Université de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :
Université Toulouse III Paul Sabatier (UT3 Paul Sabatier)

Discipline ou spécialité :

Mathématiques Appliquées

Présentée et soutenue par :
Luis Armando Salomón Hernández

le : mercredi 9 novembre 2011

Titre :

AUTOUR DE LA QUANTIFICATION FONCTIONNELLE DE PROCESSUS GAUSSIENS

Ecole doctorale :
Mathématiques Informatique Télécommunications (MITT)

Unité de recherche :
Équipe de Statistique et Probabilités

Directeur(s) de Thèse :
Jean-Claude FORT
Li-Vang LOZADA CHANG

Rapporteurs :

Bernard BERCU
Juan Antonio CUESTA-ALBERTOS

Membre(s) du jury :
Fabrice GAMBOA
Bernard BERCU
Juan Antonio CUESTA-ALBERTOS
Héctor de ARAZOZA RODRÍGUEZ
Jean-Claude FORT
Li-Vang LOZADA CHANG

UNIVERSITÉ TOULOUSE III - PAUL SABATIER

THÈSE

présentée pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE TOULOUSE
délivré par l'Université Toulouse III - Paul Sabatier

en Mathématiques Appliquées

présentée par

Luis Armando Salomón Hernández

intitulée

AUTOUR DE LA QUANTIFICATION FONCTIONNELLE DE PROCESSUS GAUSSIENS

Directeurs de thèse:
Jean-Claude Fort et Li-Vang Lozada Chang

Soutenue le 9 novembre 2011 devant le jury composé de Messieurs:

Fabrice Gamboa	Université Paul Sabatier	Président
Bernard Bercu	Université Bordeaux 1	Rapporteur
Juan Antonio Cuesta-Albertos	Université de Cantabria	Rapporteur
Héctor de Arazoza Rodríguez	Université de La Havane	Examinateur
Jean-Claude Fort	Université Paris Descartes-Paris V	Directeur
Li-Vang Lozada Chang	Université de La Havane	Directeur

Institut de Mathématiques de Toulouse
U.M.R. C.N.R.S. 5219, Université Paul Sabatier, Toulouse III

À mon père,
A mi padre,

Remerciements

Je voudrais tout d'abord exprimer la plus sincère reconnaissance et gratitude à mes directeurs Jean-Claude Fort et Li-Vang Lozada Chang pour toutes ces années de travail. Je suis très heureux d'avoir partagé tous les conseