lip}} \|X - Y\|_2. \tag{3.6}$$

In particular, if $(Y_N)_{N \geq 1}$ is a sequence of quantizers such that $\lim_{N \to \infty} \|X - Y_N\|_2 = 0$, then the distribution $\sum_{i=1}^{N} p_i^N \delta_{x_i^N}$ of Y_N converges weakly to the distribution $\mathbb{P}_X$ of X as $N \to \infty$.

This first error bound is a straightforward consequence of $|F(X) - F(Y)| \leq [F]_{\mathrm{Lip}} |X - Y|$.

2. *If Y is a stationary quantizer of X, i.e. $Y = \mathbb{E}[X|Y]$, and F is differentiable with an α-Hölder differential DF ($\alpha \in (0,1]$), then*

$$|\mathbb{E}[F(X)] - \mathbb{E}[F(Y)]| \leq [DF]_\alpha \|X - Y\|_2^{1+\alpha}. \tag{3.7}$$

In the case where F has a Lipschitz continuous derivative ($\alpha = 1$), we have. $[DF]_1 = [DF]_{\mathrm{Lip}}$. For example, if F is twice differentiable and D^2F is bounded, then $[DF]_{\mathrm{Lip}} = \|D^2F\|_\infty$.

This particular inequality comes from the Taylor expansion of F around X and the stationarity of Y.

3. *If F is a convex functional and Y is a stationary quantizer of X,*

$$\mathbb{E}[F(Y)] \leq \mathbb{E}[F(X)]. \tag{3.8}$$

This inequality is a straightforward consequence of the stationarity property and Jensen's inequality.

$$\mathbb{E}[F(Y)] = \mathbb{E}[F(\mathbb{E}[X|Y])] \leq \mathbb{E}[\mathbb{E}[F(X)|Y]] = \mathbb{E}[F(X)].$$

3.2 Functional quantization and generalized bridges

3.2.1 Generalized bridges

Let $(X_t)_{t\in[0,T]}$ be a continuous centered Gaussian semimartingale starting from 0 on $(\Omega, \mathcal{A}, \mathbb{P})$ and $\mathcal{F}^X$ its natural filtration. Fernique's theorem ensures that $\int_0^T \mathbb{E}\left[X_t^2\right] dt < +\infty$ (see Janson [13]). We aim here to compute the conditioning with respect to a finite family $\overline{Z}_T := (Z_T^i)_{i\in I}$ of Gaussian random variables, which are measurable with respect to $\sigma(X_t, t \in [0,T])$. ($I \subset \mathbb{N}$ is a finite subset of $\mathbb{N}^*$.) As Alili in [1] we settle on the case where $(Z_T^i)_{i\in I}$ are the terminal values of processes of the form $Z_t^i = \int_0^t f_i(s)dX_s$, $i \in I$, for some given finite set $\overline{f} = (f_i)_{i\in I}$ of $L^2_{loc}([0,T])$ functions. The *generalized bridge* for $(X_t)_{t\in[0,T]}$ corresponding to $\overline{f}$ with end-point $\overline{z} = (z_i)_{i\in I}$ is the process $\left(X^{\overline{f},\overline{z}}\right)_{t\in[0,T]}$ that has the distribution

$$X^{\overline{f},\overline{z}} \overset{\mathcal{L}}{\sim} \mathcal{L}\left(X \middle| Z_T^i = z_i,\ i \in I\right). \tag{3.9}$$

For example, in the case where X is a standard Brownian motion with $|I| = 1$, $\overline{f} = \{f\}$ and $f \equiv 1$, this is the Brownian bridge on $[0,T]$. If X is an Ornstein-Uhlenbeck process this is an Ornstein-Uhlenbeck bridge.

Let H be the Gaussian Hilbert space spanned by $(X_s)_{s\in[0,T]}$ and $H_{\overline{Z}_T}$ the closed subspace of H spanned by $(Z_T^i)_{i\in I}$. We denote by $H_{\overline{Z}_T}^{\perp}$ its orthogonal complement in H. Any Gaussian random variable G of H can be orthogonally decomposed into $G = \mathrm{Proj}_{\overline{Z}_T}(G) \overset{\perp\!\!\!\perp}{+} \mathrm{Proj}^{\perp}_{\overline{Z}_T}(G)$, where $\mathrm{Proj}_{\overline{Z}_T}$ and $\mathrm{Proj}^{\perp}_{\overline{Z}_T}$ are the orthogonal projections on $H_{\overline{Z}_T}$ and $H^{\perp}_{\overline{Z}_T}$. ($\mathrm{Proj}^{\perp}_{\overline{Z}_T} = Id_H - \mathrm{Proj}_{\overline{Z}_T}$). With these notations, $\mathbb{E}\left[G \middle| (Z_T^i)_{i\in I}\right] = \mathrm{Proj}_{\overline{Z}_T}(G)$.

Other definitions of generalized bridges exist in the literature, see *e.g.* [21].

3.2.2 The case of the Karhunen-Loève basis

As X is a continuous Gaussian process, it has a continuous covariance function (see [13, VIII.3]). We denote by $(e_i^X, \lambda_i^X)_{i\geq 1}$ its Karhunen-Loève eigensystem. Thus, if we define function f_i^X as the antiderivative of $-e_i^X$ that vanishes at $t = T$, *i.e.* $f_i^X(t) = \int_t^T e_i^X(s)ds$, an integration by parts yields

$$\int_0^T X_s e_i^X(s)ds = \int_0^T f_i^X(s)dX_s. \tag{3.10}$$

In other words, with the notations of Section 3.2.1, we have $Y_i := \int_0^T X_s e_i^X(s)ds = Z_T^i$.
For some finite subset $I \subset \mathbb{N}^*$, we denote by $X^{I,\overline{y}}$ and call *K-L generalized bridge* the generalized bridge associated with functions $(f_i^X)_{i\in I}$ and with end-point $\overline{y} = (y_i)_{i\in I}$. This process has the distribution $\mathcal{L}(X|Y_i = y_i, i \in I)$.
In this case, the Karhunen-Loève expansion gives the decomposition

$$X = \underbrace{\sum_{i\in I} Y_i e_i^X}_{=\mathrm{Proj}_{\overline{Z}_T}(X)} \overset{\perp\!\!\!\perp}{+} \underbrace{\sum_{i\in\mathbb{N}^*\setminus I} \sqrt{\lambda_i^X}\xi_i e_i^X}_{=\mathrm{Proj}^{\perp}_{\overline{Z}_T}(X)}, \tag{3.11}$$

where $(\xi_i)_{i\in\mathbb{N}^*\setminus I}$ are independent standard Gaussian random variables. This gives us the projections $\mathrm{Proj}_{\overline{Z}_T}$ and $\mathrm{Proj}^{\perp}_{\overline{Z}_T}$ defined in Section 3.2.1. It follows from (3.11) that a K-L generalized bridge is centered on $\mathbb{E}\left[X|Y_i = y_i, i \in I\right]$ and has the covariance function

$$\Gamma^{X|Y}(s,t) = \mathrm{cov}(X_s, X_t) - \sum_{i\in I} \lambda_i^X e_i^X(s) e_i^X(t). \tag{3.12}$$

We have $\int_0^T \Gamma^{X|Y}(t,t)dt = \sum_{i\in\mathbb{N}^*\setminus I} \lambda_i^X$.

Moreover, thanks to decomposition (3.11), if $X^{I,\overline{y}}$ is a K-L generalized bridge associated with X with terminal values $\overline{y} = (y_i)_{i\in I}$, it has the same probability distribution as the process

$$\sum_{i\in I} y_i e_i^X(t) + X_t - \sum_{i\in I}\left(\int_0^T X_s e_i^X(s)ds\right) e_i^X(t).$$

This process is then the sum of a semimartingale and a non-adapted finite-variation process.
Let us stress the fact that the second term in the left-hand side of (3.11) is the corresponding K-L generalized bridge with end-point 0, *i.e.* $\mathrm{Proj}^{\perp}_{\overline{Z}_T} = X^{I,\overline{0}}$.
In [7], an algorithm is proposed to exactly simulate marginals of a K-L generalized bridge with a linear additional cost to a prior simulation of $(X_{t_0},\cdots,X_{t_n})$, for some subdivision $0 = t_0 \le t_1 \le \cdots t_n = T$ of $[0,T]$. This was used for variance reduction issues. Note that the algorithm is easily extended to the case of (non-K-L) generalized bridges.

3.2.3 Generalized bridges as semimartingales

For a random variable L, we denote by $\mathbb{P}[\cdot|L]$ the conditional probability knowing L. We keep the notations and assumptions of previous sections. (X is a continuous Gaussian semimartingale starting from 0.) We consider a finite set $I \subset \{1,2,\cdots\}$ and $(f_i)_{i\in I}$ a set of bounded measurable functions. Let $X^{\overline{f},\overline{y}}$ be the generalized bridge associated with X with end-point $\overline{y} = (y_i)_{i\in I}$. For $i \in I$, $Z_t^i = \int_0^t f_i(s)dX_s$ and $\overline{Z}_t = (Z_t^i)_{i\in I}$.

Jirina's theorem ensures the existence of a transition kernel

$$\nu_{\overline{Z}_T|\left((X_t)_{t\in[0,s]}\right)} : \mathcal{B}(\mathbb{R}^I) \times C^0([0,s],\mathbb{R}) \to \mathbb{R}_+,$$

corresponding to the conditional distribution $\mathcal{L}\left(\overline{Z}_t \middle| \left((X_t)_{t\in[0,s]}\right)\right)$.
We now make the additional assumption $(\mathcal{H})$ that, for every $s \in [0,T)$ and for every $(x_u)_{u\in[0,s]} \in C^0([0,s],\mathbb{R})$, the probability measure $\nu_{\overline{Z}_T|\left((X_t)_{t\in[0,s]}\right)}\left(d\overline{y},(x_u)_{u\in[0,s]}\right)$ is absolutely continuous with respect to the Lebesgue measure. We denote by $\Pi_{(x_u)_{u\in[0,s]},T}$ its density. The covariance matrix of this Gaussian distribution on $\mathbb{R}^I$ writes

$$Q(s,T) := \mathbb{E}\left[\left(\overline{Z}_T - \mathbb{E}\left[\overline{Z}_T|(X_u)_{u\in[0,s]}\right]\right)\left(\overline{Z}_T - \mathbb{E}\left[\overline{Z}_T|(X_u)_{u\in[0,s]}\right]\right)^* \middle| (X_u)_{u\in[0,s]}\right].$$

If X is a martingale, we have $Q(s,T) = \left(\left(\int_s^T f_i(u)f_j(u)d\langle X\rangle_u\right)\right)_{(i,j)\in I^2}$. We recall that a continuous centered semimartingale X is Gaussian if and only if $\langle X\rangle$ is deterministic (see *e.g.* [27]). Hence, this additional hypothesis is equivalent to assume that

$$Q(s,T) \quad \text{is invertible for every } s \in [0,T). \tag{$\mathcal{H}$}$$

The following theorem follows from the same approach as the homologous result in the article by Alili [1] for the Brownian case. It is generalized to the case of a continuous centered Gaussian semimartingale starting from 0.

Theorem 3.2.1. *Under the $(\mathcal{H})$ hypothesis, for any $s \in [0,T)$, and for $\mathbb{P}_{\overline{Z}_T}$-almost every $\overline{y} \in \mathbb{R}^I$, $\mathbb{P}\left[\cdot\middle|\overline{Z}_T = \overline{y}\right]$ is equivalent to $\mathbb{P}$ on $\mathcal{F}_s^X$ and its Radon-Nikodym density is given by*

$$\frac{d\mathbb{P}\left[\cdot\middle|\overline{Z}_T = \overline{y}\right]}{d\mathbb{P}}_{|\mathcal{F}_s^X} = \frac{\Pi_{(X_u)_{u\in[0,s]},T}(\overline{y})}{\Pi_{0,T}(\overline{y})}.$$

Proof: Consider F a real bounded $\mathcal{F}_s^X$-measurable random variable and $\phi : \mathbb{R}^I \to \mathbb{R}$ a bounded Borel function.

- On one hand, preconditioning by $\overline{Z}_T$ yields

$$\mathbb{E}\left[F\phi(\overline{Z}_T)\right] = \mathbb{E}\left[\mathbb{E}\left[F\middle|\overline{Z}_T\right]\phi(\overline{Z}_T)\right] = \int_{\mathbb{R}^I} \phi(\overline{y})\mathbb{E}\left[F\middle|\overline{Z}_T = \overline{y}\right]\Pi_{0,T}(\overline{y})d\overline{y}. \tag{3.13}$$

- On the other hand, as F is measurable with respect to $\mathcal{F}_s^X$, preconditioning with respect to $\mathcal{F}_s^X$ yields

$$\mathbb{E}\left[F\phi(\overline{Z}_T)\right] = \mathbb{E}\left[F\mathbb{E}\left[\phi(\overline{Z}_T)\middle|\mathcal{F}_s^X\right]\right] = \mathbb{E}\left[F\int_{\mathbb{R}^I}\phi(\overline{y})\Pi_{(X_t)_{t\in[0,s]},T}(\overline{y})d\overline{y}\right].$$

  Now, thanks to Fubini's theorem

$$\mathbb{E}\left[F\phi(\overline{Z}_T)\right] = \int_{\mathbb{R}^I}\phi(\overline{y})\mathbb{E}\left[F\Pi_{(X_t)_{t\in[0,s]},T}(\overline{y})\right]d\overline{y}. \tag{3.14}$$

Identifying Equations (3.13) and (3.14), we see that for $\mathbb{P}_{\overline{Z}_T}$-almost surely $\overline{y} \in \mathbb{R}^I$ and for every real bounded $\mathcal{F}_s^X$-measurable random variable F,

$$\mathbb{E}\left[F\middle|\overline{Z}_T = \overline{y}\right] = \mathbb{E}\left[F\frac{\Pi_{(X_t)_{t\in[0,s]},T}(\overline{y})}{\Pi_{0,T}(\overline{y})}\right]. \tag{3.15}$$

Equation (3.15) characterizes the Radon-Nikodym derivative of the probability $\mathbb{P}\left[\cdot\middle|\overline{Z}_T = \overline{y}\right]$ on $\mathcal{F}_s^X$. □

We now can use classical filtration enlargement techniques [12, 14, 29].

Proposition 3.2.2 (Generalized bridges as semimartingales)**.** *Let us define the filtration $\mathcal{G}^X$ by $\mathcal{G}_t^X = \sigma\left(\overline{Z}_T, \mathcal{F}_t^X\right)$, the enlargement of the filtration $\mathcal{F}^X$ corresponding to the above conditioning. We consider the stochastic process $D_s^{\overline{y}} := \frac{d\mathbb{P}\left[\cdot\middle|\overline{Z}_T=\overline{y}\right]}{d\mathbb{P}}_{|\mathcal{F}_s^X} = \frac{\Pi_{(X_t)_{t\in[0,s]},T}(\overline{y})}{\Pi_{0,T}(\overline{y})}$ for $s \in [0,T)$.*

Under the $(\mathcal{H})$ hypothesis, and the assumption that $D^{\overline{y}}$ is continuous, X is a continuous $\mathcal{G}^X$-semimartingale on $[0,T)$.

Proof: $D^{\overline{y}}$ is a strictly positive martingale on $[0,T)$ which is uniformly integrable on every interval $[0,t] \subset [0,T)$. Hence, as we assumed that it is continuous, we can write $D^{\overline{y}}$ as an exponential martingale $D_s^{\overline{y}} = \exp\left(L_s^{\overline{y}} - \frac{1}{2}\left\langle L^{\overline{y}}\right\rangle_s\right)$ with $L_t^{\overline{y}} = \int_0^t \left(D_s^{\overline{y}}\right)^{-1} dD_s^{\overline{y}}$ (as $D_0^{\overline{y}} = 1$).
Now, as X is a continuous $(\mathcal{F}^X, \mathbb{P})$-semimartingale, we write $X = V + M$ its canonical decomposition (under the filtration $\mathcal{F}^X$).

- Thanks to Girsanov theorem, $\widetilde{M^{\overline{y}}} := M - \left\langle M, L^{\overline{y}}\right\rangle$ is a $\left(\mathcal{F}^X, \mathbb{P}\left[\cdot\middle|\overline{Z}_T = \overline{y}\right]\right)$-martingale.
  - A consequence is that it is a $\left(\mathcal{G}^X, \mathbb{P}\left[\cdot\middle|\overline{Z}_T = \overline{y}\right]\right)$-martingale.
  - And thus $\widetilde{M^{\overline{Z}_T}}$ is a $\left(\mathcal{G}^X, \mathbb{P}\right)$-martingale.
    For more preciseness on this, we refer to [2, Theorem 3] where the proof is based on the notion of decoupling measure.
- Moreover, conditionally to $\overline{Z}_T$, V is still a finite-variation process V, and is adapted to $\mathcal{G}^X$. □

Remark (Continuous modification). *In Proposition 3.2.2, if one only assumes that $D^{\overline{y}}$ has a continuous modification $\mathcal{D}^{\overline{y}}$, then with each one of its continuous modifications is associated a continuous $\mathcal{G}^X$-semimartingale on $[0,T)$, and all these semimartingales are modifications of each other.*

Proposition 3.2.3 (Continuity of $D^{\overline{y}}$). *If $\mathcal{F}^X$ is a standard Brownian filtration, then $D^{\overline{y}}$ has a continuous modification.*

Proof: Consider $s \in [0,T)$. Under the $(\mathcal{H})$ hypothesis, the density $\Pi_{(X_u)_{u\in[0,s]},T}$ writes

$$\Pi_{(X_u)_{u\in[0,s]},T}(\overline{y}) = (2\pi \det Q(s,T))^{-\frac{|I|}{2}} \exp\left(\left(\overline{y} - \mathbb{E}\left[\overline{Z}_T \middle| (X_u)_{u\in[0,s]}\right]\right) Q(s,T)^{-1} \left(\overline{y} - \mathbb{E}\left[\overline{Z}_T \middle| (X_u)_{u\in[0,s]}\right]\right)^*\right). \tag{3.16}$$

Let us define the stochastic process $\overline{H}$ by $\overline{H}_s := \mathbb{E}\left[\overline{Z}_T \middle| (X_u)_{u\in[0,s]}\right]$. The so-defined process $\overline{H}$ is a $\mathcal{F}^X$ local martingale. Thanks the Brownian representation theorem, $\overline{H}$ has a Brownian representation and has a continuous modification. The continuity of $s \mapsto \det Q(s,T)$ and $s \mapsto Q(s,T)^{-1}$ follows from the definition of $Q(s,T)$ and the continuity of $\overline{H}$ (up to a modification). Hence, $D^{\overline{y}}$ has a continuous modification. □

Remark.
- *The measurability assumption with respect to a Brownian filtration is satisfied in the cases of the Brownian bridge and Ornstein-Uhlenbeck processes.*
- *This hypothesis is not necessary so long as the continuity of the martingale $\overline{H}_s = \mathbb{E}\left[\overline{Z}_T \middle| (X_u)_{u\in[0,s]}\right]$ can be proved by any means.*

On the canonical decomposition

Observing that $\left\langle M, L^{\overline{y}}\right\rangle = \left\langle X, L^{\overline{y}}\right\rangle$ we can compute the canonical decomposition of $X^{\overline{f},\overline{y}}$. We have

$$L_t^{\overline{y}} = \int_0^t \frac{d\Pi_{(X_u)_{u\in[0,s]},T}(\overline{y})}{\Pi_{(X_u)_{u\in[0,s]},T}(\overline{y})},$$

and

$$\begin{aligned}\ln\left(\Pi_{(X_u)_{u\in[0,s]},T}(\overline{y})\right) = &-\frac{|I|}{2}\ln\left(2\pi\det Q(s,T)\right)\\ &-\frac{1}{2}\left(\overline{y} - \mathbb{E}\left[\overline{Z}_T \middle| (X_u)_{u\in[0,s]}\right]\right) Q(s,T)^{-1}\left(\overline{y} - \mathbb{E}\left[\overline{Z}_T \middle| (X_u)_{u\in[0,s]}\right]\right)^*.\end{aligned}$$

Using that for a semimartingale S, $d\ln S = \frac{dS}{S} - \frac{1}{2}d\left\langle \frac{1}{S}\cdot S\right\rangle$, we obtain

$$\begin{aligned}\frac{d\Pi_{(X_u)_{u\in[0,s]},T}(\overline{y})}{\Pi_{(X_u)_{u\in[0,s]},T}(\overline{y})} &= d\ln\left(\Pi_{(X_u)_{u\in[0,s]},T}(\overline{y})\right) + \begin{pmatrix}\text{finite-variation}\\ \text{process}\end{pmatrix}\\ &= -\tfrac{1}{2}d\left(\left(\overline{y} - \mathbb{E}\left[\overline{Z}_T \middle| (X_u)_{u\in[0,s]}\right]\right) Q^{-1}(s,T)\left(\overline{y} - \mathbb{E}\left[\overline{Z}_T \middle| (X_u)_{u\in[0,s]}\right]\right)^*\right) + (\text{f.-v. p.})\\ &= \left(d\mathbb{E}\left[\overline{Z}_T \middle| (X_u)_{u\in[0,s]}\right]\right) Q^{-1}(s,T)\left(\overline{y} - \mathbb{E}\left[\overline{Z}_T \middle| (X_u)_{u\in[0,s]}\right]\right)^* + (\text{f.-v. p.}).\end{aligned}$$

Hence,

$$d\left\langle X, L^{\overline{y}}\right\rangle_s = d\left\langle X, \mathbb{E}\left[\overline{Z}_T \middle| (X_u)_{u\in[0,\cdot]}\right]\right\rangle_s Q^{-1}(s,T)\left(\overline{y} - \mathbb{E}\left[\overline{Z}_T \middle| (X_u)_{u\in[0,s]}\right]\right)^*.$$

- *In the case where X is a martingale*, owing to the definition of Z_j, we have $\forall j \in I$, $\mathbb{E}\left[Z_T^j \middle| (X_u)_{u\in[0,s]}\right] = \int_0^s f_j(u)dX_u$ so that

$$\begin{aligned}d\left\langle X, L^{\overline{y}}\right\rangle_s &= \left(\overline{f}(s)Q^{-1}(s,T)\left(\overline{y} - \mathbb{E}\left[Z_T^j \middle| (X_u)_{u\in[0,s]}\right]\right)^*\right) d\langle X\rangle_s\\ &= \sum_{i\in I} f_i(s)\sum_{j\in I}\left(Q(s,T)^{-1}\right)_{ij}\left(y_j - \mathbb{E}\left[Z_T^j \middle| (X_u)_{u\in[0,s]}\right]\right) d\langle X\rangle_s.\end{aligned} \tag{3.17}$$

As a consequence, $M - \int_0^{\cdot} \sum_{i\in I} f_i(s) \sum_{j\in I} \left(Q(s,T)^{-1}\right)_{ij} \left(y_j - \mathbb{E}\left[Z_T^j \middle| (X_u)_{u\in[0,s]}\right]\right) d\langle X\rangle_s$ is a $\left(\mathcal{G}^X, \mathbb{P}\left[\cdot \middle| \overline{Z}_t = \overline{y}\right]\right)$-martingale. We have recovered Alili's result on the generalized Brownian bridge [1].

- *In the case where the Gaussian semimartingale X is a Markov process*, for every $j \in I$ there exists $g_j \in L^2([0,T])$ such that $\mathbb{E}\left[Z_T^j \middle| (X_u)_{u\in[0,s]}\right] = \int_0^s f_j(u)dX_u + g_j(s)X_s$. Indeed,

$$\mathbb{E}\left[Z_T^j \middle| (X_u)_{u\in[0,s]}\right] = \int_0^s f_j(u)dX_u + \underbrace{\mathbb{E}\left[\int_s^T f_j(u)dX_u \middle| (X_u)_{u\in[0,s]}\right]}_{:=g_j(s)X_s}.$$

Hence, if one assumes that $(g_j)_{j\in I}$ are finite-variation functions (which is is case when X is either an Ornstein-Uhlenbeck process or a Brownian bridge), we have $d\left\langle X, \mathbb{E}\left[\overline{Z}_T \middle| (X_u)_{u\in[0,\cdot]}\right]\right\rangle_s = \left(\overline{f}(s) + \overline{g}(s)\right) d\langle X\rangle_s$. And thus

$$\begin{aligned} d\left\langle X, L^{\overline{y}}\right\rangle_s &= \left(\left(\overline{f}(s) + \overline{g}(s)\right) Q^{-1}(s,T) \left(\overline{y} - \mathbb{E}\left[Z_T^j \middle| (X_u)_{u\in[0,s]}\right]\right)^*\right) d\langle X\rangle_s \\ &= \sum_{i\in I} \left(f_i(s) + g_i(s)\right) \sum_{j\in I} \left(Q(s,T)^{-1}\right)_{ij} \left(y_j - \mathbb{E}\left[Z_T^j \middle| (X_u)_{u\in[0,s]}\right]\right) d\langle X\rangle_s. \end{aligned}$$

Generalized bridges and functional stratification

With the same set of notations, we set $Y = \overline{Z}_T$ and $\widehat{Y}^\Gamma = \mathrm{Proj}_\Gamma(Y) = \sum_{i=1}^N \gamma_i \mathbf{1}_{C_i}(Y)$ a stationary quantizer of Y (where $\Gamma = \{\gamma_1, \cdots, \gamma_N\}$ and $C = \{C_1, \cdots, C_N\}$ are respectively the associated codebook and Voronoi partition).

Proposition 3.2.4 (Stratification). *Under the $(\mathcal{H})$ hypothesis, for any $s \in [0,T)$, for any $k \in \{1, \cdots, N\}$, $\mathbb{P}\left[\widehat{Y}^\Gamma = \gamma_k\right] > 0$ and the conditional probability $\mathbb{P}\left[\cdot \middle| \widehat{Y}^\Gamma = \gamma_k\right]$ is equivalent to $\mathbb{P}$ on $\mathcal{F}_s^X$.*

Proof: Obviously, if $A \in \mathcal{F}_s^X$ is such that $\mathbb{P}[A] = 0$, we have $\mathbb{P}\left[A \middle| \widehat{Y}^\Gamma = \gamma_k\right] = 0$. Conversely, $B \in \mathcal{F}_s^X$ satisfies $\mathbb{P}\left[B \middle| \widehat{Y}^\Gamma = \gamma_k\right] = 0$, then pre-conditioning by Y, we get $\mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_B | Y\right] \middle| \widehat{Y}^\Gamma = \gamma_k\right] = 0$. Thus, $\int_{\overline{y}\in C_k} \mathbb{P}\left[B | Y = \overline{y}\right] d\mathbb{P}_Y(\overline{y}) = 0$. Hence $\mathbb{P}\left[B | Y = \overline{y}\right] = 0$ for $\mathbb{P}_Y$-almost every $\overline{y} \in C_k$.
Since $\mathbb{P}_Y(C_k) > 0$, there exists at least an $\overline{y} \in C_k$ such that $\mathbb{P}\left[B | Y = \overline{y}\right] = 0$. Now thanks to Theorem 3.2.1, $\mathbb{P}[B] = 0$. □

Proposition 3.2.5 (Stratification). *Let us define the filtration $\mathcal{G}^X$ by $\mathcal{G}_t^X = \sigma\left(\mathcal{F}_t^X, \widehat{Y}^\Gamma\right)$, the enlargement of $\mathcal{F}^X$ corresponding to the conditioning with respect to $\widehat{Y}^\Gamma$. For $k \in \{1, \cdots, N\}$, we consider the stochastic process $D_s^{\gamma_k} := \frac{d\mathbb{P}\left[\cdot \middle| \widehat{Y}^\Gamma = \gamma_k\right]}{d\mathbb{P}}_{|\mathcal{F}_s^X}$ for $s \in [0,T)$.*

Under the $(\mathcal{H})$ hypothesis, and the assumption that D^{γ_k} is continuous, the conditional distribution $\mathcal{L}\left(X \middle| \widehat{Y}^\Gamma\right)$ of X knowing in which Voronoi cell $\overline{Z}_T$ falls, is the probability distribution of a $\mathcal{G}^X$-semimartingale on $[0,T)$.

Proof: Using that $\mathbb{P}\left[\cdot \middle| \widehat{Y}^\Gamma = \gamma_k\right]$ is equivalent to $\mathbb{P}$ on $\mathcal{F}_s^X$, thanks to Proposition 3.2.4, we can *mutatis mutandis* use the same arguments as for Proposition 3.2.2, $\mathbb{P}\left[\cdot \middle| \overline{Z}_T = \overline{y}\right]$ being replaced by $\mathbb{P}\left[\cdot \middle| \widehat{Y}^\Gamma = \gamma_k\right]$.

D^{γ_k} is a strictly positive martingale on $[0,T)$ uniformly integrable on every $[0,t] \subset [0,T)$. Hence, as D^{γ_k} is continuous by hypothesis, it is an exponential martingale $D_s^{\gamma_k} = \exp\left(L_s^{\gamma_k} - \frac{1}{2}\langle L^{\gamma_k}\rangle_s\right)$,

with $L_t^{\gamma_k} = \int_0^t (D_s^{\gamma_k})^{-1} dD_s^{\gamma_k}$ (as $D_0^{\gamma_k} = 1$). Now, as X is a continuous $(\mathcal{F}^X, \mathbb{P})$-semimartingale, we write $X = V + M$ its canonical decomposition (under the filtration $\mathcal{F}^X$).

- Thanks to Girsanov theorem, $\widetilde{M}^{\gamma_k} := M - \langle M, L^{\gamma_k} \rangle$ is a $\left(\mathcal{F}^X, \mathbb{P}\left[\cdot \middle| \widehat{Y}^\Gamma = \gamma_k\right]\right)$-martingale. As a consequence, it is a $\left(\mathcal{G}^X, \mathbb{P}\left[\cdot \middle| \widehat{Y}^\Gamma = \gamma_k\right]\right)$-martingale and thus $\widetilde{M}^{\widehat{Y}^\Gamma}$ is a $(\mathcal{G}^X, \mathbb{P})$-martingale.
- Moreover, conditionally to $\widehat{Y}^\Gamma$, V is still a finite-variation process V, and is adapted to $\mathcal{G}^X$. □

Proposition 3.2.6 (Continuity of D^{γ_k})**.** *If $\mathcal{F}^X$ is a Brownian filtration, then D^{γ_k} has a continuous modification.*

Proof: By definition, D^{γ_k} is a $\mathcal{F}^X$-local martingale on $[0, T]$. The conclusion is a straightforward consequence of the Brownian representation theorem. □

Considering the partition of $L^2([0, T])$ corresponding to the Voronoi cells of a functional quantizer of X, the last two propositions show that the conditional distribution of the X in each Voronoi cell (strata) is a Gaussian semimartingale with respect to its own filtration. This allows us to define the corresponding functional stratification of the solutions of stochastic differential equations driven by X.

In [7], an algorithm is proposed to simulate the conditional distribution of the marginals $(X_{t_0}, \cdots, X_{t_n})$ of X for a given subdivision $0 = t_0 < t_1 < \cdots < t_n = T$ of $[0, T]$ conditionally to a given Voronoi cell (strata) of a functional quantization of X. The simulation complexity has an additional linear complexity to an unconditioned simulation of $(X_{t_0}, \cdots, X_{t_n})$. We refer to [7] for more details.

To deal with solutions of SDE, it was proposed in [7] to simply plug these marginals in the Euler scheme of the SDE. Proposition 3.2.5 now shows that this amounts to simulate the Euler scheme of the SDE driven by the corresponding (non-Gaussian) semimartingale.

3.2.4 About the $(\mathcal{H})$ hypothesis

The martingale case

In the case where X is a continuous Gaussian martingale, the matrix $Q(s, t)$ defined in Section 3.2.3 writes $Q(s, t) = \left(\left(\int_s^t f_i(u) f_j(u) d\langle X \rangle_u\right)\right)_{(i,j) \in I^2}$.

For $1 \leq s < t \leq T$, the map $(\cdot|\cdot) : (f, g) \mapsto \int_s^t f(u) g(u) d\langle X \rangle_u$ defines a scalar product on $L^2([s, t], d\langle X \rangle)$. Hence $Q(s, t)$ is the Gram matrix of the vectors of $L^2([s, t], d\langle X \rangle)$ defined by the restrictions to $[s, t]$ of the functions $(f_i)_{i \in I}$. Thus, it is invertible if and only if these restrictions form a linearly independent family of $L^2([s, t], d\langle X \rangle)$. (Another consequence, is that if $Q(s, t)$ is invertible for some $0 \leq s < t \leq T$, then for every (u, v) such that $[s, t] \subset [u, v]$, $Q(u, v)$ is invertible).

For instance, if X is a standard Brownian motion on $[0, T]$, the functions $(f_i^X)_{i \in I}$ (associated with the Karhunen-Loève decomposition) are trigonometric functions with strictly different frequencies. Hence, they form a linearly independent family of continuous functions on every nonempty interval $[s, T) \subset [0, T)$. Moreover, the measure $d\langle X \rangle$ is proportional to the Lebesgue measure on $[0, T]$ and thus $Q(s, T)$ is invertible for any $s \in [0, T)$. *Hence, the $(\mathcal{H})$ hypothesis is fulfilled in the case of K-L generalized bridges of the standard Brownian motion.*

The standard Brownian bridge and Ornstein-Uhlenbeck processes

The Brownian bridge and the Ornstein-Uhlenbeck process are not martingales. Hence, this criterion is not sufficient and the invertibility of matrix $Q(s, T)$ has to be proved by other means.

Following from the definitions of $Q(s,T)$ and $\overline{Z}_T$, in the case of the K-L generalized bridge

$$\begin{aligned} Q(s,T)_{ij} &= \mathbb{E}\Big[\Big(\int_s^T f_i^X(u)dX_u - \mathbb{E}\Big[\int_s^T f_i^X(u)dX_u \Big| (X_u)_{u\in[0,s]}\Big]\Big) \\ &\qquad \times \Big(\int_s^T f_j^X(u)dX_u - \mathbb{E}\Big[\int_s^T f_j^X(u)dX_u \Big| (X_u)_{u\in[0,s]}\Big]\Big)^* \Big| (X_u)_{u\in[0,s]}\Big] \\ &= \operatorname{cov}\Big(\int_s^T f_i^X(u)dX_u^{(s)}, \int_s^T f_j^X(u)dX_u^{(s)}\Big), \end{aligned} \tag{3.18}$$

where $\left(X_u^{(s)}\right)_{u\in[s,T]}$ has the conditional distribution of X knowing $(X_u)_{u\in[0,s]}$.

- When X is a standard Brownian bridge on $[0,T]$, $X_u^{(s)}$ is a Brownian bridge on $[s,T]$, starting from X_s and arriving at 0.

  It is the sum of an affine function and a standard centered Brownian bridge on $[s,T]$.

- When X is a centered Ornstein-Uhlenbeck process, $X_u^{(s)}$ is an Ornstein-Uhlenbeck process on $[s,T]$ starting from X_s, with the same mean reversion parameter as X.

  It is also the sum of a deterministic function and an Ornstein-Uhlenbeck process starting from 0.

As a consequence, in these two cases, the quantity $\operatorname{cov}\left(\int_s^T f_i^X(u)dX_u^{(s)}, \int_s^T f_j^X(u)dX_u^{(s)}\right)$ can be computed by plugging either a centered Brownian bridge on $[s,T]$ or an Ornstein-Uhlenbeck starting from 0 instead of $X^{(s)}$ in Equation (3.18). This means that $Q(s,T)$ is the Gram matrix of the random variables $\left(\int_s^T f_i^X(u)dG_u\right)_{i\in I}$, where the centered Gaussian process $(G_u)_{u\in[s,T]}$ is either a standard Brownian bridge on $[s,T]$ or an Ornstein-Uhlenbeck process starting from 0 at s. Thus it is singular if and only if there exists $(\alpha_i)_{i\in I} \neq 0$ in $\mathbb{R}^I$ such that

$$\int_s^T \underbrace{\left(\sum_{i\in I} \alpha_i f_i^X(u)\right)}_{:=g(u)} dG_u = 0 \quad a.s.. \tag{3.19}$$

The case of the Brownian bridge
In the case where X is the standard Brownian bridge on $[0,T]$, functions $(f_i^X)_{i\in I}$ are C^∞ functions and G is a standard Brownian bridge on $[s,T]$. An integration by parts gives $\int_s^T G_s g'(s)ds = 0 \quad a.s.$ and thus $g' \equiv 0$ on (s,T) and thus g is constant on $[s,T]$. The functions $(f_i^X)_{i\in I}$ form a linearly independent set of functions and, as they are trigonometric functions with different frequencies, they clearly don't span constant functions, so that Equation (3.19) yields $\alpha_1 = \cdots = \alpha_n = 0$. *Hence the ($\mathcal{H}$) hypothesis is fulfilled in the case of K-L generalized bridges of the standard Brownian bridge.*

The case of Ornstein-Uhlenbeck processes
In the case where X is an Ornstein-Uhlenbeck process on $[0,T]$, G is an Ornstein-Uhlenbeck process on $[s,T]$ starting from 0. The injectivity property of the Wiener integral related to the Ornstein-Uhlenbeck process stated in Proposition 3.2.7 below, applied on $[s,T]$, shows that Equation (3.19) amounts to $g \overset{L^2([s,T],dt)}{=} 0$ and thus

$$\sum_{i\in I} \alpha_i f_i^X \overset{L^2([s,T],dt)}{=} 0. \tag{3.20}$$

Again, as $(f_i^X)_{i\in I}$ are linearly independent, we have $\alpha_1 = \cdots = \alpha_n = 0$. *Hence the ($\mathcal{H}$) hypothesis is fulfilled in the case of K-L generalized bridges of the Ornstein-Uhlenbeck processes.*

Proposition 3.2.7 (Injectivity of the Wiener integral related to centered Ornstein-Uhlenbeck processes)**.** *Let G be an Ornstein-Uhlenbeck process defined on $[0,T]$ by the SDE*

$$dG_t = -\theta G_t dt + \sigma dW_t \quad \text{with } \sigma > 0 \text{ and } \theta > 0,$$

where W is a standard Brownian motion and $G_0 \overset{\mathcal{L}}{\sim} \mathcal{N}\left(0, \sigma_0^2\right)$ is independent of W.
If $g \in L^2([0,T])$, then we have

$$\int_0^T g(s)dG_s = 0 \quad \Leftrightarrow \quad g \overset{L^2([0,T])}{=} 0.$$

Proof: The solution of the Ornstein-Uhlenbeck SDE is

$$G_t = \underbrace{G_0 e^{-\theta t}}_{\text{independent of } W} \overset{\perp\!\!\!\perp}{+} \underbrace{\int_0^t \sigma e^{\theta(s-t)} dW_s}_{:=G_t^0}.$$

The so-defined process $\left(G_t^0\right)_{t\in[0,T]}$ is a centered Ornstein-Uhenbeck process starting from 0 and satisfying the same SDE as G. Hence, we have

$$\int_0^T g(s)dG_s = -\theta G_0 \int_0^T g(s)e^{-\theta s}ds \overset{\perp\!\!\!\perp}{+} \int_0^T g(s)dG_s^0.$$

Thus, by independence, if $\int_0^T g(s)dG_s = 0$ then $\int_0^T g(s)dG_s^0 = 0$. This means that we only have to prove the proposition in the case of an Ornstein-Uhlenbeck process starting from 0.

We now assume that $\sigma_0^2 = 0$ and we temporarily make the additional assumption that $\theta T < \frac{4}{3}$. If $g \in L^2([0,T])$ and $\int_0^T g(s)dG_s = 0$, then $\theta \int_0^T g(s)G_s ds = \sigma \int_0^T g(s)dW_s$, and thus, if Γ^{OU} denotes the covariance function of G,

$$\theta^2 \int_0^T \int_0^T g(s)g(t)\Gamma^{OU}(s,t)dsdt = \sigma^2 \int_0^T g(s)^2 ds. \tag{3.21}$$

Applying Schwarz's inequality twice, we get

$$\int_0^T \int_0^T g(s)g(t)\Gamma^{OU}(s,t)dsdt \leq \int_0^T g(t)^2 dt \sqrt{\int_0^T \int_0^T \left(\Gamma^{OU}(s,t)\right)^2 dsdt}.$$

Moreover, provided that

$$\int_0^T \int_0^T \left(\Gamma^{OU}(s,t)\right)^2 dsdt < \frac{\sigma^4}{\theta^4}, \tag{3.22}$$

Equality (3.21) implies $\int_0^T g(s)^2 ds = 0$.

Now, we come to the proof of Inequality (3.22). The covariance function of the Ornstein-Uhlenbeck process starting from 0 writes

$$\Gamma^{OU}(s,t) = \frac{\sigma^2}{2\theta} e^{-\theta(s+t)} \left(e^{2\theta \min(s,t)} - 1\right).$$

If $t \in [0,T]$, we have

$$\begin{aligned} \int_0^T \left(\Gamma^{OU}(s,t)\right)^2 ds &= \int_0^t \left(\Gamma^{OU}(s,t)\right)^2 ds + \int_t^T \left(\Gamma^{OU}(s,t)\right)^2 ds \\ &= \tfrac{\sigma^4}{8\theta^3}\left(2 - 4e^{-2\theta t}\theta t - e^{-2\theta(T-t)} - 2e^{-2\theta t} + 2e^{-2\theta T} - e^{-2\theta(T+t)}\right), \end{aligned}$$

and thus

$$\int_0^T \int_0^T \left(\Gamma^{OU}(s,t)\right)^2 dsdt = \frac{\sigma^2}{16\theta^4}\left(-5 + 4\theta T + 8\theta T e^{-2\theta T} + 4e^{-2\theta T} + e^{-4\theta T}\right).$$

Consequently, the function ϕ defined by $\phi(\theta) := \int_0^T \int_0^T \left(\Gamma^{OU}(s,t)\right)^2 dsdt - \frac{\sigma^4}{\theta^4}$ writes

$$\phi(\theta) = \frac{1}{16}\frac{\sigma^4}{\theta^4}\left(-21 + 4\theta T + 8\theta e^{-2\theta T} T + 4e^{-2\theta T} + e^{-4\theta T}\right).$$

We have $\phi(\theta) < -16 + 12\theta T$ which leads to Inequality (3.22) thanks to the fact that $\theta T < \frac{4}{3}$.

We now come back to the general case where we might have $\theta T \geq \frac{4}{3}$. If this is the case, let us consider $\widetilde{T} := T - \frac{1}{\theta}$, so that $\theta\left(T - \widetilde{T}\right) < \frac{4}{3}$. For $t \in \left[\widetilde{T}, T\right]$, we have

$$G_t = \underbrace{G_{\widetilde{T}} e^{-\theta\left(t-\widetilde{T}\right)}}_{\text{independent of } (W_s)_{s\in\left[\widetilde{T},T\right]}} \overset{\perp}{+} \underbrace{\int_{\widetilde{T}}^{t} \sigma e^{\theta(s-t)} dW_s}_{:=\widetilde{G}_t^0}.$$

The so-defined process $\left(\widetilde{G}_t^0\right)_{t\in\left[\widetilde{T},T\right]}$ is a centered Ornstein-Uhenbeck process starting from 0 and satisfying the same SDE as G. Hence, by independence, if $\int_0^T g(s)dG_s = 0$, then $\int_{\widetilde{T}}^T g(s)d\widetilde{G}_s^0 = 0$.

As $\theta\left(T - \widetilde{T}\right) < \frac{4}{3}$, we can apply the result to $\left(\widetilde{G}_t^0\right)_{t\in\left[\widetilde{T},T\right]}$ so that $g_{|\left[\widetilde{T},T\right]} \overset{L^2([\widetilde{T},T])}{=} 0$. If $\widetilde{T}\theta < \frac{4}{3}$, we then have $g \overset{L^2([0,T])}{=} 0$. If it is not the case, we use the same method by using the decomposition of $\left[0, \widetilde{T}\right]$ into $\left[0, \widetilde{T} - \frac{1}{\theta}\right]$ and $\left[\widetilde{T} - \frac{1}{\theta}, \widetilde{T}\right]$ and so on. An easy induction finally shows that $g \overset{L^2([0,T])}{=} 0$.

The inverse implication is obvious. □

The case of a more general Gaussian semimartingale
In Appendix 3.A, we investigate the problem for more general Gaussian processes.

3.3 K-L generalized bridges and partial functional quantization

We keep the notations and assumptions of Section 3.2.2. As we have seen, Equation (3.11) decomposes the process X as the sum of a linear combination of $Y := (Y_i)_{i\in I}$ and an independent remainder term. We now consider $\widehat{Y}^\Gamma$ a stationary Voronoi N-quantization of Y. $\widehat{Y}^\Gamma$ can be written as a nearest neighbor projection of Y on a finite codebook $\Gamma = (\gamma_1, \cdots, \gamma_N)$.

$$\widehat{Y}^\Gamma = \mathrm{Proj}_\Gamma(Y), \qquad \text{where } \mathrm{Proj}_\Gamma \text{ is a nearest neighbor projection on } \Gamma.$$

For example, $\widehat{Y}^\Gamma$ can be a stationary product quantization or an optimal quadratic quantization of Y. We now define the stochastic process $\widetilde{X}^{I,\Gamma}$ by replacing Y by $\widehat{Y}^\Gamma$ in the decomposition (3.11). We denote $\widetilde{X}^{I,\Gamma} = \mathrm{Proj}_{I,\Gamma}(X)$.

$$\widetilde{X}^{I,\Gamma} = \sum_{i\in I} \widehat{Y}_i^\Gamma e_i^X \overset{\perp}{+} \sum_{i\in\mathbb{N}^*\setminus I} \sqrt{\lambda_i^X}\xi_i e_i^X.$$

The conditional distribution of $\widetilde{X}^{I,\Gamma}$ given that Y falls in the Voronoi cell of γ_k is the probability distribution of the K-L generalized bridge with end-point γ_k. In other words, we have quantized the Karhunen-Loève coordinates of X corresponding to $i \in I$, and not the other ones.
The so-defined process $\widetilde{X}^{I,\Gamma}$ is called a *partial functional quantization of* X.

3.3.1 Partial functional quantization of stochastic differential equations

Let X be a continuous centered Gaussian semimartingale on $[0,T]$ with $X_0 = 0$. We consider the SDE

$$dS_t = b(t, S_t)dt + \sigma(t, S_t)dX_t, \qquad S_0 = x \in \mathbb{R}, \qquad \text{and } t \in [0,T], \tag{3.23}$$

where $b(t,x)$ and $\sigma(t,x)$ are Borel functions, Lipschitz continuous with respect to x uniformly in t, σ and $|b(\cdot,0)|$ are bounded. This SDE admits a unique strong solution S.

The conditional distribution given that $Y_i = y_i$ for $i \in I$ of S is the strong solution of the stochastic differential equation $dS_t = b(t,S_t)dt + \sigma(t,S_t)dX_t^{I,\overline{y}}$, with $S_0 = x \in \mathbb{R}$, and for $t \in [0,T]$, where $X_t^{I,\overline{y}}$ is the corresponding K-L generalized bridge.

Under the $(\mathcal{H})$ hypothesis, this suggests to define the partial quantization of S from a partial quantization $\widetilde{X}^{I,\Gamma}$ of X by replacing X by $\widetilde{X}^{I,\Gamma}$ in the SDE (3.23). We define the *partial quantization* $\widetilde{S}^{I,\Gamma}$ as the process whose conditional distribution given that Y falls in the Voronoi cell of γ_k is the strong solution of the same SDE where X is replaced by the K-L generalized bridge with end-point γ_k. We write

$$d\widetilde{S}_t^{I,\Gamma} = b\left(t, \widetilde{S}_t^{I,\Gamma}\right) dt + \sigma\left(t, \widetilde{S}_t^{I,\Gamma}\right) d\widetilde{X}_t^{I,\Gamma}. \tag{3.24}$$

3.3.2 Convergence of partially quantized SDE

We start by stating some useful inequalities for the sequel. Then we recall the so-called Zador's theorem which will be used in the proof of the *a.s.* convergence of partially quantized SDE.

Lemma 3.3.1 (Gronwall inequality for locally finite measures)**.** *Consider $\mathcal{I}$ an interval of the form $[a,b)$ or $[a,b]$ with $a<b$ or $[a,\infty)$. Let μ be a locally finite measure on the Borel σ-algebra of $\mathcal{I}$. We consider u a measurable function defined on $\mathcal{I}$ such that for all $t \in \mathcal{I}$, $\int_a^t |u(s)|\mu(ds) < +\infty$. We assume that there exists a Borel function ψ on $\mathcal{I}$ such that*

$$u(t) \leq \psi(t) + \int_{[a,t)} u(s)\mu(ds), \quad \forall t \in \mathcal{I}.$$

If | *either ψ is non-negative,*
| *or $t \mapsto \mu([a,t))$ is continuous on $\mathcal{I}$ and for all $t \in \mathcal{I}$, $\int_a^t |\psi(s)|\mu(ds) < \infty$,*

then u satisfies the Gronwall inequality.

$$u(t) \leq \psi(t) + \int_{[a,t)} \psi(s)\exp(\mu([s,t)))\mu(ds).$$

A proof of this result is available in [9, Appendix 5.1].

Lemma 3.3.2 (A Gronwall-like inequality in the non-decreasing case)**.** *Consider $\mathcal{I}$ an interval of the form $[a,b)$ or $[a,b]$ with $a<b$ or $[a,\infty)$. Let μ be a locally finite measure on the Borel σ-algebra of $\mathcal{I}$. We consider u a measurable* non-decreasing *function defined on $\mathcal{I}$ such that for all $t \in \mathcal{I}$, $\int_a^t |u(s)|\mu(ds) < +\infty$. We assume that there exists a Borel function ψ on $\mathcal{I}$, and two non-negative constants $(A,B) \in \mathbb{R}_+^2$ such that*

$$u(t) \leq \psi(t) + A\int_{[a,t)} u(s)\mu(ds) + B\sqrt{\int_{[a,t)} u(s)^2\mu(ds)}, \quad \forall t \in \mathcal{I}. \tag{3.25}$$

If | *either ψ is non-negative,*
| *or $t \mapsto \mu([a,t))$ is continuous on $\mathcal{I}$ and for all $t \in \mathcal{I}$, $\int_a^t |\psi(s)|\mu(ds) < \infty$,*

then u satisfies the following Gronwall inequality.

$$u(t) \leq 2\psi(t) + 2\left(2A + B^2\right)\int_{[a,t)} \psi(s)\exp\left(\left(2A+B^2\right)\mu([s,t))\right)\mu(ds).$$

Proof: Using that for $(x,y) \in \mathbb{R}_+^2$, $\sqrt{xy} \leq \frac{1}{2}\left(\frac{x}{B} + By\right)$, we have

$$\left(\int_{[a,t)} u(s)^2 \mu(ds)\right)^{\frac{1}{2}} \leq \left(u(t)\int_{[a,t)} u(s)\mu(ds)\right)^{\frac{1}{2}} \leq \frac{u(t)}{2B} + \frac{B}{2}\int_{[a,t)} u(s)\mu(ds).$$

Plugging this in the Inequality (3.25) yields

$$u(t) \leq 2\psi(t) + \left(2A + B^2\right)\int_{[a,t)} u(s)\mu(ds).$$

Applying the regular Gronwall's inequality (Lemma 3.3.1) yields the announced result. □

Theorem 3.3.3 (Zador, Bucklew, Wise, Graf, Luschgy, Pagès)**.**

1. (Sharp rate) *Consider $r > 0$, and X be a $\mathbb{R}^d$-valued random variable such that $X \in L^{r+\eta}$ for some $\eta > 0$. Let $\mathbb{P}_X(d\xi) = \phi(\xi)d\xi + \nu(d\xi)$ be the Radon-Nikodym decomposition of the probability distribution of X. (ν and the Lebesgue's measure are singular). Then if $\phi \not\equiv 0$,*

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

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

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

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

The first statement of the theorem was first established for probability distributions with compact support by Zador [30], and extended by Bucklew and Wise to general probability distributions on $\mathbb{R}^d$ [5]. The first mathematically rigorous proof can be found in [11]. The proof of the second statement is available in [19].
The real constant $\widetilde{J}_{r,d}$ corresponds to the case of the uniform probability distribution over the unit hypercube $[0,1]^d$. We have $\widetilde{J}_{r,1} = \frac{1}{2}(r+1)^{-\frac{1}{r}}$ and $\widetilde{J}_{2,2} = \sqrt{\frac{5}{18\sqrt{2}}}$ (see [11].)

L^p convergence of partially quantized SDE

Lemma 3.3.4 (Generalized Minkowski inequality for locally finite measures)**.** *Consider $\mathcal{I}$ an interval of the form $[a,b)$ or $[a,b]$ with $a < b$ or $[a,\infty)$. Let μ be a locally finite measure on the Borel σ-algebra of $\mathcal{I}$. Then for any non-negative bi-measurable process $X = (X_t)_{t\in\mathcal{I}}$ and every $p \in [1,\infty)$,*

$$\left\|\int_{\mathcal{I}} X_t\mu(dt)\right\|_p \leq \int_{\mathcal{I}} \|X_t\|_p\mu(dt).$$

Proposition 3.3.5 (Burkholder-Davis-Gundy inequality)**.** *For every $p \in (0,\infty)$, there exist two positive real constants c_p^{BDG} and C_p^{BDG} such that for every continuous local martingale $(X_t)_{t\in[0,T]}$ null at 0,*

$$c_p^{BDG}\left\|\sqrt{\langle X\rangle_T}\right\|_p \leq \left\|\sup_{s\in[0,T]}|X_s|\right\|_p \leq C_p^{BDG}\left\|\sqrt{\langle X\rangle_T}\right\|_p.$$

We refer to [27] for a detailed proof.

Proposition 3.3.6 (L^p inequality)**.** *Let G be a standard Gaussian random variable valued in $\mathbb{R}$. There exists a constant $C_p > 0$ such that for every $M > 1$*

$$\sqrt{\frac{2}{\pi}}M^{p-1}\exp\left(-\frac{M^2}{2}\right) \leq \mathbb{E}\left[|G|^p\mathbf{1}_{|G|>M}\right] \leq C_pM^{p-1}\exp\left(-\frac{M^2}{2}\right).$$

Consequently

$$\left(\sqrt{\frac{2}{\pi}}\right)^{\frac{1}{p}} M^{\frac{1}{q}} \exp\left(-\frac{M^2}{2p}\right) \leq \left\|G\mathbf{1}_{|G|>M}\right\|_p \leq (C_p)^{1/p} M^{\frac{1}{q}} \exp\left(-\frac{M^2}{2p}\right),$$

where q is the conjugate exponent of p.

Proposition 3.3.7 (The non-standard case and L^p reverse inequality)**.** *If $H := \sigma G$ has a variance of σ^2, we obtain*

$$\begin{aligned}\left\|H\mathbf{1}_{|H|>M}\right\|_p &\leq \sigma\left\|G\mathbf{1}_{|G|>\frac{M}{\sigma}}\right\|_p = \sigma(C_p)^{1/p}\left(\tfrac{M}{\sigma}\right)^{\frac{1}{q}}\exp\left(-\tfrac{M^2}{2p\sigma^2}\right),\\ &= \underbrace{\sigma^{\frac{1}{p}}(C_p)^{1/p}M^{\frac{1}{q}}\exp\left(-\frac{M^2}{2p\sigma^2}\right)}_{:=\eta_M}.\end{aligned} \tag{3.26}$$

Conversely, for some fixed $\eta > 0$, and if $M > 1$, we have

$$M \geq \underbrace{\sqrt{-\sigma^2(p-1)\mathcal{W}\left(-\frac{q\eta^{2q}}{p\sigma^2\left(C_p^{2q/p}\sigma^{2q/p}\right)}\right)}}_{:=M_\eta} \Rightarrow \eta_M \leq \eta \tag{3.27}$$

where $\mathcal{W}$ is the Lambert $\mathcal{W}$ function.

Theorem 3.3.8 (L^p quantization of partially quantized SDE)**.** *Let X be a continuous centered Gaussian martingale on $[0,T]$ with $X_0 = 0$. Let S be the strong solution of the SDE*

$$dS_t = b(t,S_t)dt + \sigma(t,S_t)dX_t, \quad S_0 = x,$$

where $b(t,x)$ and $\sigma(t,x)$ are Borel functions, Lipschitz continuous with respect to x uniformly in t, σ and $|b(\cdot,0)|$ are bounded.
We consider $\widetilde{X}^{I,\Gamma}$ a stationary partial functional quantization of X and $\widetilde{S}^{I,\Gamma}$ the corresponding partial functional quantization of S, i.e. the strong solutions of

$$d\widetilde{S}_t^{I,\Gamma} = b\left(t,\widetilde{S}_t^{I,\Gamma}\right)dt + \sigma\left(t,\widetilde{S}_t^{I,\Gamma}\right)d\widetilde{X}_t^{I,\Gamma}, \quad \widetilde{S}_0^{I,\Gamma} = x.$$

Then, for every $p \in (0,\infty)$, $\varepsilon > 0$ and $t \in [0,T)$, there exists a positive constant $K^X_{p,\varepsilon,t,I}$ such that

$$\left\|\sup_{v\in[0,t]}\left|S_v - \widetilde{S}_v^{I,\Gamma}\right|\right\|_p \leq K^X_{p,\varepsilon,t,I}\left(\left\|Y - \widehat{Y}^\Gamma\right\|_{p+\varepsilon}\right), \tag{3.28}$$

where Y is defined from X by Equation (3.11) and $\widehat{Y}^\Gamma$ is the nearest neighbor projection on Γ.

Proof: We decompose the process X into $X_t = \sum_{i\in I} Y_i e_i^X(t) + X_t^{I,\overline{0}}$ and $\widetilde{X}^{I,\Gamma}$ into $\widetilde{X}_t^{I,\Gamma} = \sum_{i\in I}\widehat{Y}_i^\Gamma e_i^X(t) + X_t^{I,\overline{0}}$, where $\widehat{Y}^\Gamma$ is the nearest neighbor projection of Y on Γ.

For some $k \in \{1,\cdots,N\}$, conditionally to $\widehat{Y}^\Gamma = \gamma_k$, we have

$$\begin{aligned}S_t - \widetilde{S}_t^{I,\Gamma} = &\int_0^t\left(b(u,S_u) - b\left(u,\widetilde{S}_t^{I,\Gamma}\right)\right)du + \sum_{i\in I}\int_0^t\left(\sigma(u,S_u) - \sigma\left(u,\widetilde{S}_u^{I,\Gamma}\right)\right)\widehat{Y}_i^\Gamma de_i^X(u)\\ &+ \sum_{i\in I}\int_0^t\left(Y_i - \widehat{Y}_i^\Gamma\right)\sigma(u,S_u)de_i^X(u) + \int_0^t\left(\sigma(u,S_u) - \sigma\left(u,\widetilde{S}_u^{I,\Gamma}\right)\right)G_u d\langle X\rangle_u\\ &+ \int_0^t\left(\sigma(u,S_u) - \sigma\left(u,\widetilde{S}_u^{I,\Gamma}\right)\right)d\widetilde{M}_u.\end{aligned}$$

This gives (conditionally to $\widehat{Y}^\Gamma = \gamma_k$)

$$\left|S_t - \widetilde{S}_t^{I,\Gamma}\right| \leq [b]_{\mathrm{Lip}} \int_0^t \left|S_u - \widetilde{S}_u^{I,\Gamma}\right| du + [\sigma]_{\mathrm{Lip}}|I| \max_{\substack{i\in I\\ u\in[0,T]}} \left|\left(e_i^X\right)'(u)\right| \left(\max_{i\in I}\left|\widehat{Y}_i^\Gamma\right|\right) \int_0^t \left|S_u - \widetilde{S}_u^{I,\Gamma}\right| du$$
$$+ [\sigma]_{\max}|I| \max_{\substack{i\in I\\ u\in[0,T]}} \left|\left(e_i^X\right)'(u)\right| T \sum_{i\in I} \left|Y_i - \widehat{Y}_i^\Gamma\right| + \left|\int_0^t \left(\sigma(u,S_u) - \sigma\left(u, \widetilde{S}_u^{I,\Gamma}\right)\right) G_u d\langle X\rangle_u\right|$$
$$+ \left|\int_0^t \left(\sigma(u,S_u) - \sigma\left(u, \widetilde{S}_u^{I,\Gamma}\right)\right) d\widetilde{M}_u\right|.$$

As a consequence, conditionally to $\widehat{Y}^\Gamma = \gamma_k$,

$$\max_{v\in[0,t]} \left|S_v - \widetilde{S}_v^{I,\Gamma}\right| \leq [b]_{\mathrm{Lip}} \int_0^t \max_{v\in[0,u]} \left|S_v - \widetilde{S}_v^{I,\Gamma}\right| du$$
$$+ [\sigma]_{\mathrm{Lip}}|I| \max_{\substack{i\in I\\ u\in[0,T]}} \left|\left(e_i^X\right)'(u)\right| \left(\max_{i\in I}\left|\widehat{Y}_i^\Gamma\right|\right) \int_0^t \max_{v\in[0,u]} \left|S_v - \widetilde{S}_v^{I,\Gamma}\right| du$$
$$+ [\sigma]_{\max}|I| \max_{\substack{i\in I\\ u\in[0,T]}} \left|\left(e_i^X\right)'(u)\right| T \sum_{i\in I} \left|Y_i - \widehat{Y}_i^\Gamma\right| + \max_{v\in[0,t]} \left|\int_0^v \left(\sigma(u,S_u) - \sigma\left(u, \widetilde{S}_u^{I,\Gamma}\right)\right) G_u d\langle X\rangle_u\right|$$
$$+ \max_{v\in[0,t]} \left|\int_0^v \left(\sigma(u,S_u) - \sigma\left(u, \widetilde{S}_u^{I,\Gamma}\right)\right) d\widetilde{M}_u\right|.$$

To shorten the notations, we denote, for a random variable V and a non-negligible event A, $\|V\|_{p,A} := \mathbb{E}\left[V^p|A\right]^{1/p}$. Hence, using the Minkowski inequality and the generalized Minkowski inequality for locally finite measures (Lemma 3.3.4), we get

$$\left\|\max_{v\in[0,t]} \left|S_v - \widetilde{S}_v^{I,\Gamma}\right|\right\|_{p,\{\widehat{Y}^\Gamma=\gamma_k\}} \leq [b]_{\mathrm{Lip}} \int_0^t \left\|\max_{v\in[0,u]} \left|S_v - \widetilde{S}_v^{I,\Gamma}\right|\right\|_{p,\{\widehat{Y}^\Gamma=\gamma_k\}} du$$
$$+ [\sigma]_{\mathrm{Lip}}|I| \max_{\substack{i\in I\\ u\in[0,T]}} \left|\left(e_i^X\right)'(u)\right| \left(\max_{i\in I}\left|\widehat{Y}_i^\Gamma\right|\right) \int_0^t \left\|\max_{v\in[0,u]} \left|S_v - \widetilde{S}_v^{I,\Gamma}\right|\right\|_{p,\{\widehat{Y}^\Gamma=\gamma_k\}} du$$
$$+[\sigma]_{\mathrm{Lip}}|I| \max_{\substack{i\in I\\ u\in[0,T]}} \left|\left(e_i^X\right)'(u)\right| T \left\|\sum_{i\in I} \left|Y_i - \widehat{Y}_i^\Gamma\right|\right\|_{p,\{\widehat{Y}^\Gamma=\gamma_k\}} + \left\|\max_{v\in[0,t]} \left|\int_0^v \left(\sigma(u,S_u) - \sigma\left(u, \widetilde{S}_u^{I,\Gamma}\right)\right) G_u d\langle X\rangle_u\right|\right\|_{p,\{\widehat{Y}^\Gamma=\gamma_k\}}$$
$$+ \left\|\max_{v\in[0,t]} \left|\int_0^v \left(\sigma(u,S_u) - \sigma\left(u, \widetilde{S}_u^{I,\Gamma}\right)\right) d\widetilde{M}_u\right|\right\|_{p,\{\widehat{Y}^\Gamma=\gamma_k\}}.$$

Now, from the Burkholder-Davis-Gundy inequality,

$$\left\|\max_{v\in[0,t]} \left|S_v - \widetilde{S}_v^{I,\Gamma}\right|\right\|_{p,\{\widehat{Y}^\Gamma=\gamma_k\}} \leq [b]_{\mathrm{Lip}} \int_0^t \left\|\max_{v\in[0,u]} \left|S_v - \widetilde{S}_v^{I,\Gamma}\right|\right\|_{p,\{\widehat{Y}^\Gamma=\gamma_k\}} du$$
$$+ [\sigma]_{\mathrm{Lip}}|I| \max_{\substack{i\in I\\ u\in[0,T]}} \left|\left(e_i^X\right)'(u)\right| \left(\max_{i\in I}\left|\widehat{Y}_i^\Gamma\right|\right) \int_0^t \left\|\max_{v\in[0,u]} \left|S_v - \widetilde{S}_v^{I,\Gamma}\right|\right\|_{p,\{\widehat{Y}^\Gamma=\gamma_k\}} du$$
$$+ [\sigma]_{\mathrm{Lip}}|I| \max_{\substack{i\in I\\ u\in[0,T]}} \left|\left(e_i^X\right)'(u)\right| T \left\|\sum_{i\in I} \left|Y_i - \widehat{Y}_i^\Gamma\right|\right\|_{p,\{\widehat{Y}^\Gamma=\gamma_k\}}$$
$$+ \left\|\int_0^t \left|\sigma(u,S_u) - \sigma\left(u, \widetilde{S}_u^{I,\Gamma}\right)\right| |G_u| d\langle X\rangle_u\right\|_{p,\{\widehat{Y}^\Gamma=\gamma_k\}}$$
$$+ C_p^{BDG} \left\|\sqrt{\int_0^t \left(\sigma(u,S_u) - \sigma\left(u, \widetilde{S}_u^{I,\Gamma}\right)\right)^2 d\langle X\rangle_u}\right\|_{p,\{\widehat{Y}^\Gamma=\gamma_k\}}. \quad (3.29)$$

Now, from Schwarz's inequality

$$\left\|\sum_{i\in I}\left|Y_i-\widehat{Y}_i^{\Gamma}\right|\right\|_{p,\{\widehat{Y}^{\Gamma}=\gamma_k\}}\leq\left\|\sqrt{|I|}\sqrt{\sum_{i\in I}\left|Y_i-\widehat{Y}_i^{\Gamma}\right|^2}\right\|_{p,\{\widehat{Y}^{\Gamma}=\gamma_k\}}=\sqrt{|I|}\left\|Y-\widehat{Y}^{\Gamma}\right\|_{p,\{\widehat{Y}^{\Gamma}=\gamma_k\}}.$$

From the generalized Minkowski inequality

$$\begin{aligned}
&\left\|\int_0^t\left|\sigma(u,S_u)-\sigma\left(u,\widetilde{S}_u^{I,\Gamma}\right)\right||G_u|d\langle X\rangle_u\right\|_{p,\{\widehat{Y}^{\Gamma}=\gamma_k\}}\leq\int_0^t\left\|\left(\sigma(u,S_u)-\sigma\left(u,\widetilde{S}_u^{I,\Gamma}\right)\right)G_u\right\|_{p,\{\widehat{Y}^{\Gamma}=\gamma_k\}}d\langle X\rangle_u\\
&=\int_0^t\left\|\left(\sigma(u,S_u)-\sigma\left(u,\widetilde{S}_u^{I,\Gamma}\right)\right)G_u\mathbf{1}_{|G_u|\geq M}+\left(\sigma(u,S_u)-\sigma\left(u,\widetilde{S}_u^{I,\Gamma}\right)\right)G_u\mathbf{1}_{|G_u|\leq M}\right\|_{p,\{\widehat{Y}^{\Gamma}=\gamma_k\}}d\langle X\rangle_u\\
&\leq\int_0^t\left\|\left(\sigma(u,S_u)-\sigma\left(u,\widetilde{S}_u^{I,\Gamma}\right)\right)G_u\mathbf{1}_{|G_u|\geq M}\right\|_{p,\{\widehat{Y}^{\Gamma}=\gamma_k\}}d\langle X\rangle_u\\
&\qquad+\int_0^t\left\|\left(\sigma(u,S_u)-\sigma\left(u,\widetilde{S}_u^{I,\Gamma}\right)\right)G_u\mathbf{1}_{|G_u|\leq M}\right\|_{p,\{\widehat{Y}^{\Gamma}=\gamma_k\}}d\langle X\rangle_u\\
&\leq 2[\sigma]_{\max}\int_0^t\left\|G_u\mathbf{1}_{|G_u|\geq M}\right\|_{p,\{\widehat{Y}^{\Gamma}=\gamma_k\}}d\langle X\rangle_u+M[\sigma]_{\mathrm{Lip}}\int_0^t\left\|S_u-\widetilde{S}_u^{I,\Gamma}\right\|_{p,\{\widehat{Y}^{\Gamma}=\gamma_k\}}d\langle X\rangle_u.
\end{aligned}$$

We obtain, thanks to Proposition 3.3.7

$$\begin{aligned}
&\left\|\int_0^t\left|\sigma(u,S_u)-\sigma\left(u,\widetilde{S}_u^{I,\Gamma}\right)\right||G_u|d\langle X\rangle_u\right\|_{p,\{\widehat{Y}^{\Gamma}=\gamma_k\}}\\
&\leq\underbrace{2[\sigma]_{\max}\langle X\rangle_t(C_p)^{1/p}v_t^{\frac{1}{p}}M^{\frac{1}{q}}\exp\left(-\frac{M^2}{2pv_t^2}\right)}_{:=\eta_M}+M[\sigma]_{\mathrm{Lip}}\int_0^t\left\|S_u-\widetilde{S}_u^{I,\Gamma}\right\|_{p,\{\widehat{Y}^{\Gamma}=\gamma_k\}}d\langle X\rangle_u,
\end{aligned}$$

where $v_t^2=\max_{u\in[0,t]}(\mathrm{Var}(G_u))$. Moreover

$$\left\|\sqrt{\int_0^t\left(\sigma(u,S_u)-\sigma\left(u,\widetilde{S}_u^{I,\Gamma}\right)\right)^2d\langle X\rangle_u}\right\|_{p,\{\widehat{Y}^{\Gamma}=\gamma_k\}}\leq\sqrt{\int_0^t\left\|\max_{\substack{i\in I\\v\in[0,u]}}\left|S_v-\widetilde{S}_v^{I,\Gamma}\right|\right\|^2_{p,\{\widehat{Y}^{\Gamma}=\gamma_k\}}d\langle X\rangle_u}.$$

Hence, Equation (3.29) becomes

$$\begin{aligned}
\left\|\max_{v\in[0,t]}\left|S_v-\widetilde{S}_v^{I,\Gamma}\right|\right\|_{p,\{\widehat{Y}^{\Gamma}=\gamma_k\}}&\leq\underbrace{[\sigma]_{\mathrm{Lip}}|I|\max_{\substack{i\in I\\u\in[0,T]}}\left|\left(e_i^X\right)'(u)\right|\sqrt{|I|}}_{:=A_i^X}\left\|Y-\widehat{Y}^{\Gamma}\right\|_{p,\{\widehat{Y}^{\Gamma}=\gamma_k\}}+\eta_M\\
&\quad+[b]_{\mathrm{Lip}}\int_0^t\left\|\max_{v\in[0,u]}\left|S_v-\widetilde{S}_v^{I,\Gamma}\right|\right\|_{p,\{\widehat{Y}^{\Gamma}=\gamma_k\}}du\\
&\quad+[\sigma]_{\mathrm{Lip}}|I|\max_{\substack{i\in I\\u\in[0,T]}}\left|\left(e_i^X\right)'(u)\right|\left(\max_{i\in I}\left|\widehat{Y}_i^{\Gamma}\right|\right)\int_0^t\left\|\max_{v\in[0,u]}\left|S_v-\widetilde{S}_v^{I,\Gamma}\right|\right\|_{p,\{\widehat{Y}^{\Gamma}=\gamma_k\}}du\\
&\quad+C_p^{BDG}\left(\int_0^t2\left\|\max_{\substack{i\in I\\v\in[0,u]}}\left|S_v-\widetilde{S}_v^{I,\Gamma}\right|\right\|^2_{p,\{\widehat{Y}^{\Gamma}=\gamma_k\}}d\langle X\rangle_u\right)^{1/2}\\
&\quad+\underbrace{M[\sigma]_{\mathrm{Lip}}}_{:=C^{X,M}}\int_0^t\left\|\max_{v\in[0,u]}\left|S_v-\widetilde{S}_v^{I,\Gamma}\right|\right\|_{p,\{\widehat{Y}^{\Gamma}=\gamma_k\}}d\langle X\rangle_u.
\end{aligned}\tag{3.30}$$

We can then apply the "Gronwall-like" lemma 3.3.2 for locally finite measures to the non-decreasing function

$$\left\|\sup_{v\in[0,t]}\left|S_v-\widetilde{S}_v^{I,\Gamma}\right|\right\|_{p,\{\widehat{Y}^{\Gamma}=\gamma_k\}}=\mathbb{E}\left[\sup_{v\in[0,t]}\left|S_v-\widetilde{S}_v^{I,\Gamma}\right|^p\middle|\widehat{Y}^{\Gamma}=\gamma_k\right]^{1/p}$$

and with the locally finite measure μ defined by $\mu(du) = du + d\langle X\rangle_u$, and we obtain

$$\begin{aligned}\left\|\sup_{v\in[0,t]}\left|S_v - \widetilde{S}_v^{I,\Gamma}\right|\right\|_{p,\{\widehat{Y}^\Gamma=\gamma_k\}} &\leq \left(A_I^X\mathbb{E}\left[\left|Y-\widehat{Y}^\Gamma\right|^p\middle|\widehat{Y}^\Gamma=\gamma_k\right]^{1/p}+\eta_M\right)\exp\left(\left(E_I^{X,\gamma_k}+C^{X,M}\right)\mu([0,t))\right)\\ &\leq \left(A_I^X\mathbb{E}\left[\left|Y-\widehat{Y}^\Gamma\right|^p\middle|\widehat{Y}^\Gamma=\gamma_k\right]^{1/p}+\eta_M\right)\underbrace{\exp\left(E_I^{X,\gamma_k}\mu([0,t))\right)}_{:=\phi(\gamma_k)}\exp\left(C^{X,M}\mu([0,t))\right),\end{aligned}$$

where E_I^{X,γ_k} is an affine function of $\max_{i\in I}|(\gamma_k)_i|$. This yields

$$\left\|\sup_{v\in[0,t]}\left|S_v - \widetilde{S}_v^{I,\Gamma}\right|\right\|_p \leq \left(A_I^X\left\|\mathbb{E}\left[\left|Y-\widehat{Y}^\Gamma\right|^p\middle|\widehat{Y}^\Gamma\right]^{1/p}\phi\left(\widehat{Y}^\Gamma\right)\right\|_p+\eta_M\left\|\phi\left(\widehat{Y}^\Gamma\right)\right\|_p\right)\exp\left(C^{X,M}\mu([0,t))\right).$$

Now, for $\varepsilon>0$ and $\tilde{p}=1+\frac{\varepsilon}{p}$ and $\tilde{q}=\frac{\tilde{p}}{\tilde{p}-1}=1+\frac{p}{\varepsilon}$ the conjugate exponent of $\tilde{p}$, we have, thanks to Hölder's inequality

$$\begin{aligned}\mathbb{E}\left[\phi\left(\widehat{Y}^\Gamma\right)^p\mathbb{E}\left[\left|Y-\widehat{Y}^\Gamma\right|^p\middle|\widehat{Y}^\Gamma\right]\right] &\leq \left\|\phi\left(\widehat{Y}^\Gamma\right)^p\right\|_{\tilde{q}}\left\|\mathbb{E}\left[\left|Y-\widehat{Y}^\Gamma\right|^p\middle|\widehat{Y}^\Gamma\right]\right\|_{\tilde{p}}\\ &\leq \left\|\phi\left(\widehat{Y}^\Gamma\right)^p\right\|_{\tilde{q}}\mathbb{E}\left[\left|Y-\widehat{Y}^\Gamma\right|^{p+\varepsilon}\right]^{\frac{p}{p+\varepsilon}}.\end{aligned}$$

Hence,

$$\left\|\mathbb{E}\left[\left|Y-\widehat{Y}^\Gamma\right|^p\middle|\widehat{Y}^\Gamma\right]^{1/p}\phi\left(\widehat{Y}^\Gamma\right)\right\|_p \leq \left\|\phi\left(\widehat{Y}^\Gamma\right)^p\right\|_{\tilde{q}}^{1/p}\mathbb{E}\left[\left|Y-\widehat{Y}^\Gamma\right|^{p+\varepsilon}\right]^{\frac{1}{p+\varepsilon}}.$$

Now, as the so-defined function ϕ is convex and as $\widehat{Y}^\Gamma$ is a stationary quantizer of Y, we have thanks to Equation (3.8), $\left\|\phi\left(\widehat{Y}^\Gamma\right)^p\right\|_{\tilde{q}} \leq \|\phi(Y)^p\|_{\tilde{q}}$ and $\left\|\phi\left(\widehat{Y}^\Gamma\right)\right\|_p \leq \|\phi(Y)\|_p$.

Now, thanks to Proposition 3.3.7, we can ensure that $\eta_M \leq \eta := \left\|Y-\widehat{Y}^\Gamma\right\|_{p+\varepsilon}$ by taking $M = \sqrt{-v_t(p-1)\mathcal{W}\left(-\frac{q\left\|Y-\widehat{Y}^\Gamma\right\|_{p+\varepsilon}^{2q}}{pv_t^2C_p^{2q/p}v_t^{2q/p}}\right)}$ where q is the conjugate exponent of p and $\mathcal{W}$ is the Lambert $\mathcal{W}$ function. We finally have the following error bound

$$\left\|\sup_{v\in[0,t]}\left|S_v - \widetilde{S}_v^{I,\Gamma}\right|\right\|_p \leq C_{X,\varepsilon,I}\exp\left([\sigma]_{\mathrm{Lip}}\sqrt{-v_t(p-1)\mathcal{W}\left(-\frac{q\left\|Y-\widehat{Y}^\Gamma\right\|_{p+\varepsilon}^{2q}}{pv_t^2C_p^{2q/p}v_t^{2q/p}}\right)}\right)\left\|Y-\widehat{Y}^\Gamma\right\|_{p+\varepsilon}.$$

Finally, we can conclude by observing that $\mathcal{W}(u) \underset{u\to 0}{\to} 0$. □

Remark (Without the stationarity property)**.** *The last step of the demonstration of Theorem 3.3.8 (the use of Jensen's inequality) relies on the stationarity of the quantizer $\widehat{Y}$.*
Now, without this stationarity hypothesis and under the additional assumption

$$\Gamma\cap B(0,1)\neq\emptyset, \tag{$\mathcal{A}$}$$

we have for every $i\in I$

$$\left|\widehat{Y}_i\right| \leq \left|Y_i-\widehat{Y}_i\right|+|Y_i| \leq |Y_i|+\left|Y_i-\gamma_i^{k_0}\right| \leq 2|Y_i|+\left|\gamma_i^{k_0}\right| \leq 2|Y_i|+1, \quad \text{where } \gamma^{k_0}\in\Gamma\cap B(0,1).$$

Hence

$$\max_{i\in I}\left|\widehat{Y}_i\right| \leq 2\max_{i\in I}|Y_i|+1.$$

Now, we can notice that the function $\phi(x)$ defined in the demonstration of Theorem 3.3.8 writes $\phi(x) = \psi(\max_{i\in I} x_i)$ for some non-decreasing function ψ. This implies

$$\phi\left(\widehat{Y}\right) = \psi\left(\max_{i\in I}\widehat{Y}_i\right) \leq \psi\left(\max_{i\in I}(2|Y_i|+1)\right) = \phi(2|Y|+1).$$

Hence, we can obtain the same conclusion as in Theorem 3.3.8.

Corollary 3.3.9 (L^p convergence)**.** *With the same notations and hypothesis as in Theorem 3.3.8, consider $\left(\widetilde{X}^{I,\Gamma_n}\right)_{n\in\mathbb{N}}$ a sequence of partial functional quantizers of X and $\left(\widetilde{S}^{I,\Gamma_n}\right)_{n\in\mathbb{N}}$ the corresponding sequence of partial quantizers of S.*
If we make the additional assumption that the associated sequence of quantizers $\left(\widehat{Y}^{\Gamma_n}\right)_{n\in\mathbb{N}}$ is rate-optimal for the $L^{p+\varepsilon}$ convergence for some $\varepsilon > 0$, then for every $t \in [0,T)$ we have

$$\mathbb{E}\left[\sup_{u\in[0,t]}\left|S_u - \widetilde{S}_u^{I,\Gamma_n}\right|^p\right] = O\left(n^{-\frac{p}{|I|}}\right).$$

Proof: As $\left\|Y - \widehat{Y}^{\Gamma_n}\right\|_p \underset{n\to\infty}{\to} 0$, we have *a.s.* $d\left(\widehat{Y}^{\Gamma_n}, Y\right) \underset{n\to\infty}{\to} 0$. Hence, there exists $N_0 \in \mathbb{N}$ such that for every $n \geq N_0$, Γ_n verifies hypothesis $(\mathcal{A})$. From this observation, the result is straightforward consequence of Remark 3.3.2 and Zador's theorem 3.3.3, which defines the optimal convergence rate of a sequence of quantizers. □

The *a.s.* convergence of partially quantized SDE

Theorem 3.3.10 (Almost sure convergence of partially quantized SDE)**.** *Let X be a continuous centered Gaussian martingale on $[0,T]$ with $X_0 = 0$. Let S be the strong solution of the SDE*

$$dS_t = b(t,S_t)dt + \sigma(t,S_t)dX_t, \quad S_0 = x,$$

where $b(t,x)$ and $\sigma(t,x)$ are Borel functions, Lipschitz continuous with respect to x uniformly in t, σ and $|b(\cdot,0)|$ are bounded.
We consider $\left(\widetilde{X}^{I,\Gamma_k}\right)_{k\in\mathbb{N}}$ a sequence of partial functional quantizers of X and $\widetilde{S}^{I,\Gamma_n}$ the corresponding partial functional quantization of S, i.e. the strong solutions of

$$d\widetilde{S}_t^{I,\Gamma_n} = b\left(t,\widetilde{S}_t^{I,\Gamma_n}\right)dt + \sigma\left(t,\widetilde{S}_t^{I,\Gamma_n}\right)d\widetilde{X}_t^{I,\Gamma_n}, \quad \widetilde{S}_0^{I,\Gamma_n} = x.$$

We assume that the sequence of partial quantizers of X is rate-optimal for some $p > |I|$, i.e. that there exists a constant C such that

$$\mathbb{E}\left[\left|Y - \widehat{Y}^{\Gamma_n}\right|^p\right] \leq Cn^{-\frac{p}{|I|}}$$

for every $n \in \mathbb{N}^*$, where Y is defined from X by Equation (3.11) and $\widehat{Y}^{\Gamma}$ is the nearest neighbor projection on Γ. Then for every $t \in [0,T)$, $\widetilde{S}_t^{I,\Gamma_n}$ converges almost surely to S_t.

Proof: From Corollary 3.3.9, if $t \in [0,T)$, there exist three positive constants $K_{X,\varepsilon,I}$, C_t and K_t and $N_0 \in \mathbb{N}$ such that for $n \geq N_0$,

$$\mathbb{E}\left[\sup_{u\in[0,t]}\left|S_u - \widetilde{S}_u^{I,\Gamma_n}\right|^{p-\varepsilon}\right] = O\left(n^{-\frac{p}{|I|}}\right).$$

Hence, as $\frac{p}{|I|} > 1$, Beppo-Levi's theorem for series with non-negative terms implies

$$\mathbb{E}\left[\sum_{n\geq 1}\sup_{u\in[0,t]}\left|S_u - \widetilde{S}_u^{I,\Gamma_n}\right|^{p-\varepsilon}\right] < +\infty.$$

Thus $\sum_{n\geq 1}\sup_{u\in[0,t]}\left|S_u - \widetilde{S}_u^{I,\Gamma_n}\right|^{p-\varepsilon} < +\infty$ $\mathbb{P}-a.s.$ so that $\sup_{u\in[0,t]}\left|S_u - \widetilde{S}_u^{I,\Gamma_n}\right| \underset{n\to\infty}{\to} 0$ $\mathbb{P}-a.s.$. □

Remark (Extension to semimartingales). *In Theorems 3.3.8 and 3.3.10, we limited ourselves to the case where X is a local martingale. The proofs are easily extended to the case of a semimartingale X as soon as there exists a locally finite measure ν on $[0,T]$ such that for every $\omega \in \Omega$ the finite-variation part $dV(\omega)$ in the canonical decomposition of X is absolutely continuous with respect to ν. In particular, this is the case for the Brownian bridge and Ornstein-Uhlenbeck processes whose finite-variation parts are absolutely continuous with respect to the Lebesgue measure on $[0,T]$.*

3.A Injectivity properties of the Wiener integral

In this appendix, we recall some results on the definition of the Wiener integral with respect to a Gaussian process. We focus on the injectivity properties. Here, we pay special attention to the special case of the Ornstein-Uhlenbeck processes.

The covariance operator and the Cameron-Martin space
Consider X a bi-measurable centered Gaussian process on $[0,T]$ such that $\int_0^T \mathbb{E}[X_t^2]dt < \infty$ and with a continuous covariance function Γ^X on $[0,T]\times[0,T]$. We denote by $H := \overline{\operatorname{span}\{X_t, t \in [0,T]\}}^{L^2(\mathbb{P})}$ the Gaussian Hilbert space spanned by $(X_t)_{t\in[0,T]}$. The covariance operator C_X of X is defined by

$$\begin{array}{rccc} C_X: & L^2([0,T]) & \to & L^2([0,T]) \\ & y & \mapsto & C_X y = \mathbb{E}\left[(y,X)X\right]. \end{array}$$

We have $C_X y(t) = \mathbb{E}\left[(y,X)X\right](t) = \mathbb{E}\left[\int_0^T X_s y(s)ds X_t\right] = \int_0^T \Gamma^X(t,s)y(s)ds$ where $\Gamma^X(t,s) = \mathbb{E}[X_t X_s]$ is the covariance function of X.
The Cameron-Martin space of X, (or reproducing Hilbert space of C_X), which we denote by K_X, is the subspace of $L^2([0,T])$ defined by $K_X := \{t \mapsto \mathbb{E}\left[ZX_t\right], Z \in H\}$. K_X is equipped with the scalar product defined by

$$\langle k_1, k_2\rangle_X = \mathbb{E}\left[Z_1 Z_2\right] \quad \text{if} \quad k_i = \mathbb{E}\left[Z_i X.\right],\ i = 1,2,$$

so that $(K_X, \langle\cdot\rangle_X)$ is a Hilbert space, isometric with the Hilbert space $\overline{\{(y,X) : y \in L^2([0,T])\}}^H$. K_X is spanned as a Hilbert space by $\left\{C_X(y) : y \in L^2([0,T])\right\}$.

The Wiener integral
Here, we follow the same steps as Lebovits and Lévy-Véhel in [16] and Jost in [15] for the definition of a general Wiener integral. The difference here is that we use the quotient topology in order to define the Wiener integral in a more general setting.
We define the map $U : H \to K_X$ defined by $U(Z)(t) = \mathbb{E}[ZX_t]$. By definition of H and K_X, U is a bijection and for any $s \in [0,T]$, we have $U(X_s) = \Gamma^X(s,\cdot)$. Consequently, K_X is spanned by $\left(\Gamma^X(s,\cdot)\right)_{s\in[0,T]}$ as a Hilbert space. Now, we linearly map the set of the piecewise constant functions $\mathcal{E}([0,T])$ to the Cameron-Martin space K_X by

$$\begin{array}{rccc} J: & \mathcal{E}([0,T]) & \to & K_X \\ & \mathbf{1}_{|s,t|} & \mapsto & \Gamma^X(t,\cdot) - \Gamma^X(s,\cdot), \end{array}$$

where $|a,b|$ stands either for the interval $[a,b]$, (a,b), $(a,b]$ or $[a,b)$. We equip $\mathcal{E}([0,T])$ with the bilinear form $\langle\cdot,\cdot\rangle_J$ which is defined by

$$\langle f, g\rangle_J := \langle Jf, Jg\rangle_X.$$

It is a bilinear symmetric positive-*semidefinite* form.

Remark. *The so-called reproducing property shows that* $\left\langle \mathbf{1}_{|0,t|}, \mathbf{1}_{|0,s|}\right\rangle_J = \Gamma^X(t,s) + \Gamma^X(0,0) - \Gamma^X(0,s) - \Gamma^X(0,t)$. *When $X_0 = 0$ a.s., this gives* $\left\langle \mathbf{1}_{|0,t|}, \mathbf{1}_{|0,s|}\right\rangle_J = \Gamma^X(s,t)$.

Now, we define the equivalence relation $\underset{J}{\sim}$ on $\mathcal{E}([0,T])$ by $x \underset{J}{\sim} y$ if $\langle x-y, x-y\rangle_J = 0$. On the quotient space $E([0,T]) := \mathcal{E}([0,T])/\underset{J}{\sim}$, the bilinear form $\langle\cdot,\cdot\rangle_J$ is positive-definite and thus it is a scalar product on $E([0,T])$. In this context, J defines an (isometric) linear map from $E([0,T])$ to K_X. Then, considering the completion F of $E([0,T])$ associated with this scalar product, J is extended to F and $U^{-1}\circ J : F \to H$ is an (isometric) injective map that we call Wiener integral associated to X.

$$\int_0^T f(t)dX_t := U^{-1}\circ J(f).$$

Injectivity properties of the Wiener integral

As we have just seen, the Wiener integral is an (isometric) injective map from F to H. Still, for example, when dealing with a standard Brownian bridge on $[0,T]$, $\left\|\mathbf{1}_{[0,T]}\right\|_J = 0$, so that there are functions of $\mathcal{E}([0,T])$ which have a nonzero L^2 norm and a zero $\|\cdot\|_J$ norm. Injectivity only holds in the quotient space $E([0,T]) = \mathcal{E}([0,T])/\underset{J}{\sim}$ and its completion F.

It is classical background that in the special case of a standard Brownian motion, $\|\cdot\|_J$ exactly coincides with the canonical L^2 norm so that $F = L^2([0,T])$.

Study of the case of Ornstein-Uhlenbeck processes

From now, we will assume that X is a centered Ornstein-Uhlenbeck process defined on $[0,T]$ by the SDE

$$dX_t = -\theta X_t dt + \sigma dW_t \quad \text{with } \sigma > 0 \text{ and } \theta > 0,$$

where W is a standard Brownian motion and $X_0 \overset{\mathcal{L}}{\sim} \mathcal{N}\left(0,\sigma_0^2\right)$ is independent of W. We make the additional assumption that $\theta T \leq \frac{4}{3}$. The covariance function writes

$$\Gamma^X(s,t) = \frac{\sigma^2}{2\theta}e^{-\theta(s+t)}\left(e^{2\min(s,t)}-1\right) + \sigma_0^2 e^{-\theta(s+t)}.$$

Proposition 3.A.1 (Semi-norm equivalence on $\mathcal{E}([0,T])$)**.** *There exist two positive constants c and C such that for every $f \in \mathcal{E}([0,T])$, $c\|f\|_2 \leq \|f\|_J \leq C\|f\|_2$.*

Proof: Let us consider $f \in \mathcal{E}([0,T])$. We have

$$\begin{aligned}\|f\|_J^2 &= \operatorname{Var}\left(-\theta\int_0^T f(s)X_s ds + \sigma\int_0^T f(s)dW_s\right)\\ &\leq 2\operatorname{Var}\left(\theta\int_0^T f(s)X_s ds\right) + 2\operatorname{Var}\left(\sigma\int_0^T f(s)dW_s\right).\end{aligned}$$

The solution of the Ornstein-Uhlenbeck SDE is

$$X_t = \underbrace{X_0 e^{-\theta t}}_{\text{independent of } W} + \underbrace{\int_0^t \sigma e^{\theta(s-t)}dW_s}_{:=X_t^0}. \tag{3.31}$$

The so-defined process $\left(X_t^0\right)_{t\in[0,T]}$ is a centered Ornstein-Uhlenbeck process starting from 0. Hence we have

$$\begin{aligned}\|f\|_J^2 &\leq 2\operatorname{Var}\left(X_0\theta\int_0^T f(s)e^{-\theta s}ds\right) + 2\operatorname{Var}\left(\theta\int_0^T f(s)X_s^0 ds\right) + 2\operatorname{Var}\left(\sigma\int_0^T f(s)dW_s\right)\\ &\leq 2\theta^2 T\operatorname{Var}(X_0)\int_0^T f(s)^2 ds + 2\operatorname{Var}\left(\theta\int_0^T f(s)X_s^0 ds\right) + 2\operatorname{Var}\left(\sigma\int_0^T f(s)dW_s\right).\end{aligned}$$

We have seen in the proof of Proposition 3.2.7 that $\operatorname{Var}\left(\theta\int_0^T f(s)X_s^0 ds\right) \leq \operatorname{Var}\left(\sigma\int_0^T f(s)dW_s\right)$. Hence

$$\|f\|_J^2 \leq \underbrace{\left(2\theta^2 T\sigma_0^2 + 4\sigma^2\right)}_{:=C^2}\int_0^T f(s)^2 ds.$$

This is the desired inequality.

Now we write

$$\int_0^t f(s)dX_s = \underbrace{-\theta \int_0^T f(s)X_0 e^{-\theta s} ds}_{:=G_0^f} + \underbrace{\left(-\theta \int_0^T f(s)X_s^0 ds\right)}_{:=G_1^f} + \underbrace{\sigma \int_0^T f(s)dW_s}_{:=G_2^f},$$

where $\left(G_0^f, G_1^f, G_2^f\right)$ is Gaussian and G_0^f is independent of G_1^f and G_2^f. Hence

$$\begin{aligned}\operatorname{Var}\left(\int_0^t f(s)dX_s\right) &\geq \operatorname{Var}\left(G_1^f + G_2^f\right) = \operatorname{Var}\left(G_1^f\right) + \operatorname{Var}\left(G_2^f\right) + 2\operatorname{cov}\left(G_1^f, G_2^f\right) \\ &\geq \operatorname{Var}\left(G_1^f\right) + \operatorname{Var}\left(G_2^f\right) - 2\sqrt{\operatorname{Var}\left(G_1^f\right)\operatorname{Var}\left(G_2^f\right)} = \left(\sqrt{\operatorname{Var}\left(G_2^f\right)} - \sqrt{\operatorname{Var}\left(G_1^f\right)}\right)^2. \quad (3.32)\end{aligned}$$

It has been shown at the beginning of the proof of Proposition 3.2.7 that there exists a constant $K < 1$ independent of f such that $\operatorname{Var}\left(G_1^f\right) \leq K \operatorname{Var}\left(G_2^f\right)$. K was defined by

$$K = \frac{\theta^2}{\sigma^2}\sqrt{\int_0^T \int_0^T \left(\Gamma^{X^0}(s,t)\right)^2 dsdt},$$

where Γ^{X^0} is the covariance function of the Ornstein-Uhlenbeck process starting from 0. Plugging this into Equation (3.32) yields

$$\operatorname{Var}\left(\int_0^t f(s)dX_s\right) \geq \left(1-\sqrt{K}\right)^2 \operatorname{Var}\left(G_2^f\right) = \underbrace{\left(1-\sqrt{K}\right)^2 \sigma^2}_{:=c^2} \|f\|_2^2.$$

This is the wanted inequality. □

A straightforward consequence of Proposition 3.A.1 is that $\|f\|_J = 0 \Leftrightarrow \|f\|_2 = 0$ so that equivalence classes in $\mathcal{E}([0,T])$ for the relation $\underset{J}{\sim}$ are *almost surely* equal functions. Another consequence is that the sets of Cauchy sequences and convergent sequences for the two norms on $E([0,T])$ coincide, and thus the corresponding completions of $E([0,T])$ are the same. In other words, in the case of Ornstein-Uhlenbeck processes that satisfy the condition $\theta T \leq \frac{4}{3}$, we have $F = L^2([0,T])$.

Bibliography

[1] Larbi Alili. Canonical decompositions of certain generalized Brownian bridges. *Electronic communications in probability*, 7:27–35, 2002.

[2] Stefan Ankirchner, Steffen Dereich, and Peter Imkeller. Enlargement of filtrations and continuous Girsanov-type embeddings. In *Séminaire de Probabilité XV*, 2007.

[3] Kendall E. Atkinson. *The numerical solution of integral equations of the second kind.* Cambridge Monographs on Applied and Computational Mathematics, 1999.

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

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

[6] Sylvain Corlay. The Nyström method for functional quantization with an application to the fractional Brownian motion. *Preprint*, 2010.

[7] Sylvain Corlay and Gilles Pagès. Functional quantization-based stratified sampling methods. *Preprint*, 2010.

[8] Paul Deheuvels and Guennadi V. Martynov. A Karhunen-Loève decomposition of a Gaussian process generated by independent pairs of exponential random variables. *Journal of Functional Analysis*, 255(9):2363–2394, 2008.

[9] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes, characterization and convergence.* Wiley Series in Probability and Statistics, 2005.

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

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

[12] Jean Jacod. Grossissements de filtrations : exemples et applications. *Lecture Notes in Mathematics*, 1118:15–35, 1985.

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

[14] Thierry Jeulin. Semi-martingales et grossissement d'une filtration. *Lecture Notes in Mathematics*, 833, 1980.

[15] Céline Jost. Measure-preserving transformations of Volterra Gaussian processes and related bridges. *Preprint*, 2007.

[16] Joachim Lebovits and Jacques Lévy-Véhel. White noise-based stochastic calculus with respect to multi-fractional Brownian motion. *Preprint*, 2011.

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

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

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

[20] Harald Luschgy, Gilles Pagès, and Benedikt Wilbertz. Asymptotically optimal quantization schemes for Gaussian processes. *ESAIM: PS*, 14:93–116, 2010.

[21] Roger Mansuy and Marc Yor. *Random times and enlargements of filtrations in a Brownian setting.* Springer-Verlag New York, Inc., 2006.

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

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

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

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

[26] Gilles Pagès and Afef Sellami. Convergence of multi-dimensional quantized SDE's. In Catherine Donati-Martin, Antoine Lejay, and Alain Rouault, editors, *Séminaire de Probabilités XLIII*, pages 269–308. Springer, Berlin, 2010.

[27] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion.* Springer, 3rd edition, 2005.

[28] Benedikt Wilbertz. *Construction of optimal quantizers for Gaussian measures on Banach spaces.* PhD thesis, Universität Trier, 2008.

[29] Marc Yor. Grossissement d'une filtration et semi-martingales : Théoremes généraux. *Lecture Notes in Mathematics*, 649, 1978.

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

Chapter 4

Pricing vanilla options in stochastic volatility models with cubature rules based on functional quantization

Abstract

In this chapter, we propose a cubature scheme based on the functional quantization of stochastic processes for the pricing of vanilla options in stochastic volatility models.

We first put ourselves in the same framework as in [27] to clarify the main steps of the procedure. Meanwhile, we introduce a kind of variance reduction method for computing implied volatilities with functional quantization-based cubature.

Then the method is extended to the case of stochastic volatility models with embedded local volatility, often called "local stochastic volatility models". For this purpose, we propose a new quantization scheme for stochastic differential equations which we call "normal quantization". This approximation is based on a recent approach to functional quantization called "partial functional quantization" which has been introduced in [5].

We perform numerical experiments in the case of the SABR model.

Joint work with Gilles Pagès.

Keywords: Gaussian semimartingale, functional quantization, partial quantization, Karhunen-Loève, Brownian motion, Brownian bridge, Ornstein-Uhlenbeck, cubature method, option pricing, vanilla option, stochastic volatility, local stochastic volatility.

Introduction

The first historical stock price model which was able to fit market vanilla option prices is Dupire's local volatility model. Among all the models that fit market vanilla option prices, it is the only diffusion whose dynamics is driven by a unique Brownian motion, which gives it a specific relevance. The practical advantage of using this model is that the so-called Dupire's formula gives an easy way of calibration as soon as one has a continuous set of option prices at ones disposal (for every strike and maturity).

In terms of probability distributions, the fact that Dupire's model fits the market vanilla option prices simply means that it fits the marginal distributions of the spot implied by the market at future dates. However, the knowledge of these marginal distributions does not entirely characterize the transition probabilities of the spot price between these future dates.

Modeling these transition probabilities is essential to price those derivatives that don't only depend on marginal distributions of the spot price, such as forward-start options. Thus, using a local volatility model for the pricing of such options comes to make an arbitrary choice and turn a blind eye to the modeling of these transition probabilities. Such a choice is usually motivated by the tractability of Dupire's model and the difficulties raised by the problem of transition probabilities modeling.

The approach of practitioners for the modeling of these transition probabilities is the use of stochastic volatility models. A requirement for a model to be tractable is that we must be able to efficiently price the financial instruments with which it has to fit, so that we can run optimization algorithms and calibrate the parameters of the model. Usually, for accounting purposes, these instruments have the vanilla options among them. Stochastic volatility models are an active field of research, and many approaches have been developed for constraints mentioned already. The necessity to rapidly compute vanilla options prices has become a discriminatory criterion in the choice of a model, often even more than any other of its advantages and drawbacks.

A common approach in financial institutions is the use of so-called *local stochastic volatility* models. The basic idea is to first roughly calibrate a crude stochastic volatility model, with a reasonable number of parameters, whose behavior reproduces the principal features of the smile dynamics that one wants to see. The *smile dynamic* refers to these transition probability distributions expressed as "Black & Scholes implied volatility" for the corresponding forward start vanilla options. This small number of parameters is not sufficient to fit every listed vanilla option price. Then an additional local volatility term is plugged in in order to recover the perfect consistency with the vanilla option prices. Ideally, this corrective function should be close to 1.

In this chapter, we devise numerical methods for the pricing of vanilla options for a wide class of stochastic volatility models with an embedded local volatility term. First, we follow the steps of the method proposed in [27] to deal with crude stochastic volatility models. The principle is to perform a preconditioning with respect to the volatility process and a cubature method based on functional quantization. One obtains a vanilla option price in the stochastic volatility model as a weighted sum of closed-form option prices (with non-stochastic volatility).

Then we attempt to apply this method to the case of a local stochastic volatility model. In this case, the transformation of the stochastic differential equation required by our method shows a local drift and a local volatility in the obtained diffusion equation. In order to deal with it, we propose a new approximation scheme for stochastic differential equations called the "normal quantization". This technique is inspired by the new approach to functional quantization, the "partial functional quantization" recently developed in [5]. In doing so, we obtain a cubature formula for the price of a vanilla option in a local stochastic volatility model as a weighted sum of finitely many options prices in lognormal spot models. We test this new cubature formula in the case of the SABR model.

The chapter is organized as follows: the first section provides a short introduction to functional quadratic optimal quantization. Section 4.2 is devoted to the cubature schemes based on optimal quantization and the possibility of using Richardson-Romberg extrapolation methods. Section 4.3 provides a brief description of the method when applied to the simple case where no local volatility term is involved. Section 4.4 deals with the more general cases of local stochastic volatility models

and the difficulty raised by the addition of a local volatility function. We present the normal quantization scheme for stochastic differential equations which was addressed above. Then we perform numerical tests in the case of the SABR model.

4.1 Some backgrounds on functional quantization

4.1.1 Functional quantization of Gaussian processes

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, and $(E, |\cdot|)$ a separable Banach space. For $N \in \mathbb{N}^*$, the N-quantization of a E-valued random variable X consists in its approximation by a random variable Y taking at most N values. The resulting error of this discretization is measured by the L^p norm of $|X - Y|$. The minimization of the error yields the following minimization problem:

$$\min \left\{ \left\| |X - Y| \right\|_p, Y : \Omega \to E \text{ measurable, } \operatorname{card}(Y(\Omega)) \leq N \right\}. \tag{4.1}$$

A solution of (4.1) is an optimal quantizer of X. This problem, initially considered in the finite-dimensional case was first investigated for signal transmission issues [12]. Then it has been introduced in numerical probability to devise cubature methods [25]. Since the 2000's, the infinite-dimensional setting has been extensively investigated from both constructive numerical and theoretical viewpoints with a special attention paid to functional quantization, especially in the Hilbert case [21] but also in some other Banach spaces [34]. Stochastic processes are viewed as random variables taking values in their path spaces such as $L^2_T := L^2([0,T], dt)$. As applications are concerned, functional quantization has been used to design cubature schemes [27] and variance reduction methods [6, 20].

Here, we assume that X is a bi-measurable stochastic process on $[0,T]$ verifying $\int_0^T \mathbb{E}\left[|X_t|^2\right] dt < +\infty$, so this can be viewed as a random variable valued in the separable Hilbert space $L^2([0,T])$. We make the assumption that the covariance function of X, denoted by Γ^X is continuous. In [21], it is shown that in the centered Gaussian case, linear subspaces U of $L^2([0,T])$ spanned by L^2-optimal quantizers correspond to principal components of X, *i.e.* are spanned by the first eigenvectors of the covariance operator of X. Hence, the quadratic optimal quantization of Gaussian processes involves its Karhunen-Loève decomposition $(e_n^X, \lambda_n^X)_{n\geq 1}$.

If Y is a quadratic N-optimal quantizer of the Gaussian process X and $d^X(N)$ is the dimension of the subspace of $L^2([0,T])$ spanned by $Y(\Omega)$, the quadratic quantization error $\mathcal{E}_N(X)$ verifies

$$\mathcal{E}_N^2(X) = \sum_{j\geq m+1} \lambda_j^X + \mathcal{E}_N^2\left(\bigotimes_{j=1}^m \mathcal{N}\left(0, \lambda_j^X\right)\right) \text{ for } m \geq d^X(N). \tag{4.2}$$

$$\mathcal{E}_N^2(X) < \sum_{j\geq m+1} \lambda_j^X + \mathcal{E}_N^2\left(\bigotimes_{j=1}^m \mathcal{N}\left(0, \lambda_j^X\right)\right) \text{ for } 1 \leq m < d^X(N). \tag{4.3}$$

To perform optimal quantization, the decomposition is first truncated at a fixed order m and then the $\mathbb{R}^m$-valued Gaussian vector, constituted of the m first coordinates of the process on its Karhunen-Loève decomposition, is quantized. To reach optimal quantization, we have to determine the optimal rank of truncation $d^X(N)$ (the quantization dimension) and to determine the optimal $d^X(N)$-dimensional quantizer corresponding to the first coordinates $\bigotimes_{j=1}^{d^X(N)} \mathcal{N}\left(0, \lambda_j^X\right)$. Usual examples of such processes are the standard Brownian motion on $[0,T]$, the standard Brownian bridge on $[0,T]$, Ornstein-Uhlenbeck processes and the fractional Brownian motion.

Another possibility is to use a product quantization of the distribution $\bigotimes_{j=1}^{d^X(N)} \mathcal{N}(0, \lambda_j^X)$. The product quantization is the Cartesian product of the optimal quadratic quantizers of the standard one-dimensional Gaussian distributions $\mathcal{N}(0, \lambda_i^X)_{1\leq i\leq d^X(N)}$. In the case of independent marginals,

this yields a stationary quantizer, *i.e.* a quantizer Y of X which satisfies $\mathbb{E}[X|Y] = Y$. This property, shared with optimal quantizers, results in a convergence rate of a higher order by one for the quantization-based cubature scheme, as we will see in Section 4.2.1. One advantage of this setting is that the one-dimensional Gaussian quantization is a fast procedure. In [26], deterministic optimization methods (as Newton-Raphson) are shown to converge rapidly to the unique optimal quantizer of the one-dimensional Gaussian distribution. Moreover, a sharply optimized database of quantizers of standard univariate and multivariate Gaussian distributions is available on the web site `www.quantize.maths-fi.com` [28] for download. Still, we have to determine the quantization size in each direction to obtain optimal product quantization. In this case, the minimization of the distortion (4.2) comes to the following minimization problem

$$\min \left\{ \sum_{j=1}^{d} \mathcal{E}_{N_j}^2 \left(\mathcal{N}(0, \lambda_j^X) \right) + \sum_{j \geq d+1} \lambda_j^X, \ N_1 \times \cdots \times N_d \leq N, \ d \geq 1 \right\}. \tag{4.4}$$

A solution of (4.4) is called an optimal product quantizer. This problem can be solved by the "blind optimization procedure", which consists in computing the criterion for every possible decomposition $N_1 \times \cdots \times N_d$ with $N_1 \geq \cdots \geq N_d$. The result of this procedure can be kept off-line for a future use. Optimal decompositions for a wide range of values of N for both Brownian motion and Brownian bridge are available on the web site `www.quantize.maths-fi.com` [28]. Another fact on optimal functional product quantization is that it is shown to be rate-optimal.

In Figure 4.1, we display a quadratic optimal N-quantizer of the fractional Brownian motion on $[0, 1]$ with Hurst exponent $H = 0.25$ and $N = 20$ and a quadratic $5 \times 2 \times 2$-product quantizer of the stationary Ornstein-Uhlenbeck process defined by the SDE $dr_t = -r_t dt + dW_t$, on $[0, 3]$.

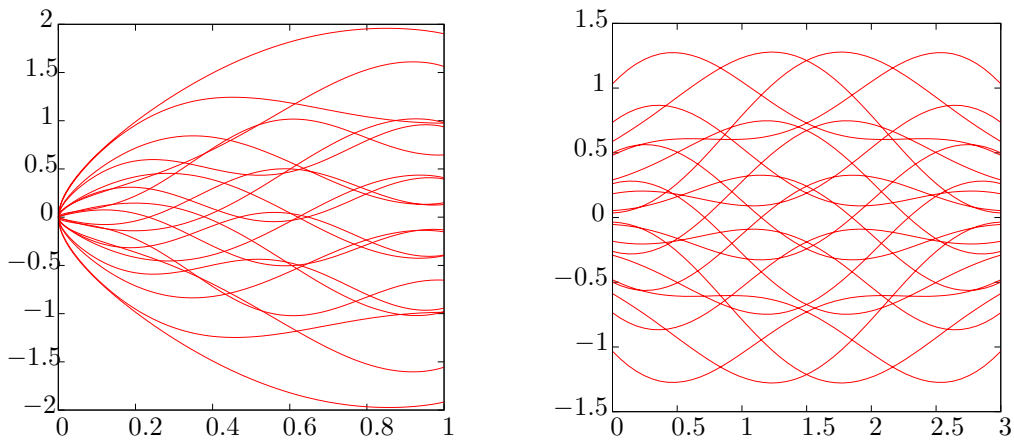

Figure 4.1: Optimal product quantizer of fractional Brownian motion on $[0, 1]$ with Hurst parameter $H = 0.25$ on the left and a quadratic $5 \times 2 \times 2$-product quantizer of the stationary Ornstein-Uhlenbeck process defined by the SDE $dr_t = -r_t dt + dW_t$, on $[0, 3]$ on the right.

In [21], the rate of convergence to zero of the quantization error is investigated. A complete solution is provided for the case of Gaussian processes under rather general conditions on the eigenvalues of the covariance operator.

From a constructive viewpoint, the numerical computation of the optimal quantization or the optimal product quantization requires a numerical evaluation of the Karhunen-Loève eigenfunctions and eigenvalues, at least the very first terms. (The quantization dimension of usual Gaussian processes increases asymptotically as the logarithm of the size of the quantizer[1], so it is most likely that it is small. For instance, the quantization dimension $d^W(N)$ of the Brownian motion with

[1] This holds in the case of the optimal product quantization. As optimal quantization concerns, it is a conjecture, supported by numerical evidence. We refer to [22, 24] for more details on asymptotics of the quantization dimension $d^X(N)$.

$N = 10000$ is 9.) The Karhunen-Loève decompositions of several usual Gaussian processes have a closed-form expression. It is the case for the standard Brownian motion, the Brownian bridge and Ornstein-Uhlenbeck processes.

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

$$e_n^W(t) := \sqrt{\frac{2}{T}} \sin\left(\pi(n-1/2)\frac{t}{T}\right), \qquad \lambda_n^W := \left(\frac{T}{\pi(n-1/2)}\right)^2, \quad n \geq 1. \tag{4.5}$$

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

$$e_n^B(t) := \sqrt{\frac{2}{T}} \sin\left(\pi n\frac{t}{T}\right), \qquad \lambda_n^B := \left(\frac{T}{\pi n}\right)^2, \quad n \geq 1. \tag{4.6}$$

3. The Ornstein-Uhlenbeck process on $[0,T]$, starting from 0, defined by the SDE $dr_t = \theta(mu - r_t)dt + \sigma dW_t$, with $\sigma \geq 0$, $\theta > 0$ and W a standard Brownian motion on $[0,T]$. (See [6]).

$$e_n^{OU}(t) := \left(\frac{1}{\sqrt{\frac{T}{2} - \frac{\sin(2\omega_{\lambda_n}T)}{4\omega_{\lambda_n}}}}\right) \sin(\omega_{\lambda_n} t), \qquad \lambda_n^{OU} := \frac{\sigma^2}{\omega_{\lambda_n}^2 + \theta^2}, \quad n \geq 1, \tag{4.7}$$

where ω_{λ_n} are the (sorted) strictly positive solutions of the equation

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

4. The stationary Ornstein-Uhlenbeck process on $[0,T]$, defined by the same SDE with $r_0 \overset{\mathcal{L}}{\sim} \mathcal{N}\left(0, \frac{\sigma^2}{2\theta}\right)$. (See [6]).

$$e_n^{OU}(t) := C_n \left(\omega_{\lambda_n} \cos(\omega_{\lambda_n} t) + \theta \sin(\omega_{\lambda_n} t)\right), \qquad \lambda_n^{OU} := \frac{\sigma^2}{\omega_{\lambda_n}^2 + \theta^2}, \quad n \geq 1, \tag{4.8}$$

where ω_{λ_n} are the (sorted) strictly positive solutions of the equation

$$2\theta\omega_n \cos(\omega_{\lambda_n} T) + \left(\theta^2 - \omega_{\lambda_n}^2\right) \sin(\omega_{\lambda_n} T) = 0,$$

and

$$\frac{1}{C_n^2} = \frac{\theta}{2}\left(1 - \cos(2\omega_{\lambda_n} T)\right) + \frac{\omega_{\lambda_n}}{2}\left(T + \frac{\sin(2\omega_{\lambda_n} T)}{2\omega_{\lambda_n}}\right) + \frac{\theta^2}{2}\left(T - \frac{\sin(2\omega_{\lambda_n} T)}{2\omega_{\lambda_n}}\right).$$

The case Ornstein-Uhlenbeck processes is derived in [6], in the general setting of an arbitrary initial variance σ_0. A procedure for the computation of ω_λ is also provided. Other examples of Karhunen-Loève expansions with closed-form expressions are derived in [7] by Deheuvels and Martynov.

In general, no closed-form expression for the Karhunen-Loève expansion is available. For instance, as far as we know, so is the case of the Karhunen-Loève expansion of the fractional Brownian motion. To fulfill the requirement of a numerical evaluation of those functions, it is possible to use numerical methods related to integral equations to solve the functional eigenvalue problem that defines the Karhunen-Loève decomposition. A review of these methods is available in [2]. In [4], the so-called Nyström method is used to compute the first terms of the Karhunen-Loève decomposition of the fractional Brownian motion for its optimal functional quantization.

4.1.2 Quantization of solutions of stochastic differential equations

An application of the quantization of a Gaussian process X is to perform a quantization of the solution of a stochastic differential equation with respect X, (as soon as we can define the corresponding stochastic integral). In the present case, we assume that X is a continuous Gaussian semimartingale on $[0,T]$. As a consequence, we have $\int_0^T \mathbb{E}[X_t^2]dt < \infty$ owing to Fernique's theorem, and X has a continuous covariance function (see [18, VIII.3]).

We can obtain a stationary quantizer of the diffusion by plugging the quantizer of the Gaussian process into the diffusion equation written in the Stratonovich sense. In [29], the *a.s.* convergence

of this quantization to σ when the quantizer size goes to infinity is proved. Moreover, the rate of convergence of the quantization error for the SDE solution is the same as for the considered Gaussian process. The article [29] is mostly specific to the Brownian motion but main results remain valid for continuous semimartingales which satisfy the Kolmogorov criterion like the Brownian bridge and Ornstein-Uhlenbeck processes.

For instance, let us consider σ the stochastic process defined as the strong solution of the following SDE, with respect to X on $[0,T]$:

$$d\sigma_t = b(t,\sigma_t)dt + \theta(t,\sigma_t)dX_t, \quad \sigma_0 \in \mathbb{R}, \tag{4.9}$$

where $b(t,x)$ and $\theta(t,x)$ are Borel functions, Lipschitz continuous with respect to x uniformly in t and $|b(\cdot,0)| + |\theta(\cdot,0)|$ is bounded on $[0,T]$. Then a unique strong solution of (4.9) exists on $[0,T]$ (see *e.g.* [19, Theorem A.3.3]).

We recall that if M and H are continuous semimartingales, the Stratonovich integral $H \circ M$ is defined by $H \circ M = H \cdot M + \frac{1}{2}\langle H, M\rangle$ where $H \cdot M$ stands for the Itô integral of H with respect to M. If $\theta(t,x)$ is differentiable with respect to x, we can write the SDE (4.9) in the Stratonovich form, $d\sigma_t = b(t,\sigma_t)dt + \theta(t,\sigma_t) \circ dX_t - \frac{1}{2}d\langle\theta(\cdot,\sigma.), X\rangle_t$. Moreover, $d\langle\theta(\cdot,\sigma.), X\rangle_t = \theta'_x(t,\sigma_t)\theta(t,\sigma_t)d\langle X\rangle_t$. Thus, we obtain

$$d\sigma_t = b(t,\sigma_t)dt - \frac{1}{2}\theta'_x(t,\sigma_t)\theta(t,\sigma_t)d\langle X\rangle_t + \theta(t,\sigma_t) \circ dX_t. \tag{4.10}$$

We recall in mind that a continuous centered semimartingale X is Gaussian if and only if $\langle X\rangle$ is deterministic (see *e.g.* [30]). Now, we replace X by the components of stationary quantizer of X. In doing so, we obtain a functional quantization of σ. As we assumed that the covariance function Γ^X is continuous, Mercer's theorem ensures that the Karhunen-Loève eigenfunctions associated with non-zero eigenvalues are continuous functions. In the cases cited above, the eigenfunctions associated with non-zero eigenvalues are C^∞ functions. In the following, we will assume that $\langle X\rangle$ and the Karhunen-Loève eigenfunctions of X are C^1 functions. The quadratic variation has a closed-form expression in the following cases:

- The standard Brownian motion: $d\langle X\rangle_t = dt$.
- The Ornstein-Uhlenbeck process defined by the SDE $dX_t = \theta(\mu - X_t)\,dt + \sigma^{0U}dW_t$ where $\sigma \geq 0$, $\theta > 0$ and W is a standard Brownian motion on $[0,T]$: $d\langle X\rangle_t = \left(\sigma^{OU}\right)^2 dt$.
- The standard Brownian bridge on $[0,T]$: we can write $dX_t = -\frac{W_T}{T}dt + dW_t$. This yields $d\langle X\rangle_t = dt$.

We refer to [3] for examples of stochastic volatility models which involve Ornstein-Uhlenbeck process. Another example is the λ-SABR model proposed by Henry-Labordère in [16] who derived short-maturity asymptotics for this model. Moreover, exponentials of Ornstein-Uhlenbeck processes are used to model the dynamics of commodity future prices [31].

Plugging the quantizer into the Stratonovich SDE

We consider $\chi = (\chi^i)_{1\leq i\leq N}$ a functional quantizer of X. χ can either be an optimal quantization or a stationary K-L product quantization. (In the second case, χ is usually indexed by a multi-index.) Here, we replace the process X by its quantizer into the Stratonovich SDE (4.10). This yields an ordinary differential equation for every $i \in \{1,\cdots,N\}$. The quantization $(\widehat{\sigma}^i)_{1\leq i\leq N}$ of σ is composed by the following ordinary differential equations

$$d\widehat{\sigma}^i_t = b(t,\widehat{\sigma}^i_t)dt - \frac{1}{2}\theta'_x(t,\widehat{\sigma}^i_t)\theta(t,\widehat{\sigma}^i_t)d\langle X\rangle_t + \theta(t,\widehat{\sigma}^i_t)\left(\chi^i\right)'(t)dt, \quad \widehat{\sigma}^i_0 = \sigma_0 > 0. \tag{4.11}$$

In some cases, these equations may have explicit solutions, but in the general case, we have to use numerical methods, like higher-order Runge-Kutta schemes or Bulirsch-Stoer methods. See the book [15] for a review of the numerical methods for solving this kind of ODE.

The lognormal case

Let us consider the case where $b(t,x) = x\mu(t)$, and $\theta(t,x) = x\gamma(t)$. Equation (4.11) becomes

$$d\widehat{\sigma}_t^i = \widehat{\sigma}_t^i \mu(t)dt - \widehat{\sigma}_t^i \frac{\gamma(t)^2}{2} d\langle X\rangle_t + \widehat{\sigma}_t^i \gamma(t) \left(\chi^i\right)'(t)dt, \quad \widehat{\sigma}_0^i = \sigma_0 > 0,$$

which leads to

$$\widehat{\sigma}_t^i = \sigma_0 \exp\left(\int_0^t \mu(s)ds + \int_0^t \gamma(s)\left(\chi^i\right)'(s)ds - \frac{1}{2}\int_0^t \gamma^2(s)d\langle X\rangle_s\right). \tag{4.12}$$

When γ is constant, (4.12) becomes

$$\widehat{\sigma}_t^i = \sigma_0 \exp\left(\int_0^t \mu(s)ds + \left(\chi^i(t) - \chi^i(0)\right)\gamma - \frac{\gamma^2}{2}\langle X\rangle_s\right). \tag{4.13}$$

In Figure 4.2, a functional N-quantizer with $N = 20$ of a log-Ornstein-Uhlenbeck process is plotted (*i.e.* the process defined by Equation (4.13) when X is an Ornstein-Uhlenbeck process). This quantizer of σ is obtained from a $5 \times 2 \times 2$ K-L product quantizer of a centered stationary Ornstein-Uhlenbeck with a mean reversion and a volatility both equal to 1. This product quantizer is then plugged into the SDE (4.13), with $\sigma_0 = 100$, $\gamma = 1$ and $\mu(s) \equiv 0$.

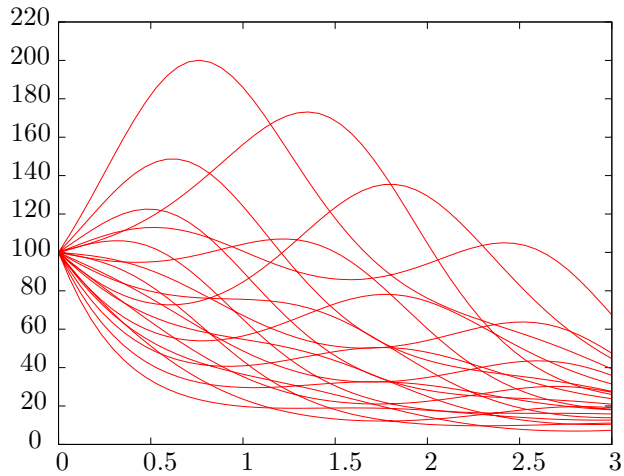

Figure 4.2: Functional quadratic $5\times2\times2$-product quantizer of the log-Ornstein-Uhlenbeck process on $[0,3]$. The parameters of the diffusion are: $\gamma \equiv 1$, $\mu \equiv 0$ and $\sigma_0 = 100$.

4.2 Quantization-based cubature

4.2.1 Basic formula and related inequalities

The idea of quantization-based cubature scheme is to approach the distribution of the random variable X by the distribution of a quantizer Y of X. As Y is a discrete random variable, we can write $\mathbb{P}_Y = \sum_{i=1}^N p_i \delta_{y_i}$. If $F : E \to \mathbb{R}$ is a Borel functional,

$$\mathbb{E}[F(Y)] = \sum_{i=1}^N p_i F(y_i). \tag{4.14}$$

Hence, if we have access to the weighted discrete distribution $(y_i, p_i)_{1\le i\le N}$ of Y, we are able to compute the sum (4.14). Now, we review some error bounds that can be derived when approaching $\mathbb{E}[F(X))]$ by the quantity (4.14). See [27] for more details on error bounds.

1. *If $X \in L^2$, Y a quantizer of X of size N and F is Lipschitz continuous, then*

$$|\mathbb{E}[F(X)] - \mathbb{E}[F(Y)]| \leq [F]_{\mathrm{Lip}} \|X - Y\|_2. \tag{4.15}$$

In particular, if $(Y_N)_{N\geq 1}$ is a sequence of quantizers such that $\lim_{N\to\infty} \|X - Y_N\|_2 = 0$, then the distribution $\sum_{i=1}^{N} p_i^N \delta_{x_i^N}$ of Y_N converges weakly to the distribution $\mathbb{P}_X$ of X as $N \to \infty$.

This first error bound is a straightforward consequence of $|F(X) - F(Y)| \leq [F]_{\mathrm{Lip}}|X - Y|$.

2. *If Y is a stationary quantizer of X, i.e. $Y = \mathbb{E}[X|Y]$, and F is differentiable with an α-Hölder differential DF ($\alpha \in (0, 1]$), then*

$$|\mathbb{E}[F(X)] - \mathbb{E}[F(Y)]| \leq [DF]_\alpha \|X - Y\|_2^{1+\alpha}. \tag{4.16}$$

When F has a Lipschitz continuous derivative ($\alpha = 1$), we have. $[DF]_1 = [DF]_{\mathrm{Lip}}$ and, if F is twice differentiable and D^2F is bounded, then we can replace $[DF]_{\mathrm{Lip}}$ by $\frac{1}{2}\|D^2F\|_\infty$.

This inequality comes from the Taylor expansion of F at X.

$$|F(X) - F(Y) - DF(Y).(X - Y)| \leq \sup_{z\in[X,Y]} \|DF(z) - DF(Y)\||X - Y| \leq [DF]_\alpha |X - Y|^{1+\alpha},$$

where $\|\cdot\|$ stands for the operator norm on $L(L^2([0, T]))$. Hence

$$|\mathbb{E}[F(X)] - \mathbb{E}[F(Y)] - \mathbb{E}\left[DF(Y).(X - Y)\right]| \leq [DF]_\alpha \mathbb{E}\left[|X - Y|^{1+\alpha}\right].$$

Moreover, by stationarity,

$$\mathbb{E}\left[DF(Y).(X - Y)\right] = \mathbb{E}\left[\mathbb{E}\left[DF(Y).(X - Y)|Y\right]\right] = \mathbb{E}\left[DF(Y).\mathbb{E}[X - Y|Y]\right] = 0. \tag{4.17}$$

As a consequence, stationarity results in a convergence rate of a higher order by one for the quantization-based cubature scheme.

3. *If F is a convex functional and Y is a stationary quantizer of X,*

$$\mathbb{E}[F(Y)] \leq \mathbb{E}[F(X)]. \tag{4.18}$$

This inequality is a straightforward consequence of the stationarity property and the Jensen inequality.

$$\mathbb{E}[F(Y)] = \mathbb{E}[F(\mathbb{E}[X|Y])] \leq \mathbb{E}[\mathbb{E}[F(X)|Y]] = \mathbb{E}[F(X)].$$

4.2.2 Richardson-Romberg extrapolation

In a general setting, an asymptotic expansion of the error of a convergent numerical method makes possible the use of convergence speeding procedures like the Richardson-Romberg extrapolation method proposed in [11].

The quadratic N-optimal quantizer, Y_N of the uniform distribution on $[a, b]$ is showed to be the equiweighted codebook $\left(a + (i - \frac{1}{2})\frac{b-a}{N}\right)_{1\leq i\leq N}$. The associated quadrature rule coincides with the classical mid-point quadrature method. In the case of a $2n$ times differentiable function, using the Euler-Maclaurin formula, we can show that the quadrature error is the sum n entirely even powers of the step-size $\frac{b-a}{N}$, which is proportional to the quadratic quantization error $\|X - Y_N\|_2$. This makes possible the use of multistep Richardson-Romberg extrapolation in this special case.

In the general setting of a non-uniform random variable X, a quadratic-optimal N-quantizer Y_N of X and a C^1 functional with Lipschitz continuous derivative, Equation (4.16) does not provide

a true asymptotic expansion which would allow us to use a Richardson-Romberg expansion, but it suggests to use a higher-order Taylor expansion of $F(X) - F(Y_N)$ to get one.

We denote by $\mathcal{E}_N = \|X - Y_N\|_2$ the quantization error. It follows from Taylor's formula that there exists a vector $\zeta \in [X, Y_N]$ such that

$$\begin{aligned}\mathbb{E}[F(X)] &= \mathbb{E}\left[F(Y_N)\right] + \underbrace{\mathbb{E}\left[\langle DF(Y_N), X - Y_N\rangle\right]}_{=0 \text{ owing to (4.17).}} + \tfrac{1}{2}\mathbb{E}\left[D^2F(Y_N)(X - Y_N)^{\otimes 2}\right] \\ &\qquad + \tfrac{1}{6}\mathbb{E}\left[\zeta(X - Y_N)^{\otimes 3}\right] + o\left(\mathbb{E}\left[|X - Y_N|^3\right]\right) \\ &= \mathbb{E}\left[F(Y_N)\right] + \tfrac{1}{2}\mathbb{E}\left[D^2F(Y_N)(X - Y_N)^{\otimes 2}\right] + O\left(\mathbb{E}\left[|X - Y_N|^3\right]\right).\end{aligned} \tag{4.19}$$

In [13], Graf *et al.* proved that the asymptotics of the L^s quantization error induced by a sequence of L^r-optimal quantizers stays rate-optimal in the case of probability distributions on $\mathbb{R}^d$, with $s < r + d$ for a class of distributions including the Gaussian distribution. In this case, this leads to $\mathbb{E}\left[|X - Y_N|^3\right] = O\left(\mathbb{E}\left[|X - Y_N|^2\right]^{\frac{3}{2}}\right)$. This holds *e.g.* for the Brownian motion.

Unfortunately, no sharp equivalence between $\mathcal{E}_N^2$ and $\mathbb{E}\left[D^2F(Y_N)(X - Y_N)^{\otimes 2}\right]$ has been established yet. Still, Equation (4.19) suggests to use a Richardson-Romberg extrapolation with respect to $\mathcal{E}_N^2$. The two-steps extrapolation between $N = k$ and $N = l$ yields the following approximation of the expectation,

$$\mathbb{E}[F(X)] \approx \frac{\mathbb{E}[F(Y_l)]\mathcal{E}_k^2 - \mathbb{E}[F(Y_k)]\mathcal{E}_l^2}{\mathcal{E}_k^2 - \mathcal{E}_l^2}. \tag{4.20}$$

Although, this kind of Richardson-Romberg extrapolation is not fully justified yet, it dramatically increases the efficiency of quantization-based cubature formulas.
When the exact value of $\mathcal{E}_k^2$ is not known, we can also rely on the asymptotic expansion with respect to the quantization level.

Remark (Romberg extrapolation with respect to the quantization level)**.** *Using the results of [21], in a hypothesis on the asymptotics the eigenvalues of the covariance operator, the rate of convergence of optimal quantizers and K-L optimal product quantizers is* $\asymp (\ln(N)^{-\alpha})$ *with,*

- $\alpha = H$ *in the case of the fractional Brownian motion with Hurst exponent* H*.*
- $\alpha = \frac{1}{2}$ *in the case of the standard Brownian bridge or Ornstein-Uhlenbeck processes defined by the SDE* ($dr_t = \theta(mu - r_t)dt + \sigma dW_t$)*.*

Replacing the distortion $\mathcal{E}_N$ *by its asymptotics* $\frac{1}{\ln(N)^\alpha}$ *as* $N \to \infty$ *in Equation (4.20) yields the following estimator of the expectation.*

$$\mathbb{E}[F(X)] \approx \frac{\mathbb{E}[F(Y_l)](\ln l)^{2\alpha} - \mathbb{E}[F(Y_k)](\ln k)^{2\alpha}}{(\ln l)^{2\alpha} - (\ln k)^{2\alpha}}. \tag{4.21}$$

Numerical experiments carried out in [33] by Wilbertz showed that using the quadratic distortion $\mathcal{E}_n^2$ when available instead of its asymptotic form generally improves the efficiency of the extrapolation.

Experiments with higher-order Richardson-Romberg extrapolations have been tested. Unfortunately, no significant results were obtained, so that we settled for the two-steps extrapolation formula.

Remark (Choice of the Romberg extrapolation couples)**.** *We have noticed that the result of the extrapolation are usually improved when choosing the two quantizer sizes* k *and* l *that immediately follow a break of quantization dimension. That is the reason why the couple* $(208 - 54)$ *is used in the following for the case of the Brownian motion.*

4.3 Vanilla option pricing in stochastic volatility models

Here, we start with a special case of stochastic volatility model. For the sake of clarity, we will handle this simple case completely before considering a more general model. We assume that, under the risk-neutral measure, the forward price of a risky asset is the solution of the system of SDE

$$\begin{cases} dF_t = F_t\sigma_t dW_t, \quad F_0 > 0, \\ d\sigma_t = b(t,\sigma_t)dt + \theta(t,\sigma_t)dW_t^\sigma, \quad \sigma_0 > 0, \\ d\langle W, W^\sigma\rangle_t = \rho dt. \end{cases} \tag{4.22}$$

4.3.1 Conditioning with respect to the volatility

SDE resolution

The Brownian motions W is decomposed into W^σ and an independent standard Brownian motion W^F.

$$dW_s = \rho dW_s^\sigma + \sqrt{1-\rho^2}dW_s^F \qquad \text{where } W^\sigma \perp\!\!\!\perp W^F.$$

Let us denote by $(\mathcal{F}_t^\sigma)_{t\geq 0}$ and $(\mathcal{F}_t^F)_{t\geq 0}$ the filtrations of the Brownian motions W^σ and W^F. The solution of SDE (4.22), $F_t = F_0 \exp\left(\int_0^t \sigma_s dW_s - \frac{1}{2}\int_0^t \sigma_s^2 ds\right)$ can be decomposed into the following product.

$$F_t = \underbrace{F_0 \exp\left(\rho \int_0^t \sigma_s dW_s^\sigma - \frac{\rho^2}{2}\int_0^t \sigma_s^2 ds\right)}_{:=A_t} \underbrace{\exp\left(\sqrt{1-\rho^2}\int_0^t \sigma_s dW_s^F - \frac{1-\rho^2}{2}\int_0^t \sigma_s^2 ds\right)}_{:=B_t}, \tag{4.23}$$

where the process $(A_t)_{t\in[0,T]}$ is adapted to $\mathcal{F}^\sigma$.

Preconditioning

In this section, the function Payoff(x,K) can either be the function $(x-K)_+$ or $(K-x)_+$, the payoff of a Call or Put option with strike K. Following the method proposed *e.g.* in [27], the preconditioning with respect to $\mathcal{F}_T^\sigma$ yields a simple expression:

$$\begin{aligned} \mathbb{E}\left[\text{Payoff}(F_T,K)\right] &= \mathbb{E}\left[\mathbb{E}\left[\text{Payoff}(F_T,K)|\mathcal{F}_T^\sigma\right]\right] \\ &= \mathbb{E}\left[\text{PrimeBS}\left(A_T, \left((1-\rho^2)\int_0^T \sigma_s^2 ds\right)^{\frac{1}{2}}, T, K\right)\right], \end{aligned}$$

where A_T is defined in Equation (4.23) and where PrimeBS(F,σ,T,K) is the closed-form expression for the price of a Call or Put option in the Black & Scholes model, with no interest rate, a forward F, a volatility σ, a maturity T and a strike K.
Now, the pricing comes to a cubature problem with respect to the volatility path $(\sigma_s)_{s\in[0,T]}$. This cubature is performed by using the functional quantizer $(\widehat{\sigma}^i)_{1\leq i\leq N}$ of σ.

$$\mathbb{E}[\text{Payoff}(F_T,K)] \approx \sum_{i=1}^N p_i \text{PrimeBS}\left(A_T^i, \left((1-\rho^2)\int_0^T \widehat{\sigma}^i(s)^2 ds\right)^{\frac{1}{2}}, T, K\right), \tag{4.24}$$

where $(A_T^i)_{1\leq i\leq N}$ denotes the quantizer of A_T deduced from $(\widehat{\sigma}^i)_{1\leq i\leq N}$.

- In this equation, $(\alpha_i)_{1\leq i\leq N}$ and $(p_i)_{1\leq i\leq N}$ are the paths of a functional quantizer of W^σ (either an optimal quantizer or an optimal product quantizer) and its weights respectively. Functions $(\widehat{\sigma}^i)_{1\leq i\leq N}$ are the paths of the quantizer of σ obtained from $(\alpha_i)_{1\leq i\leq N}$ by solving the ODE's (4.11), derived from the diffusion equation (4.22) satisfied by the volatility (written in the Stratonovich sense).

- The corresponding values of $\left(\int_0^T \widehat{\sigma}^i(s)^2 ds\right)_{1\leq i\leq N}$ needed in formula (4.24) are deduced from this quantization.

- To compute the term A_T^i, we need to evaluate the quantized version of the stochastic integral $\int_0^T \sigma_s dW_s^\sigma = \int_0^T \sigma_s \circ dW_s^\sigma - \frac{1}{2}\int_0^T d\langle \sigma, W^\sigma\rangle_t = \int_0^T \sigma_s \circ dW_s^\sigma - \frac{1}{2}\int_0^T \theta(t,\sigma_t)dt$. This leads to the quantizer

$$A_T^i = F_0 \exp\left(\rho \int_0^T \widehat{\sigma}^i(t)\alpha_i'(t)dt - \frac{\rho}{2}\int_0^T \theta(t,\widehat{\sigma}_t^i)dt - \frac{\rho^2}{2}\int_0^T \widehat{\sigma}^i(t)^2 dt\right), \quad 1 \le i \le N.$$

The corresponding integrals on $[0,T]$ may be computed by classical quadrature methods.

First numerical experiments

Here, we perform a first numerical test in the special case where we set $b(t,\sigma_t) \equiv 0$ and $\theta(t,\sigma_t) \equiv \gamma\sigma_t$ in the SDE (4.22). In this case, the volatility is lognormal. An expansion for small maturities of the implied volatility is derived in [14] for this model. We compare the results of the quantization-based cubature with this short-maturity asymptotics which is known to be very accurate for reasonable values of the parameters.

In Table 4.1, we report the option prices obtained by the short-maturity asymptotics and the quantization-based cubature formula with different sizes of the quantizer and various values of the parameters.

Parameter values	Closed-form small-maturity asymptotics	208 curves crude cubature formula	$208-54$ Richardson-Romberg extrapolation
$K=100\%$, $T=1$ $\gamma=0.3$, $\sigma_0=0.3$ $\rho=-0.5$	11.8459	11.7394	11.8296
$K=130\%$, $T=1$ $\gamma=0.3$, $\sigma_0=0.3$ $\rho=-0.5$	3.0062	2.9215	3.0001
$K=70\%$, $T=1$ $\gamma=0.3$, $\sigma_0=0.3$ $\rho=-0.8$	31.9075	31.9645	32.0730
$K=100\%$, $T=1$ $\gamma=0.5$, $\sigma_0=0.3$ $\rho=0.2$	12.2440	11.9936	12.2277
$K=130\%$, $T=0.5$ $\gamma=1$, $\sigma_0=0.3$ $\rho=0$	2.0116	1.8217	1.9458

Table 4.1: Record of the prices obtained with crude (extrapolated) functional quantization-based cubature schemes, with various values of the diffusion parameters.

We can see that despite the slow (logarithmic) rate of decay of the quantization error in the functional case, and thanks to Romberg extrapolation, we could reach a good accuracy on the option price with a reasonable size of the quantizer. We will see in the next section that it can still be improved using some kind of variance reduction method.

A kind of "variance reduction method" for computing implied volatility

In the Black & Scholes model, where the asset price follows a geometric Brownian motion with a constant volatility, the vanilla option price is an increasing function of the volatility (if the strike is not zero). Conversely, for a given vanilla option price, the Black & Scholes implied volatility is the unique value of the volatility for which the Black & Scholes formula recovers the price; *i.e.* the implied volatility associated with a given forward F_0, maturity T, strike K, and option price

P is defined by

$$P = \mathbf{PrimeBS}\left(F_0, T, K, \mathbf{ImpliedVolBS}\left(F_0, T, K, P\right)\right). \tag{4.25}$$

Although the volatility is not constant on the market, the value of implied volatility is much easily interpretable than the crude price of an option, because the magnitude of the variations of the option price are strongly depending on the parameters (maturity, strike, forward).

On the other hand, for strikes which are deep out of the money, the sensitivity of the option price to the volatility is very small, which makes it difficult to recover the value of the volatility from a given price.

Because of this computational obstacle, and as practitioners are more interested by implied volatility than true option prices, many pricing approaches for stochastic volatility model directly focus on the dynamics of the implied volatility [14, 1, 17, 10].

A fact that we have experienced, is that the accuracy of the implied volatility obtained from the quantization-based cubature method is significantly improved when using the *estimated forward* rather than the theoretical forward to compute the implied volatility, and the resulting volatility smile is more regularly shaped. This remains true when using Richardson-Romberg extrapolation methods. Thus, we obtain more accurate results when following the steps:

1. Use the functional quantization-based (extrapolated) cubature formula to get a first estimation $P_{\text{Estimated}}$ of the option price $\mathbb{E}\left[\text{Payoff}(F_T, K)\right]$.
2. Use the same (extrapolated) cubature formula to get the estimation $F_{\text{Estimated}}$ of the forward price $\mathbb{E}\left[F_T\right]$.
3. Compute the Black & Scholes implied volatility of the option corresponding to these two values,
$$\sigma_{\text{Estimated}} = \mathbf{ImpliedVolBS}\left(F_{\text{Estimated}}, T, K, P_{\text{Estimated}}\right).$$
4. Return the Black & Scholes price corresponding to the theoretical forward and this implied volatility,
$$P_{\text{Final}} = \mathbf{PrimeBS}\left(F_0, T, K, \sigma_{\text{Estimated}}\right).$$

This can be understood as a variance reduction method for the functional quantization-based cubature. In Figure 4.3, the implied volatility smile estimated by the (extrapolated) functional quantization-based cubature formula, and the value computed by the closed-form short-maturity asymptotics are depicted.

The case of long maturities or (equivalently) high volatilities

When dealing with long maturities or equivalently, high volatilities, the short-maturity asymptotics of the SABR model derived in [14] becomes inaccurate. In this setting, the results of the cubature rule should be compared to a reference Monte-Carlo simulation. In the case of the SABR model, we noticed that the quantization-based method becomes more precise when the maturity increases whereas the short-maturity asymptotics naturally becomes less and less accurate. This phenomenon is illustrated in Figure 4.4, where the implied volatility smile is depicted, computed by the three methods: Monte-Carlo simulation, short-maturity asymptotics and (extrapolated) functional quantization-based cubature formula.

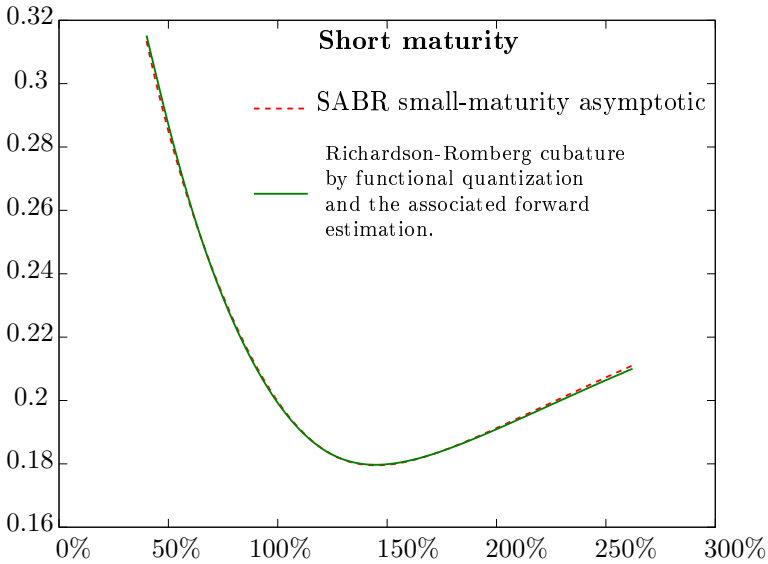

Figure 4.3: Implied volatility smile in the SABR model with $\beta = 1$, $\gamma = 0.3$, $\sigma_0 = 0.2$, $T = 1$ and $\rho = -0.5$. The continuous curve corresponds to a (208-54)-Richardson-Romberg extrapolation of the functional quantization-based cubature formula.

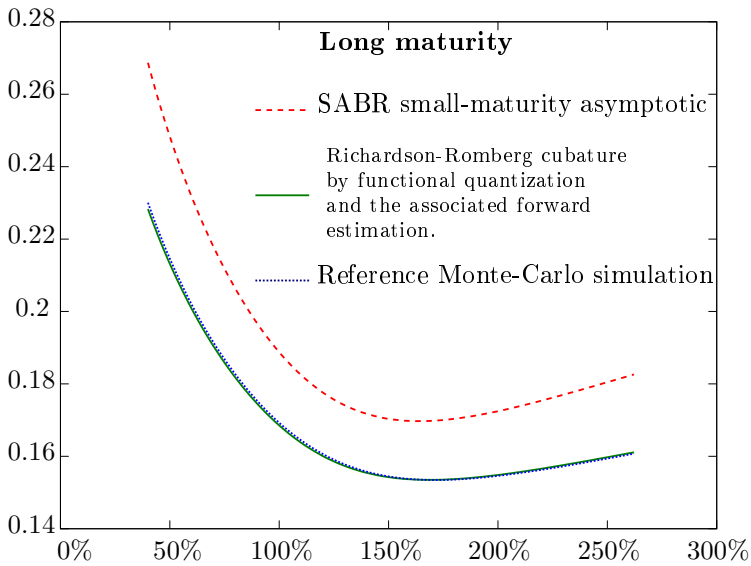

Figure 4.4: Implied volatility smile in the SABR model with $\beta = 1$, $\gamma = 0.3$, $\sigma_0 = 0.2$, $T = 20$ and $\rho = -0.5$. The continuous curve corresponds to a (208-54)-Richardson-Romberg extrapolation of the functional quantization-based formula.

4.4 Local stochastic volatility

We now consider a more general case. We assume that the forward diffusion model under the risk-neutral probability is

$$\begin{cases} dF_t = \sigma_t F_t g(t, F_t) dW_t, \\ d\sigma_t = b(t, \sigma_t)dt + \theta(t, \sigma_t)dW_t^\sigma. \end{cases} \tag{4.26}$$

where W and W^σ are standard Brownian motions. We assume that W is decomposed into $\rho dW_t^\sigma + \sqrt{1-\rho^2}dW_t^F$, where W^F is independent of W^σ. We denote respectively by $\mathcal{F}^F$ and $\mathcal{F}^\sigma$ the natural

filtrations of W^σ and W^F. Moreover, we assume that $b(t,x)$ and $\theta(t,x)$ are Borel functions, Lipschitz continuous in x uniformly in $t \in [0,T]$ and that $\theta(t,\cdot)$ is C^1 for every $t \in [0,T]$. We assume that $g(t,x)$ is a bounded Borel function such that $g(t,\cdot)$ is C^1.

This situation is very common. In the first place, many stochastic volatility models, such as the SABR model correspond to this situation. Furthermore, this is also the case of local stochastic volatility models which have been discussed in the introduction.

4.4.1 Stratonovich equation

Let us write the diffusion in the Stratonovich form, for the Brownian motion W^σ.

$$dF_t = \rho\sigma_t g(t,F_t)F_t \circ dW_t^\sigma + \sigma_t g(t,F_t)F_t\sqrt{1-\rho^2}dW_t^F - \frac{\rho}{2}d\left\langle \sigma g(\cdot,F)F, W^\sigma\right\rangle_t. \tag{4.27}$$

We now compute the quantity $\langle \sigma g(\cdot,F)F, W^\sigma\rangle$. If we denote $f(t,x) = xg(t,x)$, we have:

$$\begin{cases} d\sigma_t = b(t,\sigma_t)dt + \theta(t,\sigma_t)dW_t^\sigma, \\ df(t,F_t) = f'_x(t,F_t)dF_t + \begin{pmatrix} \text{finite-variation} \\ \text{process} \end{pmatrix}. \end{cases}$$

Thus,

$$\begin{aligned} d\langle \sigma f(\cdot,F), W^\sigma\rangle_t &= \sigma_t f'_x(t,F_t)d\langle F, W^\sigma\rangle_t + f(t,F_t)d\langle \sigma, W^\sigma\rangle_t \\ &= \sigma_t\left(g'_x(t,F_t)F_t + g(t,F_t)\right)d\langle F, W^\sigma\rangle_t + f(t,F_t)d\langle \sigma, W^\sigma\rangle_t \\ &= \rho\sigma_t^2 g(t,F_t)g'_x(t,F_t)F_t^2dt + \rho\sigma_t^2 g(t,F_t)^2F_tdt + f(t,F_t)d\langle \sigma, W^\sigma\rangle_t. \end{aligned}$$

We plug this expression into Equation (4.27) to get

$$\begin{cases} dF_t &= \sigma_t g(t,F_t)F_t\sqrt{1-\rho^2}dW_t^F + \sigma_t g(t,F_t)F_t\rho \circ dW_t^\sigma \\ &\quad -\frac{\rho^2}{2}\sigma_t^2 g(t,F_t)g'_x(t,F_t)F_t^2dt - \frac{\rho^2}{2}\sigma_t^2 g(t,F_t)^2F_tdt - \frac{\rho}{2}F_t g(t,F_t)d\langle \sigma, W^\sigma\rangle_t. \end{cases} \tag{4.28}$$

Moreover, using (4.26) we have $d\langle \sigma, W^\sigma\rangle_t = \theta(t,\sigma_t)dt$.

4.4.2 Plugging the functional quantizer

Let us now consider a $N_1 \times \cdots \times N_n$ product quantizer α of W^σ.
The path of α corresponding to the multi-index $\underline{i} := \{i_1, \cdots, i_n, \cdots\}$ has the form

$$\alpha^{\underline{i}} = \sum_{n\geq 1}\sqrt{\lambda_n^W}x_{i_n}^{N_n}e_n^W.$$

The path of σ corresponding to the multi-index $\underline{i}$ is defined as the solution of the ODE obtained when replacing W^σ by $\alpha^{\underline{i}}$ in the diffusion written in the Stratonovich sense (4.28). We replace the Brownian motion W^σ by its quantizer $\alpha^{\underline{i}}$ and σ by $\sigma^{\underline{i}}$ in Equation (4.28).

$$\begin{aligned} dF_t^{\underline{i}} &= \underbrace{F_t^{\underline{i}}\sigma_t^{\underline{i}}g\left(t,F_t^{\underline{i}}\right)\sqrt{1-\rho^2}}_{:=F_t^{\underline{i}}\theta_{\underline{i}}\left(t,F_t^{\underline{i}}\right)}dW_t^F \\ &\qquad \underbrace{\begin{aligned} &+ \sigma_t^{\underline{i}}g\left(t,F_t^{\underline{i}}\right)F_t^{\underline{i}}\rho\left(\alpha^{\underline{i}}\right)'(t)dt - \frac{\left(\rho\sigma_t^{\underline{i}}\right)^2}{2}g\left(t,F_t^{\underline{i}}\right)g'_x\left(t,F_t^{\underline{i}}\right)\left(F_t^{\underline{i}}\right)^2dt \\ &\qquad - \frac{\left(\rho^2\sigma_t^{\underline{i}}\right)^2}{2}g\left(t,F_t^{\underline{i}}\right)^2F_t^{\underline{i}}dt - \frac{\rho}{2}F_t^{\underline{i}}g\left(t,F_t^{\underline{i}}\right)\theta\left(t,\sigma_t^{\underline{i}}\right)dt \end{aligned}}_{:=F_t^{\underline{i}}\mu_{\underline{i}}\left(t,F_t^{\underline{i}}\right)dt} \\ &= F_t^{\underline{i}}\mu_{\underline{i}}\left(t,F_t^{\underline{i}}\right)dt + F_t^{\underline{i}}\theta_{\underline{i}}\left(t,F_t^{\underline{i}}\right)dW_t^F. \end{aligned} \tag{4.29}$$

In other words, $F_t^{\underline{i}}$ has a local volatility and a local drift.

4.4.3 Preconditioning

As we have seen in Section 4.4.2, we are facing a set of diffusion equations with local volatility and local drift. (See Equation (4.29).) We now try to follow the steps of Section 4.3.1.

$$\mathbb{E}\left[(F_t - K)_+\right] = \mathbb{E}\left[\underbrace{\mathbb{E}\left[(F_t - K)_+ | \mathcal{F}_T^\sigma\right]}_{=\phi((W^\sigma)_{t\in[0,T]})}\right]. \tag{4.30}$$

This expectation is approximated with the cubature formula based on functional quantization.

$$\mathbb{E}[(F_t - K)_+] \approx \sum_{i=1}^{N} p_i \phi(\alpha_i), \tag{4.31}$$

where $(\alpha_i)_{1\leq i\leq N}$ and $(p_i)_{1\leq i\leq N}$ are respectively the paths and the associated weights of a stationary functional quantizer of W^σ. The value $\phi(\alpha_i)$ in Equation (4.31) corresponds to the price of a Call or Put option with a local volatility and a local drift as in Equation (4.29). Unfortunately, no closed-form expression exists for this general case. However, we can use a common numerical method as a finite difference scheme for the associated (one-dimensional) Kolmogorov backward PDE (4.32) to overcome this difficulty.

$$\begin{cases} \frac{1}{2}\theta_{\underline{i}}(t,x)^2 x^2 \frac{\partial^2 V}{\partial x^2} + \mu_{\underline{i}}(t,x) x \frac{\partial V}{\partial x}(t,x) + \frac{\partial V}{\partial t} = 0 \\ V(T,x) = \text{Payoff}(T,x). \end{cases} \tag{4.32}$$

We obtain the same order of precision as in the previous case but with a longer computation time.

- In other words, we can see the functional quantization of the volatility as a way to approximate the value of the solution of a two-dimensional PDE with a weighted sum of the values of solutions of one-dimensional partial differential equations.
- This is easily extended to multi-factor stochastic volatility models.

In Figure 4.5 we plot the implied volatility smile in the SABR model, computed with the small-maturity asymptotic and with the quantization-based cubature formula, coupled with a simple Monte-Carlo simulation.
As an alternative for the PDE approach, we now introduce a new quantization scheme for stochastic differential equations which we call "normal quantization". This approximation is based on a recent approach to functional quantization called "partial functional quantization" which has been introduced in [5]. In Section 4.4.4 we recall some background on partial functional quantization. Then we come to the normal quantization of stochastic differential equations.

4.4.4 Normal functional quantization and stochastic differential equations

When plugging a stationary quantizer in place of the Gaussian process in a multidimensional SDE written in the Stratonovich sense, one obtains a quantizer of the solution of the SDE. It has been shown in [29] that the solutions of the quantized solutions of the ODE converge toward the solution of the SDE when the quantizer size goes to infinity. A first approach to this problem (but in a more restrictive setting) has been done in [23] for the one-dimensional setting, using the so-called Lamperti transform. This is the same transformation which was originally used by Wong and Zakai in [35]. It is also related to some articles of Doss and Sussman on the connection between stochastic and ordinary differential equations [8, 9, 32]. In [23], this transformation was used to prove some contractivity properties of the Itô map under some constraints on the parameters of the SDE.

In this section, we first briefly come back on these contractivity properties and the Lamperti transform. Then we recall the main results on partial functional quantization of stochastic differential equations, which was first introduced in [5]. We finally introduce the normal quantization of solutions of stochastic differential equations.

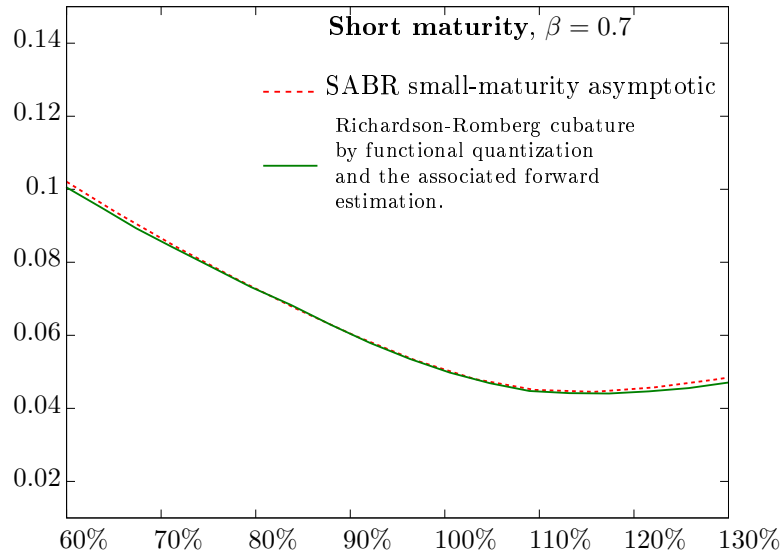

Figure 4.5: Implied volatility smile in the SABR model with $\beta = 0.7$, $\gamma = 0.3$, $\sigma_0 = 0.2$, $T = 1$ and $\rho = -0.5$. The continuous curve corresponds to a (208-54)-Richardson-Romberg extrapolation of the functional quantization-based cubature formula.

The Lamperti transform and contractivity of the Itô map

Let us consider the SDE

$$dS_t = b(t, S_t)dt + \theta(t, S_t)dX_t, \qquad S_0 = x \in \mathbb{R}, \qquad \text{and } t \in [0, T], \tag{4.33}$$

where $b(t, x)$ and $\theta(t, x)$ are Borel functions, Lipschitz continuous with respect to x uniformly in t, θ is bounded and $|b(\cdot, 0)|$ is bounded. This SDE admits a unique strong solution S.

Definition 4.4.1 (Lamperti transform)**.** *Consider $\mathcal{I}$ an open interval, $x_0 \in \mathcal{I}$ and $\theta : [0, T] \times \mathcal{I} \to \mathbb{R}_+^*$ a function satisfying the ($\mathcal{L}$) hypothesis defined bellow.*

$$\left\{\begin{array}{ll} (i) & \theta \in C^1([0, T] \times \mathcal{I}, \mathbb{R}_+^*), \\ (ii) & \forall (t, x) \in [0, T] \times \mathcal{I},\ 0 < \theta(t, x) \leq C(1 + |x|), \\ (iii) & \text{if } \mathcal{I} \neq \mathbb{R} \text{ we make the additional assumption that} \\ & \forall t \in [0, T], \quad \int_{[x_0, +\infty) \cap \mathcal{I}} \frac{d\xi}{\theta(t, \xi)} = +\infty \text{ and } \int_{(-\infty, x_0] \cap \mathcal{I}} \frac{d\xi}{\theta(t, \xi)} = +\infty. \end{array}\right. \tag{$\mathcal{L}$}$$

The Lamperti transform associated with θ and x_0 is defined by $S(t, x) = \int_{x_0}^{x} \frac{d\xi}{\theta(t, \xi)}$ on $[0, T] \times \mathcal{I}$.

Under the assumption ($\mathcal{L}$), the so-defined function S is $C^{1,2}([0, T] \times \mathcal{I})$ with

$$\frac{\partial S}{\partial t}(t, x) = -\int_{x_0}^{x} \left(\frac{1}{\theta^2} \frac{\partial \theta}{\partial t}\right)(t, \xi) d\xi,$$

$$\frac{\partial S}{\partial x}(t, x) = \frac{1}{\theta(t, x)} \quad \text{and} \quad \frac{\partial^2 S}{\partial x^2}(t, x) = -\left(\frac{1}{\theta^2} \frac{\partial \theta}{\partial x}\right)(t, x).$$

Moreover, for every $t \in [0, T]$, $x \mapsto S(t, x)$ is continuous and strictly increasing in $\mathcal{I}$.

Remark. *It follows from hypothesis (($\mathcal{L}$),(ii)) that $|S(t, x)| \geq \frac{1}{C} \log(1 + |x|)$. Hence, if $\mathcal{I} = \mathbb{R}$, claim (($\mathcal{L}$),(iii)) is a consequence of the first two assumptions.*

A consequence is that, for every $t \in [0,T]$, $S(t,\cdot)$ has a (continuous) inverse function defined on $\mathbb{R}$. This inverse function, denoted by S_t^{-1} satisfies $S_t^{-1}(x_0) = 0$ and $\left|S_t^{-1}(y)\right| \leq e^{C|y|} - 1$ for every $y \in \mathbb{R}$. Furthermore, S_t^{-1} is differentiable and $\left(S_t^{-1}\right)'(y)$ satisfies $0 < \left(S_t^{-1}\right)'(y) = \theta\left(t, S_t^{-1}(y)\right) \leq C\left(1 + |S_t^{-1}(y)|\right) \leq Ce^{C|y|}$. Hence, $\forall t \in [0,T]$, $\forall (y,y') \in \mathbb{R}^2$, $\left|S_t^{-1}(y') - S_t^{-1}(y)\right| \leq Ce^{C\max(|y|,|y'|)}|y - y'|$.

Proposition 4.4.1. *If θ is bounded on $[0,T] \times \mathcal{I}$ by $\|\theta\|_{\max}$, we easily prove that S_t^{-1} is $\|\theta\|_{\max}$-Lipschitz continuous, namely*

$$\forall t \in [0,T], \quad \forall (y,y') \in \mathbb{R}^2, \left|S_t^{-1}(y') - S_t^{-1}(y)\right| \leq \|\theta\|_{\max}|y - y'|.$$

Moreover $(t,y) \mapsto S_t^{-1}(y)$ is continuous on $[0,T]\times\mathbb{R}$, since the sets $\left\{(t,y) : S_t^{-1}(y) \gtreqless c\right\} = \left\{(t,y) : y \gtreqless S(t,c)\right\}$ are closed for every $c \in \mathbb{R}$.

Applying Itô's formula to $Y_t := S(t,F_t)$ yields

$$dY_t = \underbrace{\left(\frac{b}{\theta}(t,F_t) - \int_{x_0}^{F_t}\left(\frac{1}{\theta^2}\frac{\partial\theta}{\partial t}\right)(t,\xi)d\xi\right)dt + \frac{1}{2}\frac{\partial\theta}{\partial x}(t,F_t)d\langle X\rangle_t}_{\text{finite-variation part}} + dX_t. \tag{4.34}$$

The Lamperti transform fulfills its task to yield a SDE with a constant diffusion coefficient equal to 1.

Remark. *Under the additional assumption that the measure $d\langle X\rangle$ is absolutely continuous with respect to the Lebesgue measure on $[0,T]$, with a density $\tau_X := \frac{d\langle X\rangle}{d\lambda}$ Equation (4.34) becomes $dY_t = \beta(t,Y_t)dt + dX_t$, with*

$$\beta(t,y) := \left(\frac{b}{\theta} - \int_{x_0}^{S_t^{-1}(y)}\left(\frac{1}{\theta^2}\frac{\partial\theta}{\partial t}\right)(t,\xi)d\xi + \frac{1}{2}\frac{\partial\theta}{\partial x}\left(t, S_t^{-1}(y)\right)\tau_X(t)\right). \tag{4.35}$$

Therefore, $\beta : [0,T] \times \mathbb{R} \to \mathbb{R}$ is continuous as soon as b and τ_X are.

Study of the transformed SDE (4.34)
Let $p \in [0,\infty]$ and $h \in L_t^p$. We consider the integral equation in $(L_T^p, \|\cdot\|_p)$

$$y(t) = y_0 + \int_0^t\left(\frac{b}{\theta}(s,y(s)) - \int_{x_0}^{y(s)}\left(\frac{1}{\theta^2}\frac{\partial\theta}{\partial s}\right)(s,\xi)d\xi\right)ds + \frac{1}{2}\int_0^t\frac{\partial\theta}{\partial x}(s,y(s))d\langle X\rangle_s + h(t). \tag{4.36}$$

The existence and uniqueness of a solution for (4.36) in L_T^p follows from the approach used for Ordinary Differential Equations, assuming that the mapping $H_p : L_T^p \to L_T^p$ defined by

$$H_p(y) \equiv t \mapsto \int_0^t\left(\frac{b}{\theta}(s,y(s)) - \int_{x_0}^{y(s)}\left(\frac{1}{\theta^2}\frac{\partial\theta}{\partial s}\right)(s,\xi)d\xi\right)ds + \frac{1}{2}\int_0^t\frac{\partial\theta}{\partial x}(s,y(s))d\langle X\rangle_s + h(t)$$

is contractive for small enough T.

Remark (Contractivity of H_p)**.** *A sufficient condition for H_p to be contractive, in the case where $d\langle X\rangle$ is absolutely continuous with respect to the Lebesgue measure, is the function $\beta(t,x)$ to be Lipschitz continuous with respect to x uniformly in t. In this setting H_p will be contractive for $T[\beta]_{\text{Lip}} < 1$.*

Thus, H_p has a unique fixed point in L_T^p for $T[\beta]_{\text{Lip}} < 1$. A global solution of (4.36) in L_T^p for any fixed $T > 0$ can be constructed inductively by simply sticking pieces of solutions on intervals $[kT_0, (k+1)T_0]$ with $T[\beta]_{\text{Lip}} < 1$ starting at the appropriate values. Note that the resulting solution does not depend on p as long as $h \in L_T^p$ and that when h and τ_X are continuous, the solution y is continuous too.

Using standard Gronwall techniques and the inequality $(u+v)^p \leq 2^{p-1}(u^p + v^p)$ when $p \in [1,\infty)$, we show that the mapping $\Psi_{y_0} : L_T^p \to L_T^p$ by

"$\Psi_{y_0}(h)$ *is the unique solution of (4.36) in* L^p_T"

satisfies $Y = \Psi_{y_0}(X)$ and is $\|\cdot\|_p$-Lipschitz continuous.

Proposition 4.4.2 (Contractivity of the Itô map)**.** *The main consequence of this, is that the Itô map, defined as the composition* $S_t^{-1} \circ H_p$ *is contractive on* $[0,T]$ *for small enough* T *as soon as* θ *is bounded and* $\beta(t,x)$ *is Lipschitz continuous with respect to* x *uniformly in* t.

Partial functional quantization and generalized bridges

Here, we recall the definitions and the main results related to the partial functional quantization of continuous Gaussian semimartingales, which was introduced in [5].
Let $(X_t)_{t\in[0,T]}$ be a continuous centered Gaussian semimartingale starting from 0. In this case, the continuity assumption on the Gaussian process ensures, thanks to Fernique's theorem that $\int_0^T \mathbb{E}[X_t^2]dt < \infty$ and also ensures the continuity of the covariance function. (See [18, VIII.3]). We denote by $(e_i^X, \lambda_i^X)_{i\geq 1}$ the Karhunen-Loève eigensystem of X.

Let I be a finite subset of $\mathbb{N}$. We consider $\overline{Z}_T = (Z_T^i)_{i\in I}$ the terminal values of processes of the form $Z_t^i = \int_0^t f_i(s)dX_s,\ i \in I$, for some given finite set $\overline{f} = (f_i)_{i\in I}$ of $L^2_{loc}([0,T])$ functions.

Definition 4.4.2 (Generalized bridge)**.** *A generalized bridge for* $(X_t)_{t\in[0,T]}$ *corresponding to* $\overline{f}$ *with end-point* $\overline{z} = (z_i)_{i\in I}$ *is a process* $\left(X^{\overline{f},\overline{z}}\right)_{t\in[0,T]}$ *that has the distribution* $\mathcal{L}\left(X \middle| Z_T^i = z_i, i \in I\right)$.

For example, in the case where X is a standard Brownian motion with $|I| = 1$, $\overline{f} = \{f\}$ and $f \equiv 1$, this is the Brownian bridge on $[0,T]$. If X is an Ornstein-Uhlenbeck process this is an Ornstein-Uhlenbeck bridge.

Definition 4.4.3 (K-L generalized bridge)**.** *A K-L generalized bridge is a generalized bridge associated with the set of functions* $\left(f_i^X\right)_{i\in I}$ *defined by* $f_i^X(t) := \int_t^T e_i^X(s)ds$, *where* $(e_i^X)_{i\geq 1}$ *are the Karhunen-Loève eigenfunctions of* X.

In the case of a K-L generalized bridge, an integration by parts shows that $\int_0^T X_s e_i^X(s)ds = \int_0^T f_i^X(s)dX_s$ and thus $Y_i := \int_0^T X_s e_i^X(s)ds = Z_T^i$.

The Karhunen-Loève expansion gives the decomposition

$$X = \underbrace{\sum_{i\in I} Y_i e_i^X}_{\mathbb{E}[X|Y]} \overset{\perp}{+} \underbrace{\sum_{i\in\mathbb{N}^*\setminus I} \sqrt{\lambda_i^X}\xi_i e_i^X}_{\substack{\text{K-L generalized bridge} \\ \text{with end-point } \overline{0}}}, \tag{4.37}$$

where $(\xi_i)_{i\in\mathbb{N}^*\setminus I}$ are independent standard Gaussian random variables. It follows from (4.37) that a K-L generalized bridge is centered on $\mathbb{E}\left[X|Y_i = y_i, i \in I\right]$ and has the covariance function

$$\Gamma^{X|Y}(s,t) = \operatorname{cov}(X_s, X_t) - \sum_{i\in I} \lambda_i^X e_i^X(s) e_i^X(t). \tag{4.38}$$

We have $\int_0^T \Gamma^{X|Y}(t,t)dt = \sum_{i\in\mathbb{N}^*\setminus I} \lambda_i^X$. Moreover, thanks to decomposition (4.37), if $X^{I,\overline{y}}$ is a K-L generalized bridge associated with X with terminal values $\overline{y} = (y_i)_{i\in I}$, it has the same distribution as the process

$$\sum_{i\in I} y_i e_i^X(t) + X_t - \sum_{i\in I} \left(\int_0^T X_s e_i^X(s)ds\right) e_i^X(t).$$

This process is then the sum of a semimartingale and a non-adapted finite-variation process. Let us define the matrix $Q(s,T)$ for $s \in [0,T]$ by

$$Q(s,T) := \mathbb{E}\left[\left(\overline{Z}_T - \mathbb{E}\left[\overline{Z}_T \middle| (X_u)_{u\in[0,s]}\right]\right)\left(\overline{Z}_T - \mathbb{E}\left[\overline{Z}_T \middle| (X_u)_{u\in[0,s]}\right]\right)^* \middle| (X_u)_{u\in[0,s]}\right].$$

We make the additional assumption $(\mathcal{H})$ that

$$Q(s,T) \quad \text{is invertible for every } s \in [0,T). \tag{$\mathcal{H}$}$$

Remark (On the $(\mathcal{H})$ hypothesis). *In [5], it is proved that the $(\mathcal{H})$ hypothesis holds in the cases of K-L generalized bridges of the Brownian motion, the Brownian bridge and Ornstein-Uhlenbeck processes.*

Theorem 4.4.3 (Generalized bridges as semimartingales). *Let us assume that $\mathcal{F}^X$ is a Brownian filtration. We define the filtration $\mathcal{G}^X$ by $\mathcal{G}_t^X = \sigma\left(\overline{Z}_T, \mathcal{F}_t^X\right)$, the enlargement of the filtration $\mathcal{F}^X$ corresponding to the above conditioning. Under the $(\mathcal{H})$, the generalized bridge $X^{\overline{f},I}$ is a continuous $\mathcal{G}^X$-semimartingale on $[0,T)$ (up to a modification).*

A detailed proof, using filtration enlargement techniques is provided in [5] in a slightly more general setting.

Now, in the same set of notation, we consider the orthogonal decomposition (4.37) again and $\widehat{Y}^\Gamma = \mathrm{Proj}_\Gamma(Y)$ a stationary Voronoi N quantization of Y. (Proj_Γ is a nearest neighbor projection on Γ.)
We now define the stochastic process $\widetilde{X}^{I,\Gamma}$ by replacing Y by $\widehat{Y}^\Gamma$ in the decomposition (4.37). We denote $\widetilde{X}^{I,\Gamma} = \mathrm{Proj}_{I,\Gamma}(X)$.

$$\widetilde{X}^{I,\Gamma} = \sum_{i\in I} \widehat{Y}_i^\Gamma e_i^X \quad \overset{\perp\!\!\!\perp}{+} \sum_{i\in\mathbb{N}^*\setminus I} \sqrt{\lambda_i^X}\xi_i e_i^X. \tag{4.39}$$

The conditional distribution of $\widetilde{X}^{I,\Gamma}$ given that Y falls in the Voronoi cell of γ_k is the distribution of the K-L generalized bridge with end-point γ_k. In other words, we have quantized the Karhunen-Loève coordinates of X corresponding to $i \in I$, and not the other ones. The so-defined process $\widetilde{X}^{I,\Gamma}$ is called a *partial functional quantization of X*. Theorem 4.4.3 suggests to define the partial quantization of the solution S of the SDE (4.33) from a partial quantization $\widetilde{X}^{I,\Gamma}$ of X by replacing X by $\widetilde{X}^{I,\Gamma}$ in the SDE (4.33). We define the *partial quantization* $\widetilde{S}^{I,\Gamma}$ as the process whose conditional distribution given that Y falls in the Voronoi cell of γ_k is the strong solution of the same SDE where X is replaced by the K-L generalized bridge with end-point γ_k. We write

$$d\widetilde{S}_t^{I,\Gamma} = b\left(t, \widetilde{S}_t^{I,\Gamma}\right) dt + \theta\left(t, \widetilde{S}_t^{I,\Gamma}\right) d\widetilde{X}_t^{I,\Gamma}.$$

Theorem 4.4.4 (L^p mean quantization error of partially quantized SDE). *Let X be a continuous centered Gaussian martingale on $[0,T]$ with $X_0 = 0$. Let S be the strong solution of the SDE*

$$dS_t = b(t,S_t)dt + \theta(t,S_t)dX_t, \quad S_0 = x,$$

where $b(t,x)$ and $\theta(t,x)$ are Borel functions, Lipschitz continuous with respect to x uniformly in t, θ and $|b(\cdot,0)|$ are bounded.
We consider $\widetilde{X}^{I,\Gamma}$ a stationary partial functional quantization of X and $\widetilde{S}^{I,\Gamma}$ the corresponding partial functional quantization of S, i.e. the strong solutions of

$$d\widetilde{S}_t^{I,\Gamma} = b\left(t, \widetilde{S}_t^{I,\Gamma}\right) dt + \theta\left(t, \widetilde{S}_t^{I,\Gamma}\right) d\widetilde{X}_t^{I,\Gamma}, \quad \widetilde{S}_0^{I,\Gamma} = x.$$

Then, for every $p \in (0,\infty)$, $\varepsilon > 0$ and $t \in [0,T)$, there exists a positive constant $K_{p,\varepsilon,t,I}^X$ such that

$$\left\| \sup_{v\in[0,t]} \left| S_v - \widetilde{S}_v^{I,\Gamma} \right| \right\|_p \le K_{p,\varepsilon,t,I}^X \left(\left\| Y - \widehat{Y}^\Gamma \right\|_{p+\varepsilon} \right), \tag{4.40}$$

where Y is defined from X by Equation (4.37) and $\widehat{Y}^\Gamma$ is the nearest neighbor projection on Γ.

A detailed proof is available in [5]. It is also possible to handle the case where we have no stationarity hypothesis, when dealing with a convergent sequence of quantizers $\left(\widehat{Y}^N\right)_{N\ge1}$ for large enough N. For more details on this, see [5].

From partial functional quantization to functional quantization

If X is a continuous centered Gaussian semimartingale, we consider I a finite subset of $\mathbb{N}^*$ and $\widetilde{X}^{I,\Gamma}$ the corresponding optimal quadratic K-L partial functional quantization. Moreover, setting to 0 the other Karhunen-Loève coordinates, we denote by $\widehat{X}^{I,\Gamma}$ the associated quadratic optimal functional quantization. From Equation (4.37, we deduce

$$\left\|X-\widehat{X}^{I,\Gamma}\right\|_2^2=\left\|X-\widetilde{X}^{I,\Gamma}\right\|_2^2+\underbrace{\left\|\widetilde{X}^{I,\Gamma}-\widehat{X}^{I,\Gamma}\right\|_2^2}_{=\left\|X^{I,\overline{0}}\right\|_2^2=\sum_{i\in\mathbb{N}^*\setminus I}\lambda_i^X},$$

where $X^{I,\overline{0}}$ is the K-L generalized bridge associated with I and X and with end-point $\overline{0}$.

Now we consider the SDE (4.33) again, and respectively $\widetilde{S}^{I,\Gamma}$ and $\widehat{S}^{I,\Gamma}$ the partial functional quantization and the functional quantization of S associated with I and Γ. Using the triangle inequality, we have

$$\left\|S-\widehat{S}^{I,\Gamma}\right\|_p\leq\left\|S-\widetilde{S}^{I,\Gamma}\right\|_p+\left\|\widetilde{S}^{I,\Gamma}-\widehat{S}^{I,\Gamma}\right\|_p. \tag{4.41}$$

The term $\left\|S-\widetilde{S}^{I,\Gamma}\right\|_p$ can be controlled thanks to Theorem 4.4.4. Moreover, when $d\langle X\rangle$ is absolutely continuous with respect to the Lebesgue measure λ on $[0,T]$ with $\tau_X:=\frac{d\langle X\rangle}{d\lambda}$, and the function $\beta(t,y)$ defined in Equation (4.35) is Lipschitz continuous with respect to x uniformly in t, the term $\left\|\widetilde{S}^{I,\Gamma}-\widehat{S}^{I,\Gamma}\right\|_p$ can be controlled by $\left\|\widetilde{X}^{I,\Gamma}-\widehat{X}^{I,\Gamma}\right\|_p$ thanks to a contractivity property of the Itô map (Proposition 4.4.2). In this setting, we finally obtain

$$\left\|\widetilde{S}^{I,\Gamma}-\widehat{S}^{I,\Gamma}\right\|_p=O\left(\left\|X^{I,\overline{0}}\right\|_p\right). \tag{4.42}$$

Then, plugging this into Equation (4.41), we obtain an upper bound error for the quantization error for the solution of the stochastic differential equation.

The normal quantization of stochastic differential equations

We first briefly recall some inequalities which will be useful in the sequel.

Proposition 4.4.5 (Doob's inequality)**.** *Let M be a continuous local martingale on $\mathbb{R}_+$ with $M_0=0$. Then for every $T>0$,*

$$\mathbb{E}\left[\sup_{t\in[0,T]}M_t^2\right]\leq 4\mathbb{E}\left[\langle M\rangle_T\right].$$

Proposition 4.4.6 (Some inequalities related to the Gaussian distribution)**.** *Let G be a standard Gaussian random variable valued in $\mathbb{R}$. Consider $M>0$. We have*

$$\mathbb{E}\left[G^2\mathbf{1}_{|G|>M}\right]=\frac{2M\exp\left(-\frac{M^2}{2}\right)}{\sqrt{2\pi}}+2\mathcal{N}(-M). \tag{4.43}$$

Moreover

$$\mathcal{N}(-M)=\mathbb{P}[G>M]\leq\frac{1}{2}\exp\left(-\frac{M^2}{2}\right). \tag{4.44}$$

Thus

$$\mathbb{E}\left[G^2\mathbf{1}_{|G|>M}\right]\leq\left(\frac{2M}{\sqrt{2\pi}}+1\right)\exp\left(-\frac{M^2}{2}\right)$$

Under the additional assumption that $M>1$, we obtain

$$\mathbb{E}\left[G^2\mathbf{1}_{|G|>M}\right]\leq M\left(\frac{2}{\sqrt{2\pi}}+1\right)\exp\left(-\frac{M^2}{2}\right)\quad\text{if}\quad M>1.$$

Proof: The proof of Equality (4.43) is left to the reader. Inequality (4.44) comes from

$$\begin{aligned}\mathcal{N}(-M) = \mathbb{P}[G > M] &= \tfrac{1}{\sqrt{2\pi}} \int_M^\infty e^{-\frac{x^2}{2}} dx = \tfrac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{(x+M)^2}{2}} dx \\ &= \tfrac{1}{\sqrt{2\pi}} e^{-\frac{M^2}{2}} \int_0^\infty e^{-Mx} e^{-\frac{x^2}{2}} dx \leq \tfrac{1}{\sqrt{2\pi}} e^{-\frac{M^2}{2}} \int_0^\infty e^{-\frac{x^2}{2}} dx = \tfrac{1}{2} e^{-\frac{M^2}{2}}.\end{aligned}$$

And the proof of the last claim is straightforward. □

Proposition 4.4.7 (The non-standard case and reverse inequality)**.** *If $H := \sigma G$ has a variance of σ^2, we obtain*

$$\mathbb{E}\left[H^2 \mathbf{1}_{|H|>M}\right] = \sigma^2 \left[G^2 \mathbf{1}_{|G|>\frac{M}{\sigma}}\right] = \frac{2\sigma M}{\sqrt{2\pi}} \exp\left(-\frac{M^2}{2\sigma^2}\right) + 2\sigma^2 \mathcal{N}\left(-\frac{M}{\sigma}\right) \leq \left(\frac{2\sigma M}{\sqrt{2\pi}} + \sigma^2\right) \exp\left(-\frac{M^2}{2\sigma^2}\right).$$

And if $M > 1$, we get $\mathbb{E}\left[H^2 \mathbf{1}_{|H|>M}\right] \leq \underbrace{\left(\frac{2\sigma}{\sqrt{2\pi}} + \sigma^2\right) M \exp\left(-\frac{M^2}{2\sigma^2}\right)}_{:=\eta_M}$. Conversely, for some fixed $\eta > 0$, and if $M > 1$, we have

$$M \geq \underbrace{\sqrt{-\mathcal{W}\left(-\frac{\eta^2}{\left(\frac{2\sigma}{\sqrt{2\pi}} + \sigma^2\right)^2}\right)}}_{:=M_\eta} \quad \Rightarrow \quad \eta_M \leq \eta$$

where $\mathcal{W}$ is the Lambert $\mathcal{W}$ function.

Definition 4.4.4 (Normal quantization)**.** *Let X be a continuous centered Gaussian martingale on $[0,T]$. Let I be a finite subset of $\mathbb{N}^*$. We reconsider the decomposition (4.37), $X = \sum_{i\in I} Y_i e_i^X \overset{\perp\!\!\!\perp}{+} \sum_{i\in\mathbb{N}^*\setminus I} \sqrt{\lambda_i^X} \xi_i e_i^X$.*

Let Γ be a stationary codebook for $Y = (Y_i)_{i\in I}$ and $\widehat{Y}^\Gamma$ the corresponding Voronoi quantizer of Y. We denote by $\widehat{X}^{I,\Gamma}$ and $\widetilde{X}^{I,\Gamma}$ the corresponding functional quantization and partial functional quantization of X.

Let S be the strong solution of the SDE

$$dS_t = b(t, S_t)dt + \theta(t, S_t)dX_t, \quad S_0 = x, \tag{4.45}$$

where $b(t,x)$ and $\theta(t,x)$ are Borel functions, Lipschitz continuous with respect to x uniformly in t, θ and $|b(\cdot,0)|$ are bounded. Then we denote by

- *$\widehat{S}^{I,\Gamma}$ the corresponding functional quantization of S, obtained by replacing X by $\widehat{X}^{I,\Gamma}$ in the SDE (4.45) written in the Stratonovich form.*
- *$\widetilde{S}^{I,\Gamma}$ the corresponding partial functional quantization of S, obtained by replacing X by $\widetilde{X}^{I,\Gamma}$ in the SDE (4.45) as derived above and in [5].*

Thanks to Equation (4.39), the process $\widetilde{X}_t^{I,\Gamma}$ is decomposed into $\widetilde{X}_t^{I,\Gamma} = \widehat{X}_t^{I,\Gamma} \overset{\perp\!\!\!\perp}{+} X^{I,\overline{0}}$, where $X^{I,\overline{0}}$ is the corresponding K-L generalized bridge with end-point $\overline{0}$. We obtain

$$d\widetilde{S}_t^{I,\Gamma} = b\left(t, \widetilde{S}_t^{I,\Gamma}\right) dt + \theta\left(t, \widetilde{S}_t^{I,\Gamma}\right) d\widehat{X}_t^{I,\Gamma} + \theta\left(t, \widetilde{S}_t^{I,\Gamma}\right) dX_t^{I,\overline{0}}, \quad \widetilde{S}_0^{I,\Gamma} = x. \tag{4.46}$$

We define the normal quantization $\widehat{\widehat{S}}^{I,\Gamma}$ of this SDE by

$$d\widehat{\widehat{S}}_t^{I,\Gamma} = b\left(t, \widehat{S}_t^{I,\Gamma}\right) dt + \theta\left(t, \widehat{S}_t^{I,\Gamma}\right) d\widehat{X}_t^{I,\Gamma} + \theta\left(t, \widehat{S}_t^{I,\Gamma}\right) dX_t^{I,\overline{0}}, \quad \widehat{\widehat{S}}_0^{I,\Gamma} = x.$$

This is the same SDE as (4.46), where we have replaced $\widetilde{S}_t^{I,\Gamma}$ by $\widehat{S}_t^{I,\Gamma}$ in the drift and the volatility.

Remark. *With the same notations, $\widehat{S}^{I,\Gamma}$ is simply defined by*

$$\widehat{S}_t^{I,\Gamma} = x + \int_0^t b\left(s, \widehat{S}_s^{I,\Gamma}\right) ds + \int_0^t \theta\left(s, \widehat{S}_s^{I,\Gamma}\right) d\widehat{X}_s^{I,\Gamma} + \int_0^t \theta\left(s, \widehat{S}_s^{I,\Gamma}\right) dX_s^{I,\overline{0}}.$$

The principal interest of the normal quantization of stochastic differential equations is that it yields Gaussian processes, so that it is likely that one can obtain closed-form expressions when using it as a cubature scheme.

Theorem 4.4.8 (Quadratic error of the normal quantization scheme for SDEs)**.** *We keep the notations and assumptions of Definition 4.4.4. Then, for every $t \in [0,T)$ we have*

$$\mathbb{E}\left[\sup_{u\in[0,t]} \left|\widetilde{S}_u^{I,\Gamma} - \widehat{S}_u^{I,\Gamma}\right|^2 \middle| \widehat{Y}^\Gamma = \gamma_k\right] = O\left(\left\|\widetilde{S}^{I,\Gamma} - \widehat{S}^{I,\Gamma}\right\|_2^2\right). \tag{4.47}$$

Proof: I and Γ being fixed, to simplify notations, we will denote $\widehat{X} := \widehat{X}^{I,\Gamma}$, $\widetilde{X} := \widetilde{X}^{I,\Gamma}$, $\widehat{S} := \widehat{S}^{I,\Gamma}$, $\widetilde{S} := \widetilde{S}^{I,\Gamma}$, and $\widehat{S} := \widehat{S}^{I,\Gamma}$. We have

$$\widetilde{S}_t - \widehat{S}_t = \int_0^t \left(b\left(s,\widetilde{S}_s\right) - b\left(s,\widehat{S}_s\right)\right) ds + \int_0^t \left(\theta\left(s,\widetilde{S}_s\right) - \theta\left(s,\widehat{S}_s\right)\right) d\widehat{X}_s + \int_0^t \left(\theta\left(s,\widetilde{S}_s\right) - \theta\left(s,\widehat{S}_s\right)\right) dX_s^{I,\overline{0}}.$$

In [5], the canonical decomposition of $X^{I,\overline{0}}$ is derived. It writes $X^{I,\overline{0}} = \underbrace{\left\langle X, L^{\overline{0}}\right\rangle}_{:=\widetilde{V}} + \underbrace{X - \left\langle X, L^{\overline{0}}\right\rangle}_{:=\widetilde{M}}$, where $d\widetilde{V}_s = G_s d\langle X\rangle_s$ and $(G_s)_{s\in[0,T)}$ is a centered continuous Gaussian process.

Hence we have, conditionally to $\widehat{Y}^\Gamma = \gamma_k$,

$$\begin{aligned}
\left|\widetilde{S}_t - \widehat{S}_t\right|^2 &\le 4\left|\int_0^t \left(b\left(s,\widetilde{S}_s\right) - b\left(s,\widehat{S}_s\right)\right) ds\right|^2 + 4\left|\int_0^t \left(\theta\left(s,\widetilde{S}_s\right) - \theta\left(s,\widehat{S}_s\right)\right) \alpha_k'(s) ds\right|^2 \\
&\quad + 4\left|\int_0^t \left(\theta\left(s,\widetilde{S}_s\right) - \theta\left(s,\widehat{S}_s\right)\right) G_s d\langle X\rangle_s\right|^2 + 4\left|\int_0^t \left(\theta\left(s,\widetilde{S}_s\right) - \theta\left(s,\widehat{S}_s\right)\right) d\widetilde{M}_s\right|^2 \\
&\le 4T\left([b]_{\mathrm{Lip}}^2 + [\theta]_{\mathrm{Lip}}^2 \max_{u\in[0,T]} (\alpha_k'(u))^2\right) \int_0^t \left|\widetilde{S}_s - \widehat{S}_s\right|^2 ds \\
&\quad + 4\langle X\rangle_T \int_0^t \left(\theta\left(s,\widetilde{S}_s\right) - \theta\left(s,\widehat{S}_s\right)\right)^2 G_s^2 d\langle X\rangle_s + 4\left|\int_0^t \left(\theta\left(s,\widetilde{S}_s\right) - \theta\left(s,\widehat{S}_s\right)\right) d\widetilde{M}_s\right|^2,
\end{aligned}$$

where α_k stands for the deterministic path of $\widehat{X}$ conditionally to $\widehat{Y}^\Gamma = \gamma_k$. This yields

$$\begin{aligned}
\mathbb{E}\left[\left|\widetilde{S}_t - \widehat{S}_t\right|^2 \middle| \widehat{Y}^\Gamma = \gamma_k\right] &\le 4T\left([b]_{\mathrm{Lip}}^2 + [\theta]_{\mathrm{Lip}}^2 \max_{u\in[0,T]} (\alpha_k'(u))^2\right) \int_0^t \mathbb{E}\left[\left|\widetilde{S}_s - \widehat{S}_s\right|^2 \middle| \widehat{Y}^\Gamma = \gamma_k\right] ds \\
&\quad + 4\langle X\rangle_T \int_0^t \mathbb{E}\left[\left(\theta\left(s,\widetilde{S}_s\right) - \theta\left(s,\widehat{S}_s\right)\right)^2 G_s^2 \middle| \widehat{Y}^\Gamma = \gamma_k\right] d\langle X\rangle_s \\
&\quad + 4\mathbb{E}\left[\left|\int_0^t \left(\theta\left(s,\widetilde{S}_s\right) - \theta\left(s,\widehat{S}_s\right)\right) d\widetilde{M}_s\right|^2 \middle| \widehat{Y}^\Gamma = \gamma_k\right].
\end{aligned} \tag{4.48}$$

We now use Doob's inequality (Proposition 4.4.5) to the last term of Equation (4.48). Using that $d\left\langle \widetilde{M}\right\rangle = d\left\langle X^{I,\overline{0}}\right\rangle = d\langle X\rangle$, we get

$$\mathbb{E}\left[\sup_{u\in[0,t]} \left|\int_0^u \left(\theta\left(s,\widetilde{S}_s\right) - \theta\left(s,\widehat{S}_s\right)\right) d\widetilde{M}_s\right|^2 \middle| \widehat{Y}^\Gamma = \gamma_k\right] \le 4\int_0^t \mathbb{E}\left[\left(\theta\left(s,\widetilde{S}_s\right) - \theta\left(s,\widehat{S}_s\right)\right)^2 \middle| \widehat{Y}^\Gamma = \gamma_k\right] d\langle X\rangle_s.$$

Thus

$$4\mathbb{E}\left[\sup_{u\in[0,t]} \left|\int_0^u \left(\theta\left(s,\widetilde{S}_s\right) - \theta\left(s,\widehat{S}_s\right)\right) d\widetilde{M}_s\right|^2 \middle| \widehat{Y}^\Gamma = \gamma_k\right] \le 16[\theta]_{\mathrm{Lip}}^2 \int_0^t \mathbb{E}\left[\left(\widetilde{S}_s - \widehat{S}_s\right)^2 \middle| \widehat{Y}^\Gamma = \gamma_k\right] d\langle X\rangle_s.$$

Inserting this into (4.48) yields

$$\mathbb{E}\left[\sup_{u\in[0,t]}\left|\widetilde{S}_u-\widehat{S}_u\right|^2\middle|\widehat{Y}^\Gamma=\gamma_k\right]\leq 4T\left([b]^2_{\mathrm{Lip}}+[\theta]^2_{\mathrm{Lip}}\max_{u\in[0,T]}(\alpha_k'(u))^2\right)\int_0^t\mathbb{E}\left[\left|\widetilde{S}_s-\widehat{S}_s\right|^2\middle|\widehat{Y}^\Gamma=\gamma_k\right]ds$$
$$+16[\theta]^2_{\mathrm{Lip}}\int_0^t\mathbb{E}\left[\sup_{u\in[0,s]}\left|\widetilde{S}_u-\widehat{S}_u\right|^2\middle|\widehat{Y}^\Gamma=\gamma_k\right]d\langle X\rangle_s$$
$$+4\langle X\rangle_T\int_0^t\mathbb{E}\left[\left(\theta\left(s,\widetilde{S}_s\right)-\theta\left(s,\widehat{S}_s\right)\right)^2G_s^2\middle|\widehat{Y}^\Gamma=\gamma_k\right]d\langle X\rangle_s. \quad (4.49)$$

We now focus on the last term in (4.49). For $M>1$, we decompose the expectation on $\{|G_s|>M\}$ and $\{|G_s|\leq M\}$. We obtain, using that G_s is independent of $\widehat{Y}^\Gamma$,

$$\mathbb{E}\left[\left(\theta\left(s,\widetilde{S}_s\right)-\theta\left(s,\widehat{S}_s\right)\right)^2G_s^2\middle|\widehat{Y}^\Gamma=\gamma_k\right]\leq M^2[\theta]^2_{\mathrm{Lip}}\mathbb{E}\left[\left(\widetilde{S}_s-\widehat{S}_s\right)^2\middle|\widehat{Y}^\Gamma=\gamma_k\right]+4[\theta]^2_{\max}\mathbb{E}\left[G_s^2\mathbf{1}_{|G_s|>M}\right]$$
$$\leq M^2[\theta]^2_{\mathrm{Lip}}\mathbb{E}\left[\left(\widetilde{S}_s-\widehat{S}_s\right)^2\middle|\widehat{Y}^\Gamma=\gamma_k\right]+4[\theta]^2_{\max}M\exp\left(-\tfrac{M^2}{2v_s^2}\right)\left(\tfrac{2v_s}{\sqrt{2\pi}}+v_s^2\right),$$

where $v_s^2=\mathrm{Var}(G_s)$. If we define, for $s\in[0,T)$, $v_s^{*2}:=\sup_{u\in[0,s]}\mathrm{Var}(G_u)$ we obtain

$$\mathbb{E}\left[\sup_{u\in[0,t]}\left|\widetilde{S}_u-\widehat{S}_u\right|^2\middle|\widehat{Y}^\Gamma=\gamma_k\right]\leq 4T\left([b]^2_{\mathrm{Lip}}+[\theta]^2_{\mathrm{Lip}}\max_{u\in[0,T]}(\alpha_k'(u))^2\right)\int_0^t\mathbb{E}\left[\left|\widetilde{S}_s-\widehat{S}_s\right|^2\middle|\widehat{Y}^\Gamma=\gamma_k\right]ds$$
$$+16[\theta]^2_{\mathrm{Lip}}\int_0^t\mathbb{E}\left[\sup_{u\in[0,s]}\left|\widetilde{S}_u-\widehat{S}_u\right|^2\middle|\widehat{Y}^\Gamma=\gamma_k\right]d\langle X\rangle_s$$
$$+4\langle X\rangle_TM^2[\theta]^2_{\mathrm{Lip}}\int_0^t\mathbb{E}\left[\sup_{u\in[0,s]}\left|\widetilde{S}_u-\widehat{S}_u\right|^2\middle|\widehat{Y}^\Gamma=\gamma_k\right]d\langle X\rangle_s$$
$$+\underbrace{16\langle X\rangle_T^2[\theta]^2_{\max}\left(\frac{2v_s^*}{\sqrt{2\pi}}+v_s^{*2}\right)M\exp\left(-\frac{M^2}{2v_s^{*2}}\right)}_{:=\eta_M}. \quad (4.50)$$

We define the locally finite measure $\mu(dt)=d\langle X\rangle_t+dt$. We can ensure that

$$\eta_M\leq\eta:=\int_0^t\mathbb{E}\left[\left|\widetilde{S}_s-\widehat{S}_s\right|^2\middle|\widehat{Y}^\Gamma=\gamma_k\right]\mu(ds)$$

by setting $M=\sqrt{-\mathcal{W}\left(-\frac{\eta^2}{C_t^2}\right)}$, where $C_t:=16\langle X\rangle_T^2[\theta]^2_{\max}\left(\frac{2v_t^{*2}}{\sqrt{2\pi}}+v_t^{*2}\right)$ and where $\mathcal{W}$ is the Lambert $\mathcal{W}$ function. This finally yields

$$\mathbb{E}\left[\sup_{u\in[0,t]}\left|\widetilde{S}_u-\widehat{S}_u\right|^2\middle|\widehat{Y}^\Gamma=\gamma_k\right]\leq\Big(4T\left([b]^2_{\mathrm{Lip}}+[\theta]^2_{\mathrm{Lip}}\max_{u\in[0,T]}(\alpha_k'(u))^2\right)$$
$$-4\langle X\rangle_T^2[\theta]^2_{\mathrm{Lip}}\mathcal{W}\left(-\frac{\eta^2}{C_t^2}\right)+1\Big)\eta$$

We can conclude by observing that $\mathcal{W}(u)\underset{u\to0}{\to}0$. □

We can prove the analogous result of Theorem 4.4.8 for the general L^p framework, following the same steps except that Doob's inequality for continuous (local) martingales is replaced by the Burkholder-Davis-Gundy inequality (which holds for every $p>0$).

Remark (Extension to semimartingales)**.** *In Theorem 4.4.8, we limited ourselves to the case where X is a local martingale. The proofs are easily extended to the case of a semimartingale X*

as soon as there exists a locally finite measure ν on $[0,T]$ such that for every $\omega \in \Omega$ the finite-variation part $dV(\omega)$ in the canonical decomposition of X is absolutely continuous with respect to ν. For instance, this is the case for the Brownian bridge and Ornstein-Uhlenbeck processes whose finite-variation parts are absolutely continuous with respect to the Lebesgue measure on $[0,T]$.

Definition 4.4.5 (The lognormal quantization of stochastic differential equations)**.** *From Definition 4.4.4 of the normal quantization of stochastic differential equations, we can naturally define the lognormal quantization $\overline{S}^{I,\Gamma}$ of an SDE of the form*

$$\begin{cases} dS_t = S_t\mu(t,S_t)dt + S_t\theta(t,S_t)dX_t \\ S_0 = x, \end{cases}$$

as the exponential of the normal quantization $\widehat{(\log(S))}^{I,\Gamma}$ of the logarithm of the solution

$$d\log(S_t) = \mu(t,S_t)dt - \frac{1}{2}\theta(t,S_t)d\langle X\rangle_t + \theta(t,S_t)dX_t.$$

If the measure $d\langle X\rangle$ is absolutely continuous with respect to the Lebesgue measure λ on $[0,T]$ and $\tau_X := \frac{d\langle X\rangle}{d\lambda}$ and if we define $b(t,x) := \mu(t,x)dt - \frac{1}{2}\theta(t,x)g_X(t)$, as soon as so-defined functions b and θ satisfy the hypothesis of Definition 4.45, we have

$$\mathbb{E}\left[\sup_{u\in[0,t]}\left|\overline{S}_u - \widetilde{S}_u\right|^2 \middle| \widehat{Y}^\Gamma = \gamma_k\right] = O\left(\left\|\widetilde{S}^{I,\Gamma} - \widehat{S}^{I,\Gamma}\right\|_2^2\right). \tag{4.51}$$

Cubature method based on lognormal quantization for local volatility models

Consider the following stochastic differential equation

$$\begin{cases} dS_t = S_t b(t,S_t)dt + S_t\theta(t,S_t)dX_t, \\ S_0 = s_0, \end{cases}$$

where X is a continuous centered Gaussian semimartingale starting from 0, $b(t,x)$ and $\theta(t,x)$ are Borel functions, Lipschitz continuous with respect to x uniformly in t, θ is bounded and $|b(\cdot,0)|$ is bounded. We also assume that $d\langle X\rangle$ is absolutely continuous with respect to the Lebesgue measure λ on $[0,T]$ and $\tau_X := \frac{d\langle X\rangle}{d\lambda}$ is continuous. Moreover, we assume that the function $\beta(t,y)$ defined by Equation 4.35 is Lipschitz continuous with respect to y uniformly in t.

Then we can consider the lognormal quantization $\overline{S}^{I,\Gamma}$ of S associated with the finite set $I \subset \mathbb{N}^*$ and $\Gamma = \{\gamma_1, \cdots, \gamma_N\}$ a stationary codebook of $(Y_i)_{i\in I}$. Then we have

$$\begin{aligned}\overline{S}_T^{I,\Gamma} = S_0 \exp\Bigg(&\int_0^T b\left(t,\widehat{S}_t^{I,\Gamma}\right)dt - \int_0^T \frac{\widehat{S}_t^{I,\Gamma}\theta'_x\left(t,\widehat{S}_t^{I,\Gamma}\right)\theta\left(t,\widehat{S}_t^{I,\Gamma}\right)}{2}d\langle X\rangle_t \\ &- \int_0^T \frac{\theta\left(t,\widehat{S}_t^{I,\Gamma}\right)^2}{2}d\langle X\rangle_t + \int_0^T \theta\left(t,\widehat{S}_t^{I,\Gamma}\right)d\widehat{X}_t^{I,\Gamma} + \int_0^T \theta\left(t,\widehat{S}_t^{I,\Gamma}\right)dX_t^{I,\overline{0}}\Bigg),\end{aligned} \tag{4.52}$$

which, conditionally to $\widehat{Y}^\Gamma = \gamma_k$ is lognormal.

The first proposed cubature formula is, if one is interested by the quantity $\mathbb{E}[(S_T-K)_+]$ (or $\mathbb{E}[(K-S_T)_+]$), to make the following approximation

$$\mathbb{E}[(S_T-K)_+] \approx \mathbb{E}\left[\left(\overline{S}_T - K\right)_+\right] = \sum_{k=1}^N \mathbb{P}\left[\widehat{Y}^\Gamma = \gamma_k\right]\mathbb{E}\left[\left(\overline{S}_T - K\right)_+ \middle| \widehat{Y}^\Gamma = \gamma_k\right],$$

where the right-hand term of the expression can be derived as a weighted sum of closed-form expressions.

We recall that if $G \overset{\mathcal{L}}{\sim} \mathcal{N}(m, \Sigma^2)$ and $L = \exp(G)$, we have

$$\mathbb{E}[(L-K)_+] = \exp\left(m + \frac{\Sigma^2}{2}\right)\mathcal{N}(d_1) - K\mathcal{N}(d_2),$$

and

$$\mathbb{E}[(K-L)_+] = K\mathcal{N}(-d_2) - \exp\left(m + \frac{\Sigma^2}{2}\right)\mathcal{N}(-d_1),$$

where $d_2 = \frac{m-\ln(K)}{\Sigma}$ and $d_1 = d_2 + \Sigma$. In the setting of Formula (4.52), the conditional expectations m_k of $\log\left(\overline{S}_T^{I,\Gamma}\right)$ conditionally to $\widehat{Y}^\Gamma = \gamma_k$ are easily computed. As the conditional variances Σ_k are concerned, we devote the Appendix 4.A to the computation of $\mathrm{Var}\left(\int_0^T g(t)dX_t^{I,\overline{0}}\right)$ where $X^{I,\overline{0}}$ is the associated K-L generalized bridge and $g \in L^2([0,T])$.

Proposition 4.4.9 (Lognormal quantization cubature error)**.** *With the same set of notations and hypothesis, we have*

$$\mathbb{E}[(S_T-K)_+] - \mathbb{E}\left[\left(\overline{S}_T - K\right)_+\right] = O\left(\left\|X - \widehat{X}^{I,\Gamma}\right\|_{2+\varepsilon}\right),$$

for any $\varepsilon > 0$.

Proof: This is a straightforward consequence of Equations (4.51), (4.41), (4.42) and Theorem 4.4.4. □

Remark (Richardson-Romberg extrapolation for the lognormal cubature method)**.** *A conjecture, supported by our numerical experiments is that in the case where $\widehat{X}^{I,\Gamma}$ is an optimal K-L product quantizer of X, we have*

$$\mathbb{E}[(S_T-K)_+] - \mathbb{E}\left[\left(\overline{S}_T - K\right)_+\right] = C\left\|X - \widehat{X}^{I,\Gamma}\right\|_2^2 + o\left(\left\|X - \widehat{X}^{I,\Gamma}\right\|_2^2\right),$$

for some positive constant C. In other words, this means that as for crude quantization of S, stationarity results in a convergence rate of a higher order by one for the quantization-based cubature scheme. We conjecture that this rate is in fact a sharp rate of convergence. This highly suggests to implement Richardson-Romberg extrapolations for the lognormal quantization-based cubature formula, similar to the one we use with functional quantization (Equation (4.20)).

4.4.5 Lognormal quantization for local stochastic volatility models

The lognormal quantization-based cubature for SDE with a local drift and a local volatility, derived in the previous section provides an accurate vanilla option pricing method for this kind of models. Consequently, we have a method at our disposal to evaluate the function ϕ defined in Equation (4.30). Hence, we have found the missing link on our path to vanilla option pricing in local stochastic volatility models. In this section, we benchmark our method on the case of the SABR model with $\beta \neq 1$.

The SABR model with $\beta \neq 1$

In this section, we present numerical results when using the normal quantization method with the classical SABR model proposed in [14]. In this model, we assume that, under the risk-neutral measure, the forward price of a risky asset is the solution of the SDE

$$dF_t = F_t^\beta \sigma_t dW_t, \quad F_0 > 0, \tag{4.53}$$

where the volatility σ is the solution of the SDE

$$d\sigma_t = b(t,\sigma_t)dt + \theta(t,\sigma_t)dW_t^\sigma, \quad \sigma_0 > 0, \tag{4.54}$$

with $\beta \in [0,1]$ and $d\langle W, W^\sigma\rangle_t = \rho dt$. (The model presented in Section 4.3 corresponded to the special case where $\beta = 1$).
In Figure 4.6 we display the implied volatility smile estimated by the (extrapolated) functional quantization-based cubature formula, and the value computed by the closed-form short-maturity asymptotics.

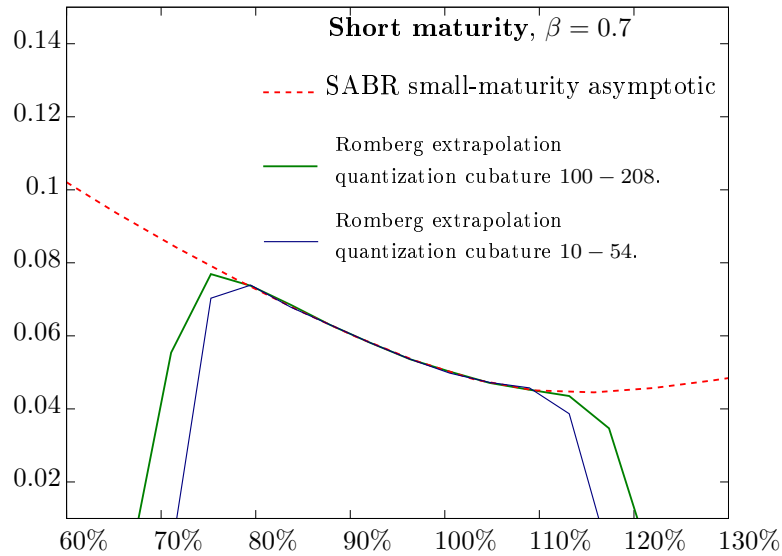

Figure 4.6: Implied volatility smile in the SABR model with $\beta = 0.7$, $\gamma = 0.3$, $\sigma_0 = 0.2$, $T = 1$ and $\rho = -0.5$.
• The dotted line curve corresponds to the SABR small-maturity asymptotic.
• The thin (blue) continuous curve corresponds to a Romberg extrapolation of the functional quantization-based cubature method. The Romberg extrapolation couples are $10 - 54$ for the lognormal quantization and $54 - 208$ for the volatility quantization.
• The thicker (green) continuous curve corresponds to a Romberg extrapolation of the functional quantization-based cubature method. The Romberg extrapolation couples are $100 - 208$ for the lognormal quantization and $54 - 208$ for the volatility quantization.

As we can see in Figure 4.6, the lognormal quantization scheme does not provide correct results for far-out-of-the-money options, for which the PDE approach is preferable. A remark is that the vanilla option payoff is not a regular functional in $L^2([0,T])$ (because it only depends on the marginal at $t = T$). We might experience better results with more regular functionals as the payoff of an Asian option. The range of strike for which we get accurate values of the implied volatility increases with the size of the lognormal quantizers that we use.

4.A Computing $\mathrm{Var}\left(\int_0^T g(t)dX_t^{I,\overline{0}}\right)$ where $X^{I,\overline{0}}$ is a K-L generalized bridge and $g \in L^2([0,T])$

Considering the decomposition

$$X_t^{I,\overline{0}} = X_t - \sum_{i\in I}\left(\int_0^T X_s e_i^X(s)ds\right) e_i^X(t),$$

We have

$$\int_0^T g(s)dX_s^{I,\overline{0}} = \int_0^T g(s)dX_s - \sum_{i\in I}\left(\int_0^T X_s e_i^X(s)ds\right)\left(\int_0^T g(s)de_i^X(s)\right).$$

Thus, as $\left(\int_0^T X_s e_i^X(s)ds\right)_{i\in I} \overset{\mathcal{L}}{\sim} \mathcal{N}\left(0, \mathrm{diag}\left(\lambda_i^X\right)_{i\in I}\right)$, we have

$$\mathrm{Var}\left(\int_0^T g(s)dX_s^{I,\overline{0}}\right) = \mathrm{Var}\left(\int_0^T g(s)dX_s\right) + \sum_{i\in I}\lambda_i^X\left(\int_0^T g(s)de_i^X(s)\right)^2 - 2\sum_{i\in I}\left(\int_0^T g(s)de_i^X(s)\right)\mathbb{E}\left[\int_0^T g(s)dX_s\int_0^T X_s e_i^X(s)ds\right].$$

Now, identifying $\int_0^T X_s e_i^X(s)ds = \int_0^T f_i^X(s)dX_s$ where f_i^X is defined as in Definition 4.4.3, we have

$$\mathbb{E}\left[\int_0^T g(s)dX_s\int_0^T X_s e_i^X(s)ds\right] = \mathbb{E}\left[\int_0^T g(s)dX_s\int_0^T f_i^X(s)dX_s\right].$$

Hence we need to derive the terms $\mathrm{Var}\left(\int_0^T g(s)dX_s\right)$ and $\mathbb{E}\left[\int_0^T g(s)dX_s\int_0^T f_i^X(s)dX_s\right]$.

4.A.1 The case of the standard Brownian motion

In the case where $X = W$ is a standard Brownian motion, we have $\mathrm{Var}\left(\int_0^T g(s)dW_s\right) = \int_0^T g(s)^2ds$ and $\mathbb{E}\left[\int_0^T g(s)dW_s\int_0^T f_i^W(s)dW_s\right] = \int_0^T g(s)f_i^W(s)ds$.

The result can be simplified using the closed-form expression of the Karhunen-Loève eigenfunctions of the standard Brownian motion which is given in Section 4.1.1. We easily obtain $f_i^W(t) = \lambda_i^W\left(e_i^W\right)'(t)$. We finally obtain

$$\begin{aligned}\mathrm{Var}\left(\int_0^T g(s)dW_s^{I,\overline{0}}\right) &= \int_0^T g^2(s)ds - \sum_{i\in I}\lambda_i^W\left(\int_0^T g(s)de_i^W(s)\right)^2\\ &= \int_0^T g^2(s)ds - \sum_{i\in I}\frac{1}{\lambda_i^W}\left(\int_0^T g(s)f_i^W(s)ds\right)^2.\end{aligned} \tag{4.55}$$

4.A.2 The case of the standard Brownian bridge

In the case where $X = B$ is a standard Brownian bridge, we write $B_t = W_t - t\frac{W_T}{T}$. This yields $\int_0^T g(s)dB_s = \int_0^T g(s)dW_s - \frac{W_T}{T}\int_0^T g(s)ds$. Thus we get

$$\mathrm{Var}\left(\int_0^T g(s)dB_s\right) = \int_0^T g(s)^2ds - \frac{1}{T}\left(\int_0^T g(s)ds\right)^2,$$

and

$$\mathbb{E}\left[\int_0^T g(s)dB_s\int_0^T f_i^B(s)dB_s\right] = \int_0^T g(s)f_i^B(s)ds - \frac{1}{T}\int_0^T g(s)ds\int_0^T f_i^B(s)ds.$$

The result can be simplified using the closed-form expression of the Karhunen-Loève eigenfunctions of the standard Brownian bridge which is given in Section 4.1.1. We easily obtain

$$f_i^B(t) = \lambda_i^B\left(e_i^B\right)' - \left(\frac{T}{\pi n}\right)\sqrt{\frac{2}{T}}\cos(\pi n),$$

and thus, using this expression, we get

$$\begin{aligned}\mathrm{Var}\left(\int_0^T g(s)dB_s^{I,\overline{0}}\right) &= \int_0^T g^2(s)ds - \frac{1}{T}\left(\int_0^T g(s)ds\right)^2 + \sum_{i\in I}\lambda_i^B\left(\int_0^T g(s)de_i^B(s)\right)^2\\ &\quad -2\sum_{i\in I}\left(\int_0^T g(s)de_i^B(s)\right)\left(\int_0^T g(s)f_i^B(s)ds - \frac{1}{T}\int_0^T g(s)ds\int_0^T f_i^B(s)ds\right)\\ &= \int_0^T g^2(s)ds - \frac{1}{T}\left(\int_0^T g(s)ds\right)^2 - \sum_{i\in I}\lambda_i^B\left(\int_0^T g(s)de_i^B(s)\right)^2\\ &\quad -2\sum_{i\in I}\left(\int_0^T g(s)de_i^B(s)\right)\left(\int_0^T g(s)ds\right)\underbrace{\left(-\left(\frac{T}{\pi n}\right)\sqrt{\frac{2}{T}}\cos(\pi n) - \frac{1}{T}\int_0^T f_i^B(s)ds\right)}_{=0},\end{aligned} \tag{4.56}$$

so that

$$\mathrm{Var}\left(\int_0^T g(s)dB_s^{I,\overline{0}}\right) = \int_0^T g^2(s)ds - \frac{1}{T}\left(\int_0^T g(s)ds\right)^2 - \sum_{i\in I}\lambda_i^B\left(\int_0^T g(s)de_i^B(s)\right)^2. \quad (4.57)$$

Bibliography

[1] Leif B. Andersen and Rupert Brotherton-Ratcliffe. Extended LIBOR market models with stochastic volatility. *Journal of Computational Finance*, 9(1):1–40, 2005.

[2] Kendall E. Atkinson. *The numerical solution of integral equations of the second kind.* Cambridge Monographs on Applied and Computational Mathematics, 1999.

[3] Lorenzo Bergomi. Smile dynamics III. *Risk*, 2007.

[4] Sylvain Corlay. The Nyström method for functional quantization with an application to the fractional Brownian motion. *Preprint*, 2010.

[5] Sylvain Corlay. Partial functional quantization and generalized bridges. *Preprint*, 2011.

[6] Sylvain Corlay and Gilles Pagès. Functional quantization-based stratified sampling methods. *Preprint*, 2010.

[7] Paul Deheuvels and Guennadi V. Martynov. A Karhunen-Loève decomposition of a Gaussian process generated by independent pairs of exponential random variables. *Journal of Functional Analysis*, 255(9):2363–2394, 2008.

[8] Halim Doss. Liens entre équations différentielles stochastiques et ordinaires. *Ann. I.H.P.*, 13(2):99–125, 1977.

[9] Halim Doss. Connections between stochastic and ordinary integral equations. In Willi Jäger, Hermann Rost, and Petre Tautu, editors, *Biological Growth and Spread*, volume 38, pages 443–448. Springer, Berlin, 1979.

[10] Jim Gatheral, Elton P. Hsu, Peter Laurence, Cheng Ouyang, and Tai-Ho Wang. Asymptotics of implied volatility in local volatility models. *Preprint*, 2009.

[11] John Arthur Gaunt and Lewis Fry Richardson. The deferred approach to the limit. *Philosophical Transactions of the Royal Society of London*, A 226:299–349, 1927.

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

[13] Siegried Graf, Harald Luschgy, and Gilles Pagès. Distortion mismatch in the quantization of probability measures. *ESAIM: PS*, 12:127–153, 2008.

[14] Patrick S. Hagan, Deep Kumar, Andrew S. Lesniewski, and Diana E. Woodward. Managing smile risk. *Wilmott magazine*, 2002.

[15] Ernst Hairer, Syvert Paul Nørsett, and Gerhard Wanner. *Solving ordinary differential equations I: nonstiff problems.* Springer-Verlag New York, Inc., 1996.

[16] Pierre Henry-Labordère. A general asymptotic implied volatility for stochastic volatility models. *Preprint*, 2005.

[17] Pierre Henry-Labordère. *Analysis, geometry and modelling in finance.* Chapman & Hall / CRC, Financial Mathematics Series, 2008.

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

[19] Shigeo Kusuoka and Daniel W. Stroock. Applications of the Malliavin calculus, part II. *J. Fac. Sci. Univ. Tokyo 32*, pages 1–76, 1985.

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

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

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

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

[24] Harald Luschgy, Gilles Pagès, and Benedikt Wilbertz. Asymptotically optimal quantization schemes for Gaussian processes. *ESAIM: PS*, 14:93–116, 2010.

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

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

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

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

[29] Gilles Pagès and Afef Sellami. Convergence of multi-dimensional quantized SDE's. In Catherine Donati-Martin, Antoine Lejay, and Alain Rouault, editors, *Séminaire de Probabilités XLIII*, pages 269–308. Springer, Berlin, 2010.

[30] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion.* Springer, 3rd edition, 2005.

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

[32] Hector J. Sussman. On the gap between deterministic and stochastic ordinary differential equations. *Ann. Probab*, 6(1):19–41, 1978.

[33] Benedikt Wilbertz. Computational aspects of functional quantization for Gaussian measures and applications. *Diploma thesis, Univ. Trier (Germany)*, 2005.

[34] Benedikt Wilbertz. *Construction of optimal quantizers for Gaussian measures on Banach spaces.* PhD thesis, Universität Trier, 2008.

[35] Eugene Wong and Moshe Zakai. On the relation between ordinary and stochastic differential equations. *International Journal of Engineering Science*, 3(2):213–229, 1965.

Chapter 5

A fast nearest neighbor search algorithm based on vector quantization

Abstract

In this chapter, we propose a new fast nearest neighbor search algorithm, based on vector quantization. Like many other branch-and-bound search algorithms [1, 10], a preprocessing recursively partitions the data set into disjoint subsets until the number of points in each part is small enough. In doing so, a search-tree data structure is built. This preliminary recursive partition of the data set is based on the vector quantization of the empirical distribution of the initial data set.

Unlike previously cited methods, this kind of partitions does not a priori allow eliminating several brother nodes in the search tree with a single test. To overcome this difficulty, we propose an algorithm to reduce the number of tested brother nodes to a minimal list that we call "friend Voronoi cells". The complete description of the method requires a deeper insight into the properties of Delaunay triangulations and Voronoi diagrams.

Keywords: vector quantization, fast nearest neighbor search, Voronoi diagram, Delaunay triangulation, principal component analysis.

Introduction

The problem of nearest neighbor search, also known as the post-office problem [7] has been widely investigated in the area of computational geometry. It is encountered for many applications, such as pattern recognition and vector quantization.

The post-office problem has been solved near optimally for the case of low dimensions. Algorithms differ in their practical efficiency on real data sets. For large dimensions, most solutions have a complexity that grows exponentially with the dimension, or require a longer query time than the obvious brute force algorithm. In fact, it has been noticed that, if n is the size of the data set and d is the dimensionality, the best choice becomes linear search when $d > K \log(n)$ for some positive constant K which depends on the chosen algorithm. This effect is known as the curse of dimensionality.

Concerning the application to (Voronoi) vector quantization, nearest neighbor projections are recognized to represent the critical part of most codebook optimization algorithms. In this case, the large amount of nearest neighbor searches we have to do shows that a preprocessing of the data set will be profitable if it reduces the average query time. Still, in some cases, the codebook is chosen so that nearest neighbor search is performed easily, (as when dealing with product quantization). Moreover, non-Voronoi quantization methods can also be designed in order to simplify the projection procedure while preserving some important properties of optimal quantizers, such as the stationarity in the quadratic case.

Let us also point out a field recently emerging under the name of dual quantization [11, 12]. In this context, the nearest neighbor search, *i.e.* the location of a point in a Voronoi partition, is replaced by the analogous procedure in the Delaunay triangulation. This localization procedure in Delaunay triangulations has been widely investigated in the practical viewpoint in terms of reduction of its computational complexity. We refer to Devillers, Pion and Teillaud for a review on this subject [2].

Many nearest neighbor search algorithms rely on a recursive partitioning of the data set resulting in a search-tree data structure [1, 10]. The method proposed by McNames in [10] improved the classical Kd-tree algorithm [1] by taking advantage of the shape of the data set thanks to principal component analysis. The "principal axis tree" algorithm performs much faster than the classical Kd-tree when the coordinates of the data set are correlated and it seems to handle better the growth of dimensionality.

In our case, the proposed algorithm uses vector quantization as a clustering method to perform this recursive partitioning and to take advantage of the geometry of the data set. It is classical background that when dealing with empirical distributions, the quadratic vector quantization problem is equivalent to the reduction of the intraclass inertia of the related partition, and the specification of the classical Lloyd algorithm in this case turns out to be the k-means clustering algorithm.

We will see that one drawback of this kind of partition is that, like other tree-based search algorithms, after determining the closest neighbor of a query in a leaf-node of the tree, the procedure has to move up to the parent node and determine whether brother nodes have to be explored or not. Unlike Kd-tree and "principal axis tree", our so-called "quantization tree" can't eliminate several brother nodes by a single test. This is the motivation for the development of our friend node algorithm.

The chapter is organized as follows. Section 5.1 is devoted to classical definitions and notations related to vector quantization. The link to the classification problem is pointed out. Section 5.2 recalls some definitions of computational geometry which will be useful in the sequel. As both the fields of vector quantization and algorithmic geometry deal with the notion of Voronoi diagram, we pay particular attention to distinguishing the corresponding definitions and notations. Section 5.3 makes a brief presentation of both the Kd-tree [1] and "principal axis tree" [10] algorithms. We deal with some optimizations that will be applicable to our quantization tree as well. Section 5.4 presents the "crude" quantization tree, *i.e.* without using any friend node algorithm. It is presented as the natural counterpart of these two branch-and-bound algorithms with a quantization-based partition of the data set. Section 5.5 presents the friend node algorithm which was discussed above.

Finally, the last section provides some performance comparisons between the different algorithms on various data sets.

5.1 Vector quantization and Voronoi tessellations

We consider $(\Omega, \mathcal{A}, \mathbb{P})$ a probability space and E a (real) finite-dimensional Euclidean space. The principle of a random variable X taking its values in E is to approach X by a random variable Y taking a finite number of values in E.

Definition 5.1.1 (quantizer)**.** *In this surrounding, the discrete random variable Y is a* quantizer *of X.*

If $X \in L^p$, the quantization error is the L^p norm of $|X - Y|$, where $|\cdot|$ denotes the Euclidean norm on E. The minimization of this error yields the following minimization problem

$$\min\{\|X - Y\|_p, Y : \Omega \to E \text{ measurable, } \operatorname{card}(Y(\Omega)) \leq N\}. \tag{5.1}$$

Definition 5.1.2 (Voronoi partition)**.** *Consider $N \in \mathbb{N}^*$, $\Gamma = \{\gamma_1, \cdots, \gamma_N\} \subset E$ and let $C = \{C_1, \cdots, C_N\}$ be a Borel partition of E. C is a Voronoi partition associated with Γ if $\forall i \in \{1, \cdots, N\}$, $C_i \subset \{\xi \in E, |\xi - \gamma_i| = \min\limits_{j \in \{1, \cdots, N\}} |\xi - \gamma_j|\}$.*

If $C = \{C_1, \cdots, C_N\}$ is a Voronoi partition associated with $\Gamma = \{\gamma_1, \cdots, \gamma_N\}$, it is clear that $\forall i \in \{1, \cdots, N\}, \gamma_i \in C_i$. C_i is called Voronoi slab associated with γ_i in C and γ_i is the center of the slab C_i.
We denote $C_i = \operatorname{slab}_C(\gamma_i)$. For every $a \in \Gamma$, $W(a|\Gamma)$ is the closed subset of E defined by $W(a|\Gamma) = \left\{y \in E,\ |y - a| = \min\limits_{\gamma \in \Gamma} |y - \gamma|\right\}$.

Definition 5.1.3 (Nearest neighbor projection)**.** *Consider $\Gamma \subset E$ a finite subset of E. A nearest neighbor projection onto Γ is an application $\operatorname{Proj}_\Gamma$ that satisfies*

$$\forall x \in E, \quad \left|x - \operatorname{Proj}_\Gamma(x)\right| = \min_{\gamma \in \Gamma} |x - \gamma|.$$

To be more precise, if $\operatorname{Proj}_\Gamma$ is a measurable nearest neighbor projection onto Γ, there exists a Voronoi partition $C = \{C_1, \cdots, C_N\}$ associated to Γ such that $\operatorname{Proj}_\Gamma = \sum\limits_{i=1}^{N} \gamma_i \mathbf{1}_{C_i}$.

Proposition 5.1.1. *Let X be an E-valued L^p random variable, and Y taking its values in the fixed point set $\Gamma = \{y_1, \cdots, y_N\} \subset E$ where $N \in \mathbb{N}$. Set $\widehat{X}^\Gamma$ the random variable defined by $\widehat{X}^\Gamma := \operatorname{Proj}_\Gamma(X)$ where $\operatorname{Proj}_\Gamma$ is a nearest neighbor projection on Γ, called a Voronoi Γ-quantizer of X.*
Then we clearly have $\left|X - \widehat{X}^\Gamma\right| \leq |X - Y|$ a.s.. Hence $\left\|X - \widehat{X}^\Gamma\right\|_p \leq \|X - Y\|_p$.

A consequence of this proposition is that solving the minimization problem (5.1) amounts to solving the simpler minimization problem

$$\min\left\{\|X - \operatorname{Proj}_\Gamma(X)\|_p,\ \Gamma \subset E, \operatorname{card}(\Gamma) \leq N\right\}. \tag{5.2}$$

The quantity $\|X - \operatorname{Proj}_\Gamma(X)\|_p$ is called the mean L^p-quantization error. When this minimum is reached, we refer to L^p-optimal quantization.
The problem of the existence of a minimum has been investigated for decades on its numerical and theoretical aspects in the finite-dimensional case [5]. For every $N \geq 1$, the L^p-quantization error is Lipschitz continuous and reaches a minimum. An N-tuple that achieves the minimum has pairwise distinct components, as soon as $\operatorname{card}(\operatorname{supp}(\mathbb{P}_X)) \geq N$. This result stands in the general case of a random variable valued in a reflexive separable Banach space [8]. If $\operatorname{card}(X(\Omega))$ is infinite,

this minimum strictly decreases to 0 as N goes to infinity. The asymptotic rate of convergence, in the case of non-singular distributions is ruled by the Zador theorem [5]. A non-asymptotic upper bound for the quantization error is also available [9].
We now focus on the quadratic case ($p = 2$). For a L^2 random variable X, we now denote by $\mathcal{C}_N(X)$ the set of L^2-optimal quantizers of X of level N and $e_N(X)$ the minimal quadratic distortion that can be achieved when approximating X by a quantizer of level N. A quantizer Y of X is stationary (or self-consistent) if $Y = \mathbb{E}[X|Y]$.

Proposition 5.1.2 (Stationarity of L^2-optimal quantizers)**.** *A (quadratic) optimal quantizer is stationary.*

The stationarity is a particularity of the quadratic case. In other L^p cases, a similar property involving the notion of p-center occurs. A proof is available in [6].

Definition 5.1.4 (Centroidal projection)**.** *Let $C = \{C_1, \cdots, C_N\}$ be a Borel partition of E. Let us define for $1 \leq i \leq N$, $G_i = \begin{cases} \mathbb{E}[X|X \in C_i] & \text{if } \mathbb{P}[X \in C_i] \neq 0, \\ 0 & \text{in the other case,} \end{cases}$ the centroids associated with X and C.*

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

Lemma 5.1.3 (Huyghens, variance decomposition)**.** *Let X be a E-valued L^2 random variable, $N \in \mathbb{N}^*$ and $C = (C_i)_{1 \leq i \leq N}$ a Borel partition of E. Consider* $\mathrm{Proj}_{C,X} = \sum_{i=1}^{N} G_i \mathbf{1}_{C_i}$ *the associated centroidal projection. Then one has,*

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

The variance of the probability distribution X decomposes itself as the sum of the ***intraclass inertia*** (1) *and the* ***interclass inertia*** (2).

Proof:

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

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

5.2 Backgrounds on theory of polytopes

Let E be a d-dimensional vector space and E^* its dual.

Definition 5.2.1 (k-flat)**.** *A k-flat is a k-dimensional affine subspace E.*

Definition 5.2.2 (convex polyhedron and convex polytope)**.** *A convex polyhedron is the intersection of a finite subset of closed halfspaces. If it is bounded, it is a convex polytope.*

Definition 5.2.3 (cell)**.** *A cell is the intersection of a finite set of flats and open halfspaces. And thus, equivalently, it is the relative interior of a convex polyhedron. If $R \subset E$, we denote by* $\mathrm{cell}(R)$ *the relative interior of the convex hull of R.*

Definition 5.2.4 (Leibniz halfspace). *For $(a,b) \in E^2$ let us denote by $H(a,b) := \left\{x \in \mathbb{R}^d \middle| |x-a| \leq |x-b|\right\}$ the Leibniz halfspace associated to (a,b).*

Definition 5.2.5 (simplex). *A simplex is* cell(R) *where R is a set of affinely independent points.*

- A 2-dimensional simplex is the interior of a triangle.
- A 3-dimensional simplex is the interior of a tetrahedron.

Definition 5.2.6 (circumsphere). *A circumsphere of a set $R \subset E$ is a sphere S of E such that $R \subset S$.*

Definition 5.2.7 (face). *A face of a convex polyhedron P is the relative interior of the intersection of a hyperplane supporting P with the closure of P.*

Proposition 5.2.1. *Let P be a convex polyhedron, a face of P is a cell, and a face of a face of P is a face of P.*

Definition 5.2.8 (k-face). *A k-face is a a face whose affine closure has dimension k.*

Definition 5.2.9 (cell complex). *A cell complex is a finite collection of pairwise disjoint cells so that the face of every cell is in the collection.*

Definition 5.2.10 (opposite k-faces). *Two distinct k-cells of a cell complex are opposite if they have a common $(k-1)$-face.*

Definition 5.2.11 (triangulation). *Let S be a finite point set of E. A triangulation T of S is a cell complex whose union is the convex hull of S and whose set of 0-cells is S.*

Definition 5.2.11 is a non-standard definition because cells are not required to be simplices. This formalism is due to Fortune [4].

Definition 5.2.12 (proper triangulation). *A proper triangulation is a triangulation all whose cells are simplices.*

Any triangulation can be completed to a proper triangulation by subdividing non-simplicial cells.

5.2.1 Voronoi diagrams and Delaunay triangulations

Voronoi diagram

Let E be a d-dimensional Euclidean space, and S a finite subset of E. In the following, elements of S will be called *sites*.

Definition 5.2.13 (Voronoi cell). *For a nonempty subset of S, $R \subset S$, the Voronoi cell of R, denoted $V(R)$ is the set of all points in E that are equidistant from all sites in R, and closer to every site of R than to any site not in R.*

Proposition 5.2.2.

- *Clearly, is $r \in S$, $V(\{r\})$ is the set of all points strictly closer to r than to any other site. In particular, it is the interior of the Voronoi slab associated to r in S. (See the definition of a Voronoi slab in Section 5.1.)*
- *$V(R)$ may be empty.*
- *Any point of E lies in $V(R)$ for some $R \subset S$.*

Definition 5.2.14 (Voronoi diagram). *The Voronoi diagram V is the collection of all nonempty Voronoi cells $V(R)$ for $R \subset S$.*

Delaunay triangulation

Definition 5.2.15 (Delaunay cell)**.** *If $R \subset S$, and $V(R)$ is a nonempty Voronoi cell, then the Delaunay cell $D(R)$ is* cell(R).

Definition 5.2.16 (Delaunay triangulation)**.** *The Delaunay triangulation D of S is the collection of Delaunay cells $D(R)$, where R varies over subsets of S with $V(R)$ nonempty.*

Proposition 5.2.3 (Empty circumsphere property)**.** *For $R \subset S$,* cell(R) *is a Delaunay cell if and only if there is a circumsphere of R that contains no site of $S \backslash R$ in its interior.*

Proof: This follows from the definition of the Voronoi cell $V(R)$, which is nonempty if and only if R admits an empty circumsphere. □

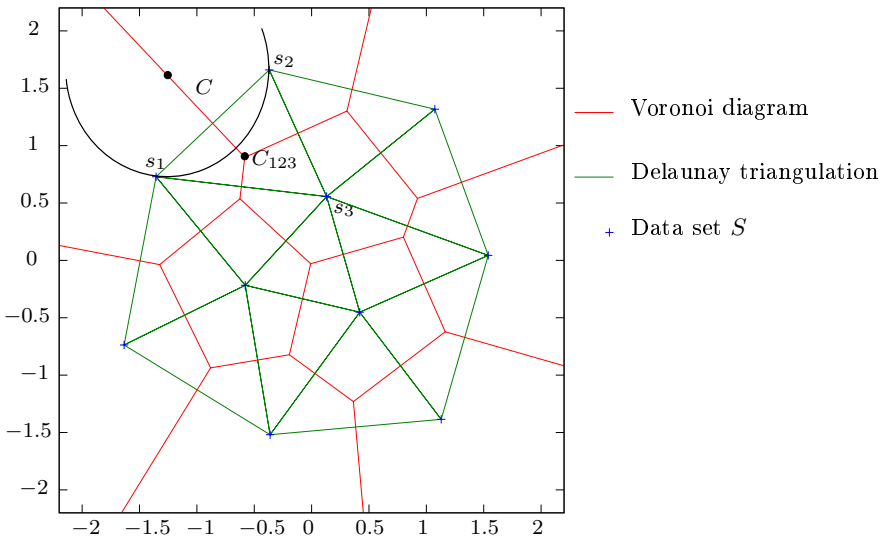

Figure 5.1: Voronoi diagram and Delaunay triangulation of a data set S of size 10. We have $C \in V_S(\{s_1, s_2\})$. So C is the center of an empty circumsphere of $\{s_1, s_2\}$. The point C_{123} is the center of the circumsphere of the Delaunay triangle $\{s_1, s_2, s_3\}$.

Theorem 5.2.4. *Let S be a set of n points in E with Voronoi diagram V and Delaunay triangulation D. Then*

1. *V is a cell complex that partitions E.*
2. *D is a triangulation of S.*
3. *V and D are linked by the following duality relation:*

   For $R, R' \subset S$, $V(R)$ is a face of $V(R')$ if and only if $D(R')$ is a face of $D(R)$.
4. *$V(R)$ is unbounded if and only if every site of R is on the boundary of the convex hull of S.*

We refer to [4] for a detailed proof.

Locality

Definition 5.2.17 (locally Delaunay)**.** *We consider two opposite d-cells* cell(R) *and* cell(R') *in a triangulation T with circumspheres C and C'.* cell(R) *and* cell(R') *are locally Delaunay if $R' \backslash R \cap C = \emptyset$. This is equivalent to $R \backslash R' \cap C' = \emptyset$.*
A triangulation is locally Delaunay if every pair of opposite d-cells is locally Delaunay.

Lemma 5.2.5 (Delaunay and locally Delaunay). *A triangulation is Delaunay if and only if it is locally Delaunay.*

We refer to [4] for a detailed proof.

Definition 5.2.18 (General position). *Let S be a nonempty finite set of sites in E. S is in general position if no $d+1$ points of S are affinely dependent and if no $d+2$ points of S lie on a common sphere.*

Definition 5.2.19 (Incircle list). *In the following, if S is a finite nonempty set of sites, D is a Delaunay triangulation of S and $x \in E$ is a fixed point, we call incircle list and denote by $ICL_D(x)$ the set of d-cells of D whose circumsphere contains x.*

If S is in general position, no Delaunay cell of S is degenerated. Every cell of the triangulation is a simplex and for any $R \subset S$, $V(R)$ has dimension $d+1-|R|$.

Computing the Delaunay triangulation and the Voronoi diagram

Whereas the Voronoi diagram was defined before the Delaunay triangulation, it has been recognized that it is easier to devise algorithms in terms of Delaunay triangulation, especially because of the locality property 5.2.5.

A common data structure for Delaunay triangulations is a graph structure where each simplex is a "node". The node contains the indices of the $d+1$ sites of the simplex and the pointers to the adjacent simplices. Null pointers are used when the simplices lie on the boundary of the triangulation. Cells of lower dimension are not directly represented in the graph structure. Another convenient convention is that the kth pointer stored in the node corresponds to the facet obtained by deleting the kth site in the node. Moreover the order is chosen so that the orientation of every simplex in the triangulation remains always positive.

Here, we present the principles of incremental algorithms for Delaunay triangulations. In this kind of algorithms, sites are added one by one, and the Delaunay triangulation is modified to include each new site. Many other algorithms have been designed for computing the Delaunay triangulation, especially in dimension 2. Moreover, computing the Delaunay triangulation of the Voronoi diagram in the one-dimensional case simply amounts to sorting the data set. An advantage of incremental algorithms is that they are valid in any dimension. Moreover, for another purpose in the following, we will need a new algorithm (the friend node algorithm presented in Section 5.5) that requires a stage which is very similar to the insertion of a new point in the Delaunay triangulation. Hence we will focus here on incremental algorithms.
Let $S = (s_1, \cdots, s_N)$ be a nonempty finite set of sites of E of cardinal N. We define the sets $S_k := (s_1, \cdots, s_k)$ for $k \in \{1, \cdots, N\}$. Now, for a fixed integer $i < N$, let us consider D_i the Delaunay triangulation of S_i. We inspect the situation of s_{i+1} with respect to the Delaunay triangulation D_i. From this analysis, the Delaunay triangulation will be modified locally to build a new Delaunay triangulation D_{i+1} of S_{i+1}. When all the sites of S will be processed, we will have the complete Delaunay triangulation D of S.

Three situations can occur, if S is in general position:

1. s_{i+1} lies in the interior convex hull of S_i.
2. s_{i+1} does not lie in any circumsphere of any simplex of D_i.
3. s_{i+1} lies outside of the convex hull of S_i but belongs to a circumsphere of a simplex of D_i.

(1) In the first situation, let us denote by $\mathcal{S} := ICL_{D_i}(s_{i+1})$ and $F_1, \cdots, F_p$ the external faces of $\mathcal{S}$ of any dimension $k < d$. We can show that the cell complex defined by

$$D_{i+1} := (D_i \backslash \mathcal{S}) \ \cup \Big\{ \text{cell}(F_j, s_{i+1})_j, \ 1 \leq j \leq p \Big\} \cup \Big\{ \{s_{i+1}\} \Big\}$$

is the Delaunay triangulation associated to S_{i+1}. In a more general setting, we have the following property:

Proposition 5.2.6 (star-shaped incircle list)**.** *Let S be a nonempty finite set of sites of E and $x \in E$ that lies on the convex hull of S. Consider C the union of the d-cells of $ICL_D(x)$ and of all its faces. Then C is star-shaped from x, that is for any point $p \in C$, $[x, p] \subset C$.*

(2) The second situation is the simplest. If $F_1, \cdots, F_p$ are the external faces of the triangulation D_i (of any dimension $k < d$) that are visible from s_{i+1}. We can show that the cell complex defined by

$$D_{i+1} := D_i \ \cup \Big\{ \text{cell}(F_j, s_{i+1})_j, \ 1 \leq j \leq p \Big\} \cup \Big\{ \{s_{i+1}\} \Big\}$$

is the Delaunay triangulation associated to S_{i+1}.

(3) In the third situation, if we denote by $\mathcal{S} = ICL_{D_i}(s_{i+1})$ the set of elements of D_i whose circumsphere contains s_{i+1} and $F_1, \cdots, F_p$ are the external faces) of this set which are not visible from s_{i+1} and $F_{p+1}, \cdots, F_{p+q}$ are the external faces of D_i that are not faces of elements of $\mathcal{S}$ and that are visible from x_{i+1}. We can show that the cell complex defined by

$$D_{i+1} := (D_i \backslash \mathcal{S}) \ \cup \Big\{ \text{cell}(F_j, s_{i+1})_j, \ 1 \leq j \leq p \Big\} \cup \Big\{ \{s_{i+1}\} \Big\}$$

is the Delaunay triangulation associated to S_{i+1}.

The first triangulation D_{d+1} is made of a simple simplex defined by the $d+1$ first inserted points.

One important modification of the incremental algorithm consists in inserting sites in a random order. Its expected running time is better than the worst-case running time for the incremental algorithm.

The worst-case complexity of computing the Delaunay triangulation of n points in a d-dimensional Euclidean space E is $O\left(n \log(n) + n^{\lceil \frac{d}{2} \rceil}\right)$.

On the practical implementation

The first step is the Localization. It consists in finding whether the new site x is in the convex hull of S or not, and if it is the case, in what Delaunay cell of the triangulation T_S x lies. A survey on localization methods is available in [2]. When x is inside of the convex hull of S, the localization procedure return the index of the the Delaunay cell where it lies. This corresponds to the Situation (1). When x is outside of this convex hull, the localization returns a Null pointer. This corresponds to Situations (2) and (3).

The second step consists in finding the list of the Delaunay cells whose circumsphere contains x (the incircle list). In Situation (1), this list contains at least the Delaunay cell where x is located. Owing to Proposition 5.2.6, we know that the union of these Delaunay cells is star-shaped so that it can be determined locally by testing connected cells in the graph structure presented above.

The last step consists in deleting the Delaunay cells of the incircle list and connecting the new site to the external faces of the incircle list or the visible faces of the convex hull of S depending on the situation (1), (2) or (3).

5.3 Classical examples of fast nearest neighbor search algorithms in low dimensions

Given a set of n points, $\{x_1, \cdots, x_n\} \subset E$, the nearest neighbor problem is to find the point that is closest to a query point $q \in E$. Many algorithms have been proposed to avoid the large computational cost of the obvious brute force algorithm. When one has to perform a large amount of nearest neighbor searches, a preprocessing of the data set will be profitable if it reduces the average query time.

The problem is optimally solved in the one-dimensional case. The best algorithm consists in sorting the data set by the unique coordinate of its points during a preprocessing stage. (Approximative cost of $O(n\ln(n))$). The search algorithm consists in a simple binary search whose cost is $\frac{\ln(n)}{\ln(2)} + O(1)$.

In the case of low dimensions, most fast search algorithms still have an approximative preprocessing cost of $O(n\log(n))$ and an average search cost in $O(\log(n))$ in low dimension. The criterion of choice among them relies on

- their effective speed on real data sets,
- the required memory,
- the sensitivity of the speed to the dimensionality.

A first obvious optimization called *partial distance search* (P.D.S.) consists in a simple modification of the brute force search: during the calculation of the distance, if the partial sum of square differences exceeds the distance to the nearest neighbor found so far, the calculation is aborted. This almost always speeds up the nearest neighbor search procedure.

5.3.1 The Kd-tree algorithm

The Kd-tree algorithm is the archetype of the branch-and-bound nearest neighbor search tree. It is very popular because of its simplicity.

Building the tree:

- Every point of the data set is associated to the root node.
- The data set is being partitioned into two subsets of cardinal $\left\lfloor \frac{n}{2} \right\rfloor + 1$ or $\left\lfloor \frac{n}{2} \right\rfloor$. The first group corresponds to large values of the first coordinate of the sites, and the second one corresponds to small values.
- Each subset is associated to a child node of the root node.
- The process is repeated on each child node recursively using the coordinate axis in a cyclic order, until there are less than two points in each node.

Searching in the tree: Let q be the query point.

- The search procedure begins by searching in what child node q is (depending of its first coordinate).
- This child node is then searched, and the process is repeated recursively until a terminal node is reached.
- A trivial nearest neighbor search is performed in the terminal node. (Partial Distance Search optimization can be used.)
- The procedure moves up to the parent of the terminal node.
- If the distance d_2 between q and the hyperplane that splits the data set is smaller than the distance $d_{\min}$ to the nearest neighbor found so far, the other child node is searched.
- The procedure continues its way back to the root node.

Complexity: Except in one dimension where the search complexity is logarithmic (it amounts to a binary search), the worst case for the Kd-tree corresponds to the case where every node of the tree is explored. Then the worst-case complexity is time exponential. The distances to every point is computed. The complexity of the preprocessing is $O(d \times n\log(n))$.

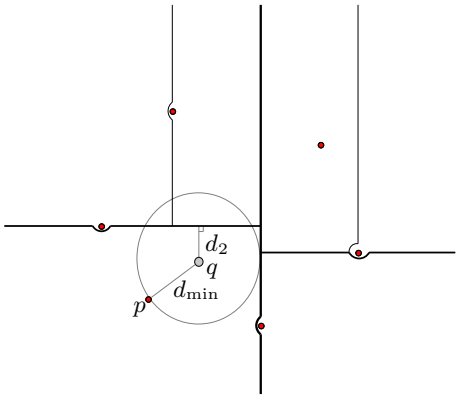

Figure 5.2: K-d tree elimination condition: if the distance d_2 between the query point q and the brother node is smaller than the distance $d_{\min}$ to the nearest neighbor found so far, say p, the brother node has to be explored.

5.3.2 The principal axis tree algorithm

The Principal Axis Tree (PAT) is a generalization of the Kd-tree proposed by McNames in [10]. Instead of using a coordinate axis to sort the data set, its principal axis is used at each step. Moreover, the number of child nodes in the tree can be greater than 2 at each generation.

Building the tree:

- Every point of the data set is associated to the root node.
- The data set is being partitioned in n_c subsets whose cardinality is $\left\lfloor \frac{n}{n_c} \right\rfloor + 1$ or $\left\lfloor \frac{n}{n_c} \right\rfloor$ along its principal axis.
- Each subset is associated to a child node of the root node.
- The process is repeated on each child node recursively until there are less than n_c points in each node.
- At each step, the principal axis, and maximal and minimal values of subset projections on the principal axis are kept in memory.

Optimizing the elimination condition:

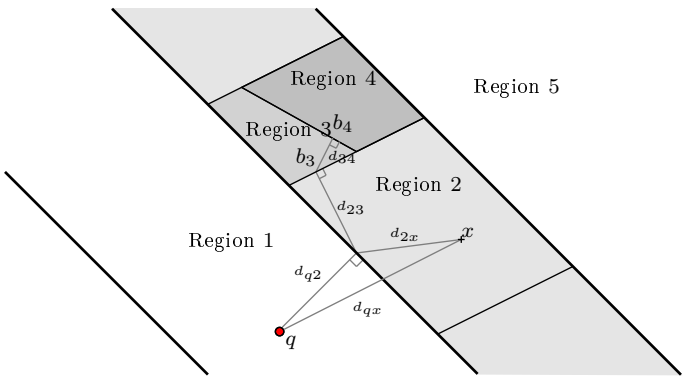

Figure 5.3: Elimination condition of the principal axis tree.

We refer here to Figure 5.3. We can improve the lower bound to the points that belong to child nodes of brother nodes. For any point q in region 1 and x in region 2, we have $d^2(q,x) \geq d_{q2}^2 + d_{2x}^2$. This result is then used again to get a lower bound to points in region 3, and 4 and so on.

$$\begin{array}{rcll} d_{2x}^2 & \geq & d_{23}^2 & \forall x \in \text{ Region 3,} \\ d^2(q,x) & \geq & d_{q2}^2 + d_{23}^2 + d_{34}^2 & \forall x \in \text{ Region 4.} \end{array}$$

Searching in the tree: Let q be the query point.

- The search process begins by searching in which child node q is (by computing its projection on principal axis).
- This child node is then searched, and the process is repeated recursively until a terminal node is reached.
- A partial distance search is then performed in the terminal node.
- The procedure moves up to the parent of the terminal node.
- The elimination condition is checked to decide if brother nodes have to be searched or not.
- The procedure continues its way back to the root node.

Choice of parameter n_c: For normal or uniform random data sets (and distribution of query points), best overall performances are obtained with $n_c = 7$ (independently from dimensionality for $d < 10$). (The same optimal value is obtained by McNames in [10].) In the case where the data set is an optimal quantizer of those distributions, best performance is obtained with $n_c = 13$.

Complexity: Space storage is $O(n)$. Except in the one-dimensional setting where the search complexity is logarithmic (it comes to a binary search), the worst case for the principal axis tree corresponds to the case where every node of the tree is explored. Then the worst-case complexity is time exponential (2^n comparisons of coordinates). n distances are computed. The complexity of the preprocessing is $O(d \times n \log(n))$.

Algorithm performance: On a 5000 points Gaussian data set in $\mathbb{R}^2$, the depth of the tree is 4.

- 27 (partial) distances,
- 15 scalar products,
- 9 binary searches

are performed in average.

Why using this space partitioning? The idea is that good empirical performance of PAT are due to the fact that it takes advantage of the shape of the data set. Yet obviously when both query point distribution and data sets lie on a smaller dimension ($k < d$) subspace of E, one retrieves the same complexity as when using the same algorithm on a k-dimensional space. This intrinsic dimension is often less than the spatial dimension of the space. In a more general setting, PAT takes advantage of large correlations in the data set coordinates.

However if one uses the same number of child nodes n_c in Kd-tree and PAT tree, we see that

- Preprocessing time is longer for PAT than for Kd-tree.
- The first traversal of the tree to a terminal node is more costly (projections have to be computed).

But PAT is still faster because its geometrical partition of the space fits the data set in a more relevant way. To be precise, it happens less often than one has to search a brother node with PAT than with Kd-tree.

In [3], the same space decomposition was proposed for the nearest neighbor search problem (but using the only 2 child node at each generation). They justify the use of this decomposition using a heuristic criterion, according to which the best possible decomposition of the data set into two subsets for branch-and-bound nearest neighbor search is to split the data set with respect to its projection on the principal axis.

5.4 A new quantization-based tree algorithm

As we have seen in previous sections, a good space decomposition that fits to the data distribution may lead to a faster branch-and-bound nearest neighbor search algorithm, if less brother nodes have to be explored. The traversal of the tree can be a little more expensive if it is compensated by the gain due to the fact that less nodes are explored.

Principal component analysis and optimal quantization are two types of projection of a probability distribution. Similar inertia decompositions hold in the quadratic case (Huyghens lemma).

PAT is based on a recursive space decomposition based on the principal component analysis of the underlying data set. The initial idea here is to design a branch-and-bound algorithm based on a recursive quantization of the empirical distribution of the underlying data set.

5.4.1 The crude quantization tree algorithm

Building the tree:

- Every point of the data set is associated to the root node.
- The data set is being partitioned into n_c subsets corresponding to the Voronoi cells of an optimized quantizer of the empirical distribution of the data set.
- Each subset is associated to a child node of the root node.
- The process is repeated on each child node recursively until there are less than a certain number of points in each node.

Some other computations are done during the preprocessing that will be detailed further on.

Remark. *One notices that the resulting search tree is not balanced and may have some longer branches.*

Searching in the tree: Let q be the query point.

- By performing trivial nearest neighbor searches in the node quantizer the search algorithm traverses the tree to a terminal node where a trivial partial distance search is performed.
- The procedure moves up to the parent of the terminal node.
- The elimination condition, (developed further on) is checked to decide whether brother nodes have to be searched or not.
- The procedure continues its way back to the root node.

Consistency of the space decomposition:

Implementing only the way down to the terminal node (with $n_c = 7$ in both principal axis tree and quantization tree), we naturally do not obtain always the index of the nearest neighbor. But we have noticed that the result is more often the right one with the quantization tree than with the principal axis tree.

For instance, in dimension 2, on a 5000 points Gaussian data set, on a million Gaussian query points, we notice:

- 56 percent of false results with PAT.
- 16 percent of false results with the quantization tree.

Similar results are obtained with other values of the parameters and other data set distributions. This empirical test makes us reasonably optimistic about the performance of a branch-and-bound tree based on this decomposition.

Still, the cost of the way through the search tree is more expansive with the quantization tree (as described above).

- For the "quantization tree", we have to perform trivial nearest neighbor search to find the right child node.
- For "principal axis tree", we only compute a projection and perform a binary search.

Moreover, it was proved in [13] that in the case of Gaussian distributions, the affine subspace spanned by stationary quantizers corresponds to the first principal components of the considered Gaussian distribution. (This result, extended to the infinite-dimensional case in [8] allows us to efficiently compute optimal quadratic quantizers of bi-measurable Gaussian processes.) Hence, in this case, this shows that the quantization tree with two branches at each generation is related to the principal axis tree.

First elimination condition If the center of the Voronoi cell corresponding to the current node is A, the first rough method to decide whether a brother node with center B has to be explored or not is compute the distance d_2 of the query point Q to the Leibniz halfspace $H(B, A)$. Then the node corresponding to point B is explored if d_2 is smaller than the distance to the nearest neighbor found so far, d_1. We have $d_2 = \frac{AB}{2} - AQ\cos\alpha$ where α is the angle between AQ and AB and $QB^2 = QA^2 + AB^2 - 2AQAB\cos\alpha$ so that $\Rightarrow \cos\alpha = \frac{QA^2+AB^2-QB^2}{2AQAB}$. This yields $d_2 = \frac{QB^2-QA^2}{2AB}$. Hence, the computation of the distance to the Leibniz halfspace requires one subtractions QA^2-QB^2, (QA^2 and QB^2 can be computed during the search in the quantizer in the parent node), and one multiplication by $\frac{1}{2AB}$. ($\frac{1}{2AB}$ can be computed during the preprocessing.)

Thus it is clear that the nearest brother node corresponds to the second nearest neighbor in the quantizer, and the second nearest to the third nearest neighbor, and so on. Hence, brother nodes have to be explored in the order defined by the distances of its centers to the query point.

We can also use the same optimization of the lower bound proposed by McNames in [10] and presented in Section 5.3.2. Referring to Figure 5.4, the lower bounds d_i are recursively incremented when exploring brother nodes.

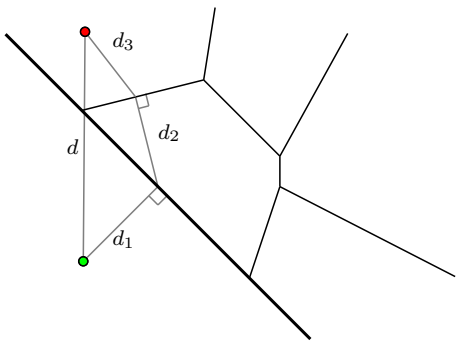

Figure 5.4: Optimization of the elimination condition for the quantization tree $d^2 \geq d_1^2 + d_2^2 + d_3^2$.

Performance of this first quantization tree algorithm. This first algorithm has been implemented and its empirical performances have been compared to the two previously exposed PAT and Kd-tree in terms of empirical performances.

Intermediate performances between our implementations of Kd-tree and PAT were obtained in small dimensions. Although, as we will see further in empirical tests, it seems to handle better the increase of dimensionality. The preprocessing time, that requires small quantizer computations is also more costly than both PAT and Kd-tree.

5.4.2 Optimizations for the quantization tree

To reduce the average query time, we are now proposing a new optimization procedure which reduces the number of brother nodes to be checked.

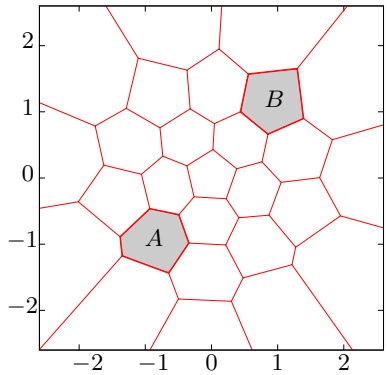

Figure 5.5: Cell B is "hidden" from cell A.

Let us consider the Voronoi diagram plotted in Figure 5.5. In this figure, we obviously know that when the query point is in a cell A, its nearest neighbor cannot be in cell B, because cell B is "hidden" by closer cells. One has to give a precise mathematical sense to "hidden" in this sentence. However, in the quantization tree as it has been described, the distance of query point to $H(a, b)$ has to be computed.

A first idea is to compute for each $1 \leq i \leq n_c$ a list of "friends" among brother nodes in which the nearest neighbor can be when q is in cell i.

This list has to be large enough to ensure that it contains the nearest neighbor but as small as possible in order to reduce the computations of elimination conditions.

As concerns the choice of the parameter n_c, we have to take into consideration that increasing n_c makes the depth of the tree smaller but also makes the nearest neighbor search slower for each generation of the search tree.

How can we obtain a friend Voronoi cells list? The first observation about obtaining such a friend list is that it is not a simple problem. Indeed, this list is a priori not reduced to adjacent cells in the Voronoi diagram. Moreover, in some cases, the minimal friend list can be quiet large. So is the case for unbounded Voronoi cells for example.

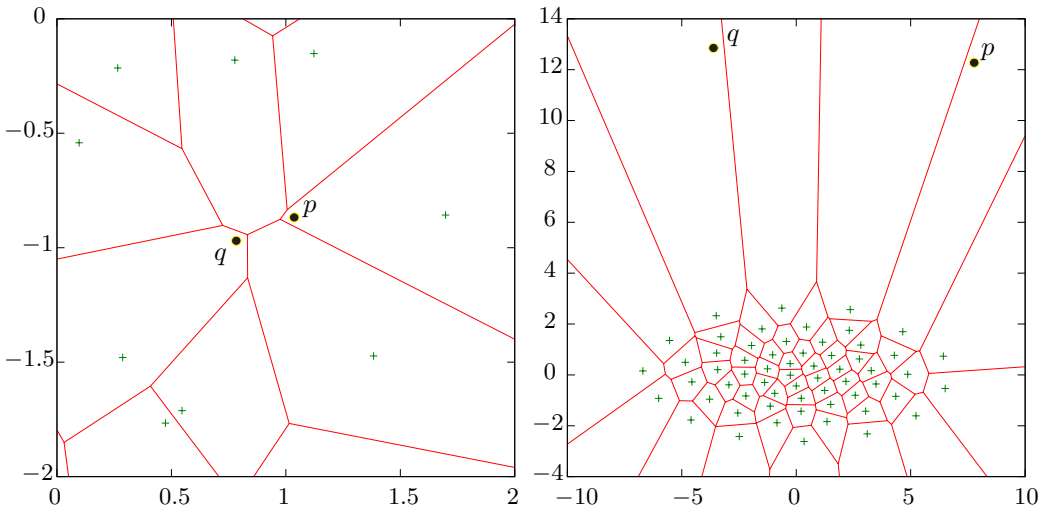

Figure 5.6: In these cases, the nearest neighbor of the query point q may be p although p is not in an adjacent Voronoi cell.

A procedure to obtain such a friend Voronoi list is proposed in Section 5.5.

5.5 Some optimizations for the quantization tree algorithm

In Section 5.2.1, basic definitions about Voronoi diagrams and Delaunay triangulations that are prerequisites to this section have been recalled.

Remark (Voronoi slabs and Voronoi cells)**.** *From their respective definitions, one can easily deduce the following properties:*

- *Let $S \subset E$ be a finite set of sites, let C be an associated Voronoi partition and consider $s \in S$. Then it is clear that $V(\{s\}) = \overset{\circ}{\widehat{\mathrm{slab}_C(s)}}$.*
- *The points of the Voronoi cells $V(R)$ with $R \subset S$ and $\operatorname{card} R > 1$ belong to the boundaries of Voronoi slabs.*
- *As a consequence, for $s \in S$, as the boundary $\underline{V(\{s\})}$ is constituted of its faces of lower dimensions, a previous remark yields $\overline{V(\{s\})} = \overline{\mathrm{slab}(s)}$ and $\partial\,\mathrm{slab}_S(s) = \partial V_S(\{s\})$.*

Notations: In the following of this section, if $S \subset E$ is a finite set of sites in E, one will denote by T_S the Delaunay triangulation of S, DG_S the Delaunay graph of S, V_S its Voronoi diagram. For $R \subset S$, $V_S(R)$ will represent the Voronoi cell of R in S. If C_S is a Voronoi partition associated to S, and $s \in S$, $\mathrm{slab}_S(s)$ will denote the Voronoi slab associated to S is the Voronoi partition C.

Proposition 5.5.1. *An obvious property is if S is a finite set of sites of E, and $p \in S$,*

$$V_S(\{p\}) = \bigcap_{s \in S, s \neq p} H(p,s).$$

Proposition 5.5.2. *If S is a finite set of sites of E, and $p \in S$, $V_S(\{p\}) = \bigcap_{\{s,p\} \in DG_S} H(p,s)$.*

Lemma 5.5.3. *Let $S \subset E$ be a nonempty finite set of sites in E and $x \in E \backslash S$. Consider $s \in S$, the following assertions are equivalent:*

1. $\{x,s\} \in DG_{S \cup \{x\}}$.
2. $V_S(\{s\}) \cap V_{S \cup \{x\}}(\{x\}) \neq \emptyset$.
3. $V_S(\{s\}) \cap H(x,s) \neq \emptyset$.

Proof: See Figure 5.7 for an illustration of the proof.

- (1. $\Rightarrow$ 2.) Assume that $\{x,s\} \in DG_{S \cup \{x\}}$ then by definition, it is equivalent to $V_{S \cup \{x\}}(\{x,s\}) \neq \emptyset$.

  $V_{S \cup \{x\}}(\{x,s\})$ is $(d-1)$-face of $V_{S \cup \{x\}}(x)$. Moreover, by definition of Voronoi cells, $V_{S \cup \{x\}}(\{x,s\}) \subset V_S(\{s\})$, which is open. As a consequence, $\forall y \in V_{S \cup \{x\}}(\{x,s\})$, $\forall \varepsilon > 0$, $B(y,\varepsilon) \cap V_{S \cup \{x\}}(x) \neq \emptyset$. And for small enough ε, $B(y,\varepsilon) \subset V_S(\{s\})$. We conclude that $V_S(\{s\}) \cap V_{S \cup \{x\}}(\{x\}) \neq \emptyset$.
- (2. $\Rightarrow$ 3.) is obvious owing to Proposition 5.5.1.
- (3. $\Rightarrow$ 1.) If $y \in V_S(\{s\}) \cap H(x,s)$, let us show that $V_{S \cup \{x\}}(\{x,s\}) \neq \emptyset$.

  Consider the segment $[s,y]$. By convexity, $[s,y] \subset V_S(\{s\})$. Thus every point of $[s,y]$ is closer to s than to any other point of S. On the other hand, it can either be closer to s than to x, or closer to x than to s or at the same distance.

  We now define the applications $f : [0,1] \to [s,y] \subset E$ by $f(\lambda) = \lambda s + (1-\lambda) y$ and $\Delta : E \to \mathbb{R}$ by $\Delta(p) = d(p,x) - d(p,s)$.

  $\Delta \circ f$ is a continuous function with $\Delta \circ f(0) > 0$, $\Delta \circ f(1) < 0$. The intermediate value theorem shows that there exists λ^* such that $\Delta \circ f(\lambda^*) = 0$ and thus $f(\lambda^*) \in V_{S \cup \{x\}}(\{x,s\})$. $\square$

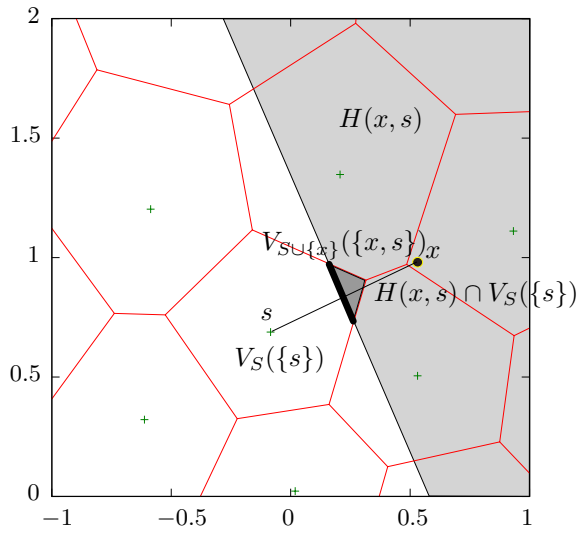

Figure 5.7: If the query point q lies on the dark gray region $H(x,s)\cap V_S(\{s\})$ its nearest neighbor may be x.

The first modification made in the quantization tree algorithm is to assume that the points of the quantizer at each generation are points of the underlying codebook Γ. (In order to fulfill this requirement, we project an optimal quantizer onto the codebook.)

Corollary 5.5.4. *Let $\Gamma = \{\gamma_1, \cdots, \gamma_n\}$ be a codebook of E. $S = \{s_1, \cdots, s_p\} \subsetneq \Gamma$ be a subset of Γ. Let Proj_Γ be a nearest neighbor projection on Γ. Γ is being partitioned into p subsets $\Gamma^1, \cdots, \Gamma^p$ with $\Gamma^i = \Gamma \cap \mathrm{slab}_S(s_i)$, by their nearest neighbor projection on S.*
Consider $q \in E$. If $q \in \mathrm{slab}_S(s)$ for some $s \in S$ and $t = \mathrm{Proj}_\Gamma(s)$ then $\{t, s\} \in DG_{S\cup\{t\}}$.

Proof: This is a straightforward consequence of the previous lemma. □

Notation: Let S be a set of sites in E. For a point t in E, we denote $PI_S(t) = \Big\{s \in S,\ \{s,t\} \in DG_{S\cup\{t\}}\Big\}$. The notation PI stands for "Pseudo-Insertion".

From an algorithmic viewpoint, the Delaunay graph of S being computed, $PI_S(t)$ stands for the sets of points in S, that are connected to t when updating the Delaunay graph to take account of this new point.

Implementing a procedure that computes $PI_S(t)$ is very similar to the insertion procedure of a point t in T_S.

First friend node algorithm: This leads to a first method to compute a friend list:

For every point p of the underlying codebook,

- Compute $s = \mathrm{Proj}_S(p)$ and $PI_S(p)$.
- Then for every point $s' \in PI_S(p)$, insert s in the set of friends of node s'.

This method gives a first algorithm to compute friend list. Still, when the data set is large, it is very expensive because one has to deal with all the points of the data set.

In fact it is possible to compute an acceptable friend list by using Lemma 5.5.3 without using the points of the underlying data set.

Fast friend node algorithm: In this section, another method to compute friend node lists is devised which does not need to deal with the complete underlying data set but only the underlying codebook.

When keeping the same notations, the principle of the method is to compute for every $\mathrm{slab}_S(s)$, $s \in S$, of the Voronoi partition C_S, the set $UPI_S(s) := \bigcup_{p \in \mathrm{slab}_S(s)} PI_S(p)$. It is the union of all the pseudo-insertions of points of $\mathrm{slab}_S(s)$. If one is able to compute this set, the resulting friend nodes algorithm simply writes:

For every point $s \in S$,

- Compute $UPI_S(s)$.
- Then for every point $s' \in UPI_S(s)$, insert s in the set of friends of node s'.

The question is: how can we compute $UPI_S(s)$?

Lemma 5.5.5. *With the same notations, one has* $UPI_S(s) = \bigcup_{p \in \partial\, \mathrm{slab}_S(s)} PI_S(p)$. *In other words, we have to check points of the boundary* $\partial\, \mathrm{slab}_S(s)$ *of* $\mathrm{slab}_S(s)$.

Remark. *Let us recall that, thanks to Proposition 5.2.2,* $(\partial\, \mathrm{slab}_S(s) = \partial V_S(\{s\})$.

Proof: Consider $x \in \mathrm{slab}_S(s)$ such as $s' \in PI_S(x)$. Let us define x^*, such that $\{x^*\} = [x, s'] \cap \partial V_S(s)$.

- One has $H(x^*, s') \supset H(x, s')$. So $V_S(\{s'\}) \cap H(x^*, s') \supset V_S(\{s'\}) \cap H(x, s')$, hence $V_S(\{s'\}) \cap H(x, s') \neq \emptyset \Rightarrow V_S(\{s'\}) \cap H(x^*, s') \neq \emptyset$ that is equivalent to $s' \in PI(x^*)$ thanks to the Lemma 5.5.3.
- Finally, $\forall x \in \mathrm{slab}_S(s), \forall s' \in PI_S(x), \exists x^* \in \partial\, \mathrm{slab}_S(s)$ such that $s' \in PI_S(x^*)$. □

Remark. *As there are not a finite number of sites on the boundaries, this does not give an effective method for computing* $UPI_S(s)$ *yet.*

As seen in Section 5.2.1, computing the set $PI_S(x)$ corresponds almost to the same algorithm as the insertion procedure in an incremental triangulation algorithm, that is:

- Localization of x in the triangulation,
- Computation of the set $ICL(x)$,
- $UI_S(x)$ is the set of points that belong to a cell of $ICL(x)$ plus, if x is outside the convex hull of S, the points of the external faces of T_S that are visible from x.

Lemma 5.5.6. *Let S be a nonempty finite set of sites in E. We consider the circumsphere C of Delaunay d-cell of the Delaunay triangulation T_S. We denote by c its center and r its radius. Let s be a site of S.*
If $V_S(\{s\}) \subset C \neq \emptyset$ *then* $c + \frac{r}{|s-c|}(s - c) \in V_S(\{s\})$.

The proof is straightforward. This leads to an algorithm to compute sets $(UPI_S(s))_{s \in S}$.

- For every Delaunay d-cell D of T_S
  - Compute the center c and radius r of its circumsphere.
  - For every site $s \in S$ that is not in D, compute $p := c + r\frac{s-c}{|s-c|} \in V_S(\{s\})$, and check if the site s is the nearest neighbor of p in S. If so is the case, then the points of the Delaunay d-cells D belong to $UPI_S(s)$.
- Then deal with unbounded Voronoi cells:
  - For every external face F of the Delaunay triangulation, compute a normal vector u_F directed toward the exterior of the convex hull of S.

– For two distinct external faces F_1 and F_2 of the Delaunay triangulation, if $\langle u_{F_1}, u_{F_2} \rangle > 0$ then for every $(s_1, s_2) \in F_1 \times F_2$, $s_1 \in UPI_S(s_2)$ and $s_2 \in UPI_S(s_1)$.

In Figure 5.8, we present some friend Voronoi lists in the 2-dimensional case.

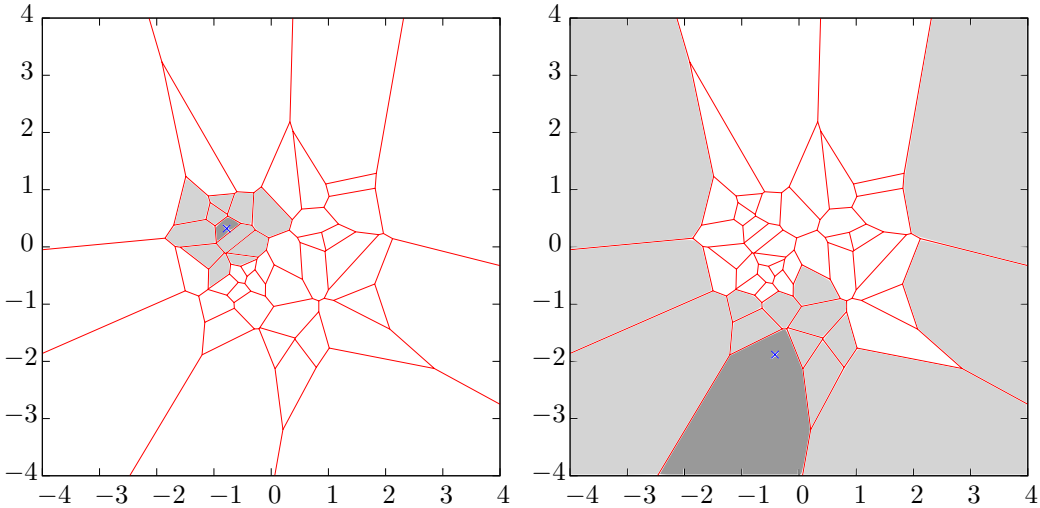

Figure 5.8: Examples of friend Voronoi cells in a two-dimensional Voronoi diagram in the case of a bounded Voronoi cell (left) and in the unbounded case (right). In both case, the dark gray region is the considered Voronoi cell and the light gray regions are the friend Voronoi cells.

5.6 Test with real data sets

To perform the following tests, the quantization tree algorithm and the friend-node optimization have been implemented in the `C++` programming language. Because of the additional features related to computational geometry that we needed, as the pseudo-insertion procedure, we had to implement a Delaunay triangulation. All the figures presented in this chapter were generated with this implementation of the Voronoi diagram with which we performed the following tests.

5.6.1 Tests on Gaussian and uniform data sets

In Tables 5.1, 5.2 and 5.3, we report the execution time for 10 millions independent nearest neighbor queries on data sets of size 5000 generated with a Gaussian pseudorandom generator and with a uniform pseudorandom generator on the hypercube. The best overall performances were obtained with $n_c = 35$ children by node for the quantization tree. The tests were performed with an Intel Pentium Dual CPU at 2GHz. We noticed that in dimensions $d = 2$ and $d = 3$, we had intermediate performances between the "principal axis tree" and the Kd-tree algorithms. In dimension 4, the performance of the "principal axis tree" and the "quantization tree" are close one to each other. Finally, it seems that the quantization tree has a better behaviour in dimensions greater than 5 where it significantly outperforms the two other implemented methods.

Remark (Computational cost or the preprocessing for the friend cell algorithm)**.** *An important fact that we have experienced is that, in higher dimensions, the friend cells list becomes larger and there is no more competitive advantage in using it in dimension higher than* 7 *(when having less than* 30 *branches per generation in the quantization tree). Moreover, as it requires to compute Delaunay triangulations during the preprocessing, whose complexity exponentially increases with*

	$d=2$	$d=3$	$d=4$	$d=5$	$d=6$	$d=7$	$d=8$
Quantization tree	1.76s	2.75s	5.35s	8.93s	15.99s	28.06s	52.31s
Principal axis tree	1.21s	1.86s	4.49s	10.87s	20.14s	41.56s	82.30s
Kd-tree	1.88s	3.71s	8.54s	17.13s	31.06s	60.67s	118.93s

Table 5.1: Execution time of 10 millions random queries on a data set of 5000 points, generated with a Gaussian pseudorandom generator.

	$d=2$	$d=3$	$d=4$	$d=5$	$d=6$	$d=7$	$d=8$
Quantization tree	2.59s	3.87s	6.46s	11.90s	27.54s	45.78s	84.63s
Principal axis tree	1.33s	2.44s	4.94s	12.78s	41.02s	62.33s	119.88s
Kd-tree	2.82s	5.20s	11.32s	24.20s	47.51s	87.61s	164.52s

Table 5.2: Execution time of 10 millions random queries on a data set of 10000 points, generated with a Gaussian pseudorandom generator.

	$d=2$	$d=3$	$d=4$	$d=5$	$d=6$	$d=7$	$d=8$
Quantization tree	1.62s	2.30s	3.75s	6.47s	10.33s	15.91s	32.62s
Principal axis tree	0.74s	1.52s	2.81s	6.71s	16.53s	28.03s	47.53s
Kd-tree	1.54s	2.82s	5.46s	10.64s	18.50s	31.60s	55.71s

Table 5.3: Execution time of 10 millions random queries on a data set of 5000 points, generated with a uniform pseudorandom generator.

the dimension, the computational cost of the friend cell preprocessing makes it useless in higher dimensions.

Bibliography

[1] Jon Louis Bentley. Multidimensional binary search trees used for associative searching. *Commun. ACM*, 18(9):509–517, 1975.

[2] Olivier Devillers, Sylvain Pion, and Monique Teillaud. Walking in a triangulation. *Internat. J. Found. Comput. Sci.*, 13:181–199, 2002.

[3] Wim D'Haes, Dirk van Dyck, and Xavier Rodet. An efficient branch and bound seach algorithm for computing k nearest neighbors in a multidimensional vector space. *IEEE Advanced Concepts for Intelligent Vision Systems (ACIVS)*, 2002.

[4] Steven Fortune. Voronoi diagrams and Delaunay triangulations. In Ding-Zhu Du and Frank Hwang, editors, *Computing in Euclidean Geometry*, pages 193–233. World Scientific, 1992.

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

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

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

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

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

[10] James McNames. A fast nearest-neighbor algorithm based on a principal axis search tree. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(9):964–976, 2001.

[11] Gilles Pagès and Benedikt Wilbertz. Intrinsic stationarity for vector quantization: Foundation of dual quantization. *Preprint*, 2010.

[12] Gilles Pagès and Benedikt Wilbertz. Sharp rate for the dual quantization problem. *Preprint*, 2010.

[13] Thaddeus Tarpey, Luning Li, and Bernard D. Flury. Principal points and self-consistent points of elliptical distributions. *Ann. Stat.*, 23(1):103–112, 1995.

Appendix A

Accurate quadratic quantization of the one-dimensional Gaussian distribution

Abstract

In this chapter, we detail properties of the deterministic methods available to compute optimal quadratic quantization of one-dimensional distributions. We focus on the case of the one-dimensional standard Gaussian distribution for which we provide a database of optimal quantizers for a wide range of values of the quantizer size. This precomputed database allows a faster computation of the functional quantization of Gaussian processes, and is also useful for other purposes. The methods used to compute this database are available with an arbitrary precision. The numerical values provided in the database have a relative precision of 10^{-32}.

Keywords: Gaussian distribution, elliptic distribution, vector quantization, log-concave, numerical integration, arbitrary precision, floating point, Lloyd algorithm, Newton-Raphson.

Introduction

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. The principle of the quantization of a $\mathbb{R}^d$-valued random variable X is to approximate X by a random variable Y taking at most N values. The discrete random variable Y is a quantizer of X. The resulting error is the L^p norm of $|X - Y|$ where $|\cdot|$ is the Euclidean norm on $\mathbb{R}^d$ and $p \geq 1$. The minimization of the error yields the following minimization problem

$$\min\{\|X - Y\|_p, Y : \Omega \to \mathbb{R}^d \text{ measurable }, \text{card}(Y(\Omega)) \leq N\}. \tag{A.1}$$

This problem has been initially investigated for its application to signal transmission issues [8]. More recently, quantization has been introduced in numerical probability to devise quadrature methods and variance reduction algorithms [23, 5]. The case of other metrics and spaces (as infinite-dimensional functional spaces) has been investigated in the literature [9, 14].
Optimal quantization has also been used for automatic classification issues [26], and as a grid generation method [6, 27].

Definition A.0.1 (Voronoi partition). *Consider $N \in \mathbb{N}^*$, $\Gamma = \{\gamma_1, \cdots, \gamma_N\} \subset \mathbb{R}^d$ and let $C = \{C_1, \cdots, C_N\}$ be a Borel partition of $\mathbb{R}^d$. C is a Voronoi partition associated with Γ if $\forall i \in \{1, \cdots, N\}$, $C_i \subset \{\xi \in \mathbb{R}^d, |\xi - \gamma_i| = \min_{j \in \{1, \cdots, N\}} |\xi - \gamma_j|\}$.*

If $C = \{C_1, \cdots, C_N\}$ is a Voronoi partition associated with $\Gamma = \{\gamma_1, \cdots, \gamma_N\}$, it is clear that $\forall i \in \{1, \cdots, N\}, \gamma_i \in C_i$. C_i is called Voronoi slab associated with γ_i in C and γ_i is the center of the slab C_i.
One denotes $C_i = \text{slab}_C(\gamma_i)$, and for every $a \in \Gamma$, $W(a|\Gamma)$ is the closed subset of $\mathbb{R}^d$ defined by $W(a|\Gamma) = \left\{y \in \mathbb{R}^d, |y - a| = \min_{\gamma \in \Gamma} |y - \gamma|\right\}$.

Definition A.0.2 (Nearest neighbour projection). *Let us consider the fixed point set $\Gamma = \{\gamma_1, \cdots, \gamma_N\} \subset \mathbb{R}^d$ and $C = \{C_1, \cdots, C_N\}$ a Voronoi partition associated with Γ. The nearest neighbour projection on Γ is the application $\text{Proj}_\Gamma := \sum_{1=1}^{N} \gamma_i \mathbf{1}_{C_i}$.*

Proposition A.0.1. *Let X be an $\mathbb{R}^d$-valued L^p random variable, and Y taking its values in the fixed point set $\Gamma = \{y_1, \cdots, y_N\} \subset \mathbb{R}^d$ where $N \in \mathbb{N}$. Set $\widehat{X}^\Gamma$ the random variable defined by $\widehat{X}^\Gamma := \text{Proj}_\Gamma(X)$ where Proj_Γ is a nearest neighbour projection on Γ, called a Voronoi Γ-quantizer of X.*
Then we clearly have $\left|X - \widehat{X}^\Gamma\right| \leq |X - Y|$ a.s.. Hence $\left\|X - \widehat{X}^\Gamma\right\|_p \leq \|X - Y\|_p$.

A consequence of this remark is that solving the minimization problem (A.1) amounts to solving the simpler minimization problem

$$\min\left\{\|X - \text{Proj}_\Gamma(X)\|_p,\ \Gamma \subset \mathbb{R}^d,\ \text{card}(\Gamma) \leq N\right\}. \tag{A.2}$$

The quantity $\|X - \text{Proj}_\Gamma(X)\|_p$ is called mean L^p-quantization error of L^p distorsion. When this minimum is reached, we refer to L^p-optimal quantization.
The problem of the existence of a minimum has been investigated for decades on its numerical and theoretical aspects in the finite-dimensional case [18, 9]. For every $N \geq 1$, the L^p-quantization error is Lipschitz continuous and reaches a minimum. An N-tuple that achieves the minimum has pairwise distinct components, as soon as $\text{card}(\text{supp}(\mathbb{P}_X)) \geq N$. This result stands in the general case of a random variable valued in a reflexive separable Banach space [14]. If $\text{card}(X(\Omega))$ is infinite, this minimum strictly decreases to 0 as N goes to infinity. The asymptotic rate of convergence for non-singular distributions is ruled by the Zador theorem [9]. A non-asymptotic upper bound for the quantization error is also available [15].
We now focus on the quadratic case ($p = 2$). For a L^2 random variable X, let us now denote by $\mathcal{C}_N(X)$ the set of L^2-optimal quantizers of X of level N and $e_N(X)$ the minimal quadratic

distortion that can be achieved when approximating X by a quantizer of level N. A quantizer Y of X is stationary (or self-consistent) if $Y = \mathbb{E}[X|Y]$.

Proposition A.0.2 (Stationarity of L^2-optimal quantizers)**.** *A (quadratic) optimal quantizer is stationary.*

The stationarity is a particularity of the quadratic case. In other L^p cases, a similar property involving the notion of p-center occurs. A proof is available in [10].

A.1 Optimization algorithms for quadratic vector quantization

Various algorithms have been developed to obtain numerically an optimal N-grid with a minimal quadratic quantization error in the finite-dimensional setting. A review of these methods is available in [23]. The most common method is the Lloyd algorithm [13]. Another algorithm is a stochastic gradient method, which is suggested by the fact that the L^2-quantization error function is differentiable at any N-tuple having pairwise distinct components and a $\mathbb{P}_X$-negligible Voronoi tessellation boundary, and has an integral representation. The algorithm is investigated in [19]. Let us also mention an evolutionary algorithm approach by Hamida and Mrad in [17].

The Lloyd algorithm.
If Γ^0 is a quantizer of X of size N, whose points are all distinct, we define the sequence of quantizers $(\Gamma^s)_{s\in\mathbb{N}}$ by

$$\begin{aligned}\Gamma^{s+1} &= \mathbb{E}[X|\operatorname{Proj}_{\Gamma^s}(X)](\Omega)\\ &= (\mathbb{E}[X|X\in C_i(\Gamma^s)])_{1\le i\le N}\,,\end{aligned}$$

where $C_i(\Gamma^s)$ is the ith cell of the Voronoi tessellation associated with Γ^s.

Proposition A.1.1. *The sequence* $(\|X-\operatorname{Proj}_{\Gamma^s}(X)\|_2)_{s\in\mathbb{N}}$ *is decreasing.*

However, it is not proved that $\operatorname{Proj}_{\Gamma^s}(X)$ converges in the general case. When it does, the limit verifies the stationarity property $\widehat{X} = \mathbb{E}\left[X\middle|\widehat{X}\right]$.

Proof: For every $s\in\mathbb{N}$, let us denote by $C^s = \{C_1^s,\cdots,C_N^s\}$ the Voronoi partition associated with Γ^s that verifies

$$\operatorname{Proj}_{\Gamma^s} = \sum_{i=1}^N \Gamma_i^s \mathbf{1}_{C_i^s}.$$

If we define $G_i^s = \mathbb{E}[X|X\in C_i^s]$, Lloyd's algorithm simply writes $\Gamma_i^{s+1} = G_i^s$. Thanks to Proposition A.0.1

$$\|X-\operatorname{Proj}_{\Gamma^{s+1}}(X)\|_2 \le \|X-\mathbb{E}[X|\operatorname{Proj}_{\Gamma^s}(X)]\|_2. \tag{A.3}$$

Moreover, by definition of the conditional expectation, $\mathbb{E}[X|\operatorname{Proj}_{\Gamma^s}(X)]$ is the best L^2 approximation of X by a function of $\operatorname{Proj}_{\Gamma^s}(X)$, hence

$$\|X-\mathbb{E}[X|\operatorname{Proj}_{\Gamma^s}(X)]\|_2 \le \|X-\operatorname{Proj}_{\Gamma^s}(X)\|_2. \tag{A.4}$$

The combination of both Inequalities (A.3) and (A.4) yields the conclusion. □

More results are available in the one-dimensional setting. If the density of X is logarithmically concave, there exists a unique optimal quantizer, and the Lloyd algorithm converges toward this quantizer exponentially fast [11]. A deeper study of the properties of the Lloyd algorithm is provided by Du and Emelianeko in [7].

As concern the L^p setting with $p\neq 2$, a method similar to Lloyd's algorithm using conditional p-centers instead of conditional expectations can be used (see *e.g.* [25]).

A.2 Explicit Newton-Raphson procedure

Let X be an $\mathbb{R}^d$-valued L^2 random variable. Let us denote by D_N^X the squared quadratic quantization error associated with a codebook $\Gamma = \{\gamma_1, \cdots, \gamma_N\}$ of size N with respect to X.

$$\begin{array}{rccc} D_N^X: & \left(\mathbb{R}^d\right)^N & \to & \mathbb{R}_+ \\ & (\gamma_1, \cdots, \gamma_N) & \mapsto & \mathbb{E}\Big[\min\limits_{1\le i\le N} |X-\gamma_i|^2\Big]. \end{array}$$

The distortion function D_N^X is $|\cdot|$-differentiable at N-quantizers $\Gamma = \{\gamma_1, \cdots, \gamma_N\} \in \left(\mathbb{R}^d\right)^N$ with pairwise distinct components and a $\mathbb{P}_X$-negligible Voronoi partition boundary $\bigcup\limits_{i=1}^N \partial C_i(\Gamma)$. Moreover, for such a quantizer Γ,

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

It is proved in [9] that every optimal quantizer satifies $\mathbb{P}_X\left(\bigcup\limits_{i=1}^N \partial C_i(\Gamma)\right) = 0$. (In particular, a Voronoi quantizer associated with a critical point of D_N^X is a stationary quantizer.)

We also notice that for such a quantizer (verifying the boundary condition), Equation (A.5) can be differentiated again. Thus, as soon as the starting point is close enough to a local minimum of the quadratic distortion, we can use a Newton-Raphson minimization procedure. (An N-optimal codebook has N pairwise distinct points as soon as the approximated distribution does weight more than N points.)

In the one-dimensional setting, if X is a $\mathbb{R}$-valued random variable and if $\mathbb{P}_X$ is absolutely continuous with respect to the Lebesgue measure, the Newton-Raphson method takes a simple form. We denote by f the probability density function and F the probability cumulative function of X. Consider $x = \{x_1, \cdots, x_N\}$, $x_1 < \cdots < x_N$ a codebook. Set $x_{j\pm\frac{1}{2}} := (x_j + x_{j\pm1})/2$, $j = 1, \cdots, N-1$, $x_{\frac{1}{2}} = -\infty$ and $x_{N+\frac{1}{2}} = \infty$. The quadratic distortion of the Voronoi quantization of X corresponding to this codebook writes

$$D_N^X(x) = \sum_{j=1}^N \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} (x_j - \xi)^2 f(\xi)d\xi. \tag{A.6}$$

The Newton-Raphson procedure for optimal quadratic codebook computation is used in [19] for the Gaussian case.

A.2.1 The Hessian matrix of the quadratic distortion

With these notations, and thanks to (A.5)

$$\frac{\partial D_N^X(x)}{\partial x_i} = 2x_i\left(F\left(x_{i+\frac{1}{2}}\right) - F\left(x_{i-\frac{1}{2}}\right)\right) - 2\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \xi f(\xi)d\xi. \tag{A.7}$$

The computation of the second order derivatives is then straightforward.

$$\left\{\begin{array}{ll} \frac{\partial^2 D_N^X(x)}{\partial x_i^2} & = 2\left(F\left(x_{i+\frac{1}{2}}\right) - F\left(x_{i-\frac{1}{2}}\right)\right) + x_i\left(f\left(x_{i+\frac{1}{2}}\right) - f\left(x_{i-\frac{1}{2}}\right)\right) - x_{i+\frac{1}{2}}f\left(x_{i+\frac{1}{2}}\right) + x_{i-\frac{1}{2}}f\left(x_{i-\frac{1}{2}}\right) \\ & = 2\left(F\left(x_{i+\frac{1}{2}}\right) - F\left(x_{i-\frac{1}{2}}\right)\right) + \frac{1}{2}f\left(x_{i+\frac{1}{2}}\right)(x_i - x_{i+1}) + \frac{1}{2}f\left(x_{i-\frac{1}{2}}\right)(x_{i-1} - x_i) \\ \frac{\partial^2 D_N^X}{\partial x_i \partial x_{i+1}} & = \left(x_i - x_{i+\frac{1}{2}}\right)f\left(x_{i+\frac{1}{2}}\right) = \frac{1}{2}(x_i - x_{i+1})f\left(x_{i+\frac{1}{2}}\right) \\ \frac{\partial^2 D_N^X}{\partial x_i \partial x_{i-1}} & = \left(x_{i-\frac{1}{2}} - x_i\right)f\left(x_{i-\frac{1}{2}}\right) = \frac{1}{2}(x_{i-1} - x_i)f\left(x_{i-\frac{1}{2}}\right). \end{array}\right. \tag{A.8}$$

The other coefficients of the Hessian matrix are equal to zero.

A.2.2 Tridiagonal Newton-Raphson

We are now able to implement a Newton-Raphson method to find a zero of dD_n^X in $\mathbb{R}^N$, starting from $x^{(0)} \in \mathbb{R}^N$. We compute recursively

$$x^{(n+1)} = x^{(n)} - \left[d^2 D_N^X \left(x^{(n)}\right)\right]^{-1} \cdot dD_N^X(x^{(n)}).$$

In practice, we do not have to explicitly invert matrix $d^2 D_N^X(x^{(n)})$. Indeed, this comes to $x^{(n+1)} = x^{(n)} - \gamma_n$, where γ_n is the solution of the tridiagonal linear system $\left[d^2 D_N^X \left(x^{(n)}\right)\right] \gamma_n = dD_N^X \left(x^{(n)}\right)$. The resolution of a tridiagonal linear system has a $O(N)$ complexity [24]. Still, in some cases, the algorithm proposed in [24] may fail. The convergence is only guaranteed for matrices having the diagonal dominance property, which is not verified here. An implementation for the general case, using elimination with partial pivoting and row interchanges is available in the Lapack project [1].

A.2.3 Accurate computation of the Hessian matrix

When filling the coefficients of the Hessian matrix, we have interest to take care how expressions of Equation (A.8) are evaluated. For the evaluation of the term $F\left(x_{i+\frac{1}{2}}\right) - F\left(x_{i-\frac{1}{2}}\right)$, we can take advantage of the possible symmetry of the probability distribution. For example, if X has a symmetric probability density function, then $F(b) - F(a) = F(-a) - F(-b)$. This is the case for the standard normal univariate distribution. As floating point numbers have a constant relative precision on their range of definition (except for denormalized numbers), we always have interest to compute this difference on the side where the magnitude of the value of F is the smallest.
At this point, the accuracy of the estimation of the Hessian matrix relies on the accuracy of the evaluation of the cumulative distribution function F and the probability density function f.

A.3 Explicit Lloyd algorithm in the elliptic case

We now stand in the case where the probability density function f has the form $f(x) = g(x^2)$. (This is the case for an elliptic distribution.) We consider G a primitive of g.

A.3.1 Closed-form position of the centroids in the elliptic case

With these hypothesis

$$\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} x f(x) dx = \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} x g(x^2) dx = \frac{1}{2} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} 2x g(x^2) dx = \frac{1}{2}\left(G\left(x_{i+\frac{1}{2}}^2\right) - G\left(x_{i-\frac{1}{2}}^2\right)\right). \quad \text{(A.9)}$$

For instance, when X is distributed according to the standard univariate Gaussian distribution, we have $g(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t}{2}\right)$ and $G(t) = -\frac{2}{\sqrt{2\pi}} \exp\left(-\frac{t}{2}\right)$. Equation (A.9) becomes

$$\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = \frac{1}{\sqrt{2\pi}} \left(\exp\left(-\frac{x_{i-\frac{1}{2}}^2}{2}\right) - \exp\left(-\frac{x_{i+\frac{1}{2}}^2}{2}\right)\right).$$

The Lloyd algorithm takes a very simple form.

$$\begin{cases} x^{(0)} & = x, \\ x_i^{(n+1)} & = \frac{1}{2} \frac{1}{F\left(x_{i+\frac{1}{2}}^{(n)}\right) - F\left(x_{i-\frac{1}{2}}^{(n)}\right)} \left(G\left({x_{i+\frac{1}{2}}^{(n)}}^2\right) - G\left({x_{i-\frac{1}{2}}^{(n)}}^2\right)\right). \end{cases}$$

A.3.2 Accurate computation of the centroids

Here, the same remark as in Section A.2.3 holds: one can take advantage of the possible symmetry of the distribution for the computation of the term $F\left(x_{i+\frac{1}{2}}^{(n)}\right) - F\left(x_{i-\frac{1}{2}}^{(n)}\right)$.

A.3.3 Convergence of the algorithm

Du and Emelianeko proved in [7] that, in the one-dimensional case, for any starting point $x^{(0)}$, that the Lloyd algorithm converges globally. Moreover, it is well known [11] that in the case of a logarithmic concave density function, the Lloyd algorithm converges globally to a unique fixed point, which is the optimal quadratic codebook of the distribution.
As a consequence, the convergence of the algorithm toward an optimal quadratic codebook is guaranteed for the standard normal distribution, which is an elliptic distribution, with a log-concave probability density function. However, the Student t distribution with μ degrees of freedom, which is elliptic, does not have a log-concave probability density function, for any value of parameter μ. Thus, even though the convergence of the algorithm is guaranteed, the limit may depend of the starting point $x^{(0)}$.

A.4 Explicit Levenberg-Marquardt method in the Gaussian case

A.4.1 On the Levenberg-Marquardt algorithm

The Levenberg-Marquardt algorithm was initially developed in [12, 16] as a numerical method for minimizing a function of several variables. It is particularly adapted to the minimization of functions of the form

$$F(x) = \sum_{i=1}^{N} (y_i - f_i(x))^2 . \tag{A.10}$$

In vector notations, Equation (A.10) writes $F(x) = \|y - f(x)\|^2 = (y - f(x))^t (y - f(x))$, where $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ and $f = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix}$. Differentiating this expression gives

$$\frac{\partial F}{\partial x}(x) = -(y - f(x))^t J_f(x) - J_f^t(x)(y - f(x)) = -2J_f^t(x)(y - f(x)), \tag{A.11}$$

where J_f is the Jacobian matrix of f.

The optimization consists in an iterative procedure. At each step, the current approximation of the minimum argument x is replaced by $x+\delta$. Approximating $\frac{\partial F}{\partial x}(x+\delta)$ by $-2J_f^t(x)(y - f(x+\delta))$ and $f(x+\delta)$ by $f(x) + J_f(x)\delta$ in Equation (A.11) yields

$$\frac{\partial F}{\partial x}(x+\delta) = -2J_f^t(x)(y - f(x) - J_f(x)\delta).$$

Now, setting the left-hand side to zero yields $J_f^t J_f \delta = J_f^t(y - f(x))$. This linear system of equations can be solved numerically. (This corresponds to the Gauss-Newton iteration.) The idea of the Levenberg-Marquardt algorithm is to replace this equation by

$$\left(J_f^t J_f + \lambda \operatorname{diag}(J_f^t J_f)\right)\delta = J_f^t(y - f(x)), \tag{A.12}$$

for some $\lambda \geq 0$. The value of λ changes at each iteration. If it grows, the method is closer to the gradient algorithm, whereas when λ is close to zero, this comes to the Gauss-Newton algorithm. An usual choice for the sequence $(\lambda_n)_{n\in\mathbb{N}}$ is, to multiply or divide λ by a fixed value $v > 1$ depending on whether the new value of F is higher or smaller than the previous one. If $F(x+\delta) > F(x)$, $\lambda_{n+1} = v\lambda_n$ and if $F(x-\delta) < F(x)$, $\lambda_{n+1} = \frac{\lambda_n}{v}$.

A.4.2 Application to optimized quantization

We observe that the L^p quantization error has the form of function F in (A.10). Indeed, with the notations of Section A, with $\Gamma = \{\gamma_1, \cdots, \gamma_N\} \subset \mathbb{R}^d$,

$$\|X - \mathrm{Proj}_\Gamma(X)\|_p^p = \sum_{i=1}^N \mathbb{E}\left[|X-\gamma_i|^p \mathbf{1}_{X\in C_i}\right] = \sum_{i=1}^N f_i(\Gamma)^2,$$

with $f_i(\Gamma) := \sqrt{\mathbb{E}\left[|X-\gamma_i|^p \mathbf{1}_{X\in C_i}\right]}$.

Moreover, function f_i only depends on the elements of Γ which are connected to γ_i in the Delaunay triangulation of Γ (and γ_i itself). In the one-dimensional case, where Γ is the sorted sequence $\{x_1 < \cdots < x_N\}$, this comes to say that f_i only depends on x_{i-1}, x_i and x_{i+1}.

A.4.3 Computation of the Jacobian matrix J_f in the one-dimensional elliptic case

Denoting $x = (x_1, \cdots, x_N)$, we have $f_i(x) = \sqrt{\int_{x_{i-\frac12}}^{x_{i+\frac12}} (x_i-\xi)^2 f(\xi)d\xi}$.

Now, the non-zero partial derivatives are

- $\frac{\partial f_i}{\partial x_i}(x) = \frac{1}{2f_i(x)}\Bigg(\frac12 f\left(x_{i+\frac12}\right)\left(x_i - x_{i+\frac12}\right)^2 - \frac12 f\left(x_{i-\frac12}\right)\left(x_i - x_{i-\frac12}\right)^2$
$\qquad +2x_i\left(F\left(x_{i+\frac12}\right) - F\left(x_{i-\frac12}\right)\right) - \left(G\left(x_{i+\frac12}^2\right) - G\left(x_{i-\frac12}^2\right)\right)\Bigg)$,
- $\frac{\partial f_i}{\partial x_{i+1}}(x) = \frac{1}{2f_i(x)}\left(\frac12\left(x_i - x_{i+\frac12}\right)^2 f\left(x_{i+\frac12}\right)\right)$,
- $\frac{\partial f_i}{\partial x_{i-1}}(x) = \frac{1}{2f_i(x)}\left(\frac12\left(x_i - x_{i-\frac12}\right)^2 f\left(x_{i-\frac12}\right)\right)$.

We still need an explicit formula for the quantity $f_i(x)$.

$$\begin{aligned}
\int_{x_{i-\frac12}}^{x_{i+\frac12}} (x_i-\xi)^2 f(\xi)d\xi &= x_i^2 \int_{x_{i-\frac12}}^{x_{i+\frac12}} f(\xi)d\xi + \int_{x_{i-\frac12}}^{x_{i+\frac12}} \xi^2 f(\xi)d\xi - 2x_i \int_{x_{i-\frac12}}^{x_{i+\frac12}} \xi f(\xi)d\xi \\
&= x_i^2\left(F\left(x_{i+\frac12}\right) - F\left(x_{i-\frac12}\right)\right) - x_i\left(G\left(x_{i+\frac12}^2\right) - G\left(x_{i-\frac12}^2\right)\right) + \int_{x_{i-\frac12}}^{x_{i+\frac12}} \xi^2 f(\xi)d\xi \\
&= x_i^2\left(F\left(x_{i+\frac12}\right) - F\left(x_{i-\frac12}\right)\right) - x_i\left(G\left(x_{i+\frac12}^2\right) - G\left(x_{i-\frac12}^2\right)\right) \\
&\qquad + \frac12\left(G\left(x_{i+\frac12}^2\right)x_{i+\frac12} - G\left(x_{i-\frac12}^2\right)x_{i-\frac12}\right) - \frac12\int_{x_{i-\frac12}}^{x_{i+\frac12}} G(\xi^2)d\xi.
\end{aligned}$$

In the standard Gaussian case, *i.e.* when $f(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$, $F = \mathcal{N}$, $g(t) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t}{2}\right)$ and $G(t) = -\frac{2}{\sqrt{2\pi}}\exp\left(-\frac{t}{2}\right)$, we obtain

$$\begin{aligned}
\int_{x_{i-\frac12}}^{x_{i+\frac12}} (x_i-\xi)^2 f(\xi)d\xi = (x_i^2+1)\left(\mathcal{N}\left(x_{i+\frac12}\right) - \mathcal{N}\left(x_{i-\frac12}\right)\right) + \frac{2x_i}{\sqrt{2\pi}}\left(\exp\left(-\frac{x_{i+\frac12}^2}{2}\right) - \exp\left(-\frac{x_{i-\frac12}^2}{2}\right)\right) \\
- \frac{1}{\sqrt{2\pi}}\left(\exp\left(-\frac{x_{i+\frac12}^2}{2}\right)x_{i+\frac12} - \exp\left(-\frac{x_{i-\frac12}^2}{2}\right)x_{i-\frac12}\right). \qquad \text{(A.13)}
\end{aligned}$$

A.5 Starting point for optimization algorithms

In the general case of a $\mathbb{R}^d$-valued random variable X, a natural starting point for quadratic quantization optimization is a set of N independent copies of X. Still, in a more specific setting, one can take advantage on informations that we have on the distribution (symmetry, asymptotics of the distribution tails).

As concern the Gaussian distribution, the specification of Pagès and Sagna's results on the asymptotics of L^r-optimal quantizer radius [22] to the quadratic quantization of the standard one-dimensional Gaussian probability distribution provides an equivalent of the radius of an optimal quadratic quantizer.

$$\rho_N \underset{N\to\infty}{\sim} \sqrt{6\ln(N)}.$$

So that a natural simple starting point would be N equidistant abscissas between $-\sqrt{6\ln(N)}$ and $\sqrt{6\ln(N)}$.

However, numerical tests presented in [22] show that the convergence of $\frac{\rho_N}{\sqrt{6\ln(N)}}$ to 1 is very slow. (The ratio increases with N and is close to 0.86 for $N = 10000$).

A.6 One-dimensional Gaussian quantizer database

A.6.1 High precision computation

On the website `www.quantize.maths-fi.com` [21], we provide a sharply optimized database of Gaussian quantizers for a wide range of sizes and dimensions. This precomputed database allows a faster computation of (product) functional quantization of Gaussian processes [20, 5], and is also useful for other purposes [3]. The one-dimensional database provides 32 significant figures of the optimal quadratic quantization of the Gaussian distribution. The numerical computations used an implementation of multiple precision floating point numbers available in the ARPREC library [2], developed by Bailey *et al.* available at `http://crd.lbl.gov/~dhbailey/mpdist/`.
The ARPREC libraries make available basic operations for arbitrary precision floating point numbers together with an implementation of usual special functions. The Gaussian cumulative distribution function is computed using the Chiarella and Reichel formula [4].

$$\mathcal{N}(t) = -\frac{e^{-\frac{t^2}{2}}\alpha t}{\sqrt{2}\pi}\left(\frac{1}{t^2} + \sum_{k\geq 1}\frac{e^{-k^3\alpha^2}}{k^2\alpha^2 + t^2/2}\right) + \frac{1}{1 - e^{-\frac{\sqrt{2}\pi t}{\alpha}}} + E,$$

where $t < 0$ and $|E| < \frac{3}{2}e^{-\pi^2/\alpha^2}$. The parameter α is chosen small enough to ensure that the error E is sufficiently small. (Given an absolute precision of p digits, if $\alpha \leq \widetilde{\alpha}$ where $\widetilde{\alpha}$ is defined by $10^{-p} = 5e^{-\pi^2/\widetilde{\alpha}^2}$, we have $|E| < 10^{-p}$.)
Moreover, provided that $t > -\sqrt{2p\log 10}$, the formula is also accurate to a relative error of 10^{-p}.

A.6.2 The Gaussian database available on www.quantize.maths-fi.com

The files are in text format and is in the form of a matrix. In every case, filenames are `N_d_nopti` where N is the quantizer size and d is the dimension. For a given size N, the text files are organized as follows. It presents in the form of a matrix $G = (G_{i,j})$ with $N+1$ rows and $d+3$ columns.

- On row $i, i = 1, \cdots, N$: Element i of the grid and its companion parameters.

$$G_{i,1} = (\text{weight of the Voronoi cell } i) = \mathbb{P}[\mathcal{N}(0, I_d) \in C_i(G)].$$

$$\{G_{i,j}, j = 2, \cdots, d+2\} = (\text{coordinates of element } i)\,.$$

$$G_{i,d+2} = \Big(\text{conditional local } L^2 \text{ distorsion of the cell } i\Big) = \frac{\int_{C_i(G)} |z - G_{i,d}|^2 \mathcal{N}(0, I_d)(dz)}{G_i}.$$

$$G_{i,d+3} = \Big(\text{conditional local } L^1 \text{ distorsion of the cell } i\Big) = \frac{\int_{C_i(G)} |z - G_{i,d}| \mathcal{N}(0, I_d)(dz)}{G_i}.$$

- On last row ($i = N + 1$):

$$G_{N+1,j} = 0 \quad \text{for } j = 1, \cdots, d+1.$$

$$G_{N+1,d+2} = (\text{quadratic distorsion of the quantization grid}).$$

$$G_{N+1,d+2} = \Big(L^1 \text{ distorsion of the quantization grid}\Big).$$

- In particular we can verify that

$$\sum_{i=1}^{N} G_{i,1} = 1, \qquad \sum_{i=1}^{N} G_{i,1} G_{i,d+2} = G_{N+1,d+2}, \quad \text{and} \quad \sum_{i=1}^{N} G_{i,1} G_{i,d+3} = G_{N+1,d+3}.$$

A.6.3 A numerical example of optimal grid with 32 significant figures

In the following table, we provide numerical values for the points and the corresponding weights of a quadratic N-optimal quantizer of the univariate standard Gaussian distribution with $N = 20$. These numerical values have a relative precision of 10^{-32}.

Points	Weights
$-2.9079606783683841963010327564498190$	$4.7524748771175782543725573022952456 \times 10^{-3}$
$-2.2787139395025731289271610321795226$	$1.4574794203175608186209279080806336 \times 10^{-2}$
$-1.8569773889164496515754309935113130$	$2.6167977870684019356572298202272326 \times 10^{-2}$
$-1.5234142919930826826479444306240991$	$3.8153342170807815367544210261608918 \times 10^{-2}$
$-1.2384672136365404555283444318004377$	$4.9623785627087823397837755897103597 \times 10^{-2}$
$-0.9836424445338500002671635082479855 3$	$5.9948614410758147973469475590025768 \times 10^{-2}$
$-0.7485332893063160913785642774016703 6$	$6.8675125400963623653532164782940411 \times 10^{-2}$
$-0.5264881518717473396127020579235336 4$	$7.5478203095746321193432533376470300 \times 10^{-2}$
$-0.3127913756230818079470407729157103 3$	$8.0132035254861491696064717890529078 \times 10^{-2}$
$-0.1037625817938238686483686592721159 2$	$8.2493647088797570920965007615948020 \times 10^{-2}$
$0.1037625817938238686483686592721159 2$	$8.2493647088797570920965007615948020 \times 10^{-2}$
$0.3127913756230818079470407729157103 3$	$8.0132035254861491696064717890529078 \times 10^{-2}$
$0.5264881518717473396127020579235336 4$	$7.5478203095746321193432533376470300 \times 10^{-2}$
$0.7485332893063160913785642774016703 6$	$6.8675125400963623653532164782940411 \times 10^{-2}$
$0.9836424445338500002671635082479855 3$	$5.9948614410758147973469475590025768 \times 10^{-2}$
$1.2384672136365404555283444318004377$	$4.9623785627087823397837755897103597 \times 10^{-2}$
$1.5234142919930826826479444306240991$	$3.8153342170807815367544210261608918 \times 10^{-2}$
$1.8569773889164496515754309935113130$	$2.6167977870684019356572298202272326 \times 10^{-2}$
$2.2787139395025731289271610321795226$	$1.4574794203175608186209279080806336 \times 10^{-2}$
$2.9079606783683841963010327564498190$	$4.7524748771175782543725573022952456 \times 10^{-3}$

We notice that the reported values are exactly symmetric with respect to 0, as expected. More optimal quantization grids are available on the website `www.quantize.maths-fi.com` [21], in the form presented in Section A.6.2.

Bibliography

[1] Edward Anderson, Zhaojun Bai, Christian H. Bischof, Laura Susan Blackford, James W. Demmel, Jack J. Dongarra, Jeremy J. DuCroz, Anne Greenbaum, Sven Hammarling, Alan McKenney, and Danny C. Sorensen. *LAPACK User's Guide.* Society for Industrial and Applied Mathematics, 1999.

[2] David H. Bailey, Yozo Hida, Xiaoye S. Li, Brandon Thompson, Karthik Jeyabalan, and Alex Kaiser. The ARPREC libraries, 2010.

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

[4] Carl Chiarella and Alex Reichel. On the evaluation of integrals related to the error function. *Mathematics of Computation*, 22:137–143, 1968.

[5] Sylvain Corlay and Gilles Pagès. Functional quantization-based stratified sampling methods. *Preprint*, 2010.

[6] Qiang Du and Max Gunzburger. Grid generation and optimization based on centroidal Voronoi tessellations. *Applied Mathematics and Computation*, 133:591–607, 2002.

[7] Qiang Du, Maria Emelianenko, and Lili Ju. Convergence of the Lloyd algorithm for computing centroidal Voronoi tessellations. *SIAM J. Numer. Anal.*, 44(1):102–119, 2006.

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

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

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

[11] John C. Kieffer. Uniqueness of locally optimal quantizer for vector quantizer design. *IEEE Trans. Comm.*, 29:42–47, 1983.

[12] Kenneth Levenberg. A method for the solution of certain non-linear problems in least squares. *The Quarterly of Applied Mathematics*, 2:164–168, 1944.

[13] Stuart P. Lloyd. Least squares quantization in pcm. *Information Theory, IEEE Transactions on*, 28(2):129–137, 1982.

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

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

[16] Donald W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics*, 11:431–441, 1963.

[17] Moez Mrad and Sana Ben Hamida. Optimal quantization: evolutionary algorithm *vs.* stochastic gradient. *Proceedings of the 9th Joint Conference on Information Sciences*, 2006.

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

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

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

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

[22] Gilles Pagès and Abass Sagna. Asymptotics of the maximal radius of an L^r-optimal sequence of quantizers. *Bernoulli*, 2008.

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

[24] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical recipes in C++: The art of scientific computing.* Cambridge University Press, February 2002.

[25] Abass Sagna. Universal L^s-rate-optimality of L^r-optimal quantizers by dilatation and contraction. *Preprint*, 2009.

[26] Thaddeus Tarpey, Luning Li, and Bernard D. Flury. Principal points and self-consistent points of elliptical distributions. *Ann. Stat.*, 23(1):103–112, 1995.

[27] Desheng Wang and Qiang Du. Mesh optimization based on the centroidal Voronoi tesselation. *International Journal of Numerical Analysis and modeling*, 2:100–113, 2005.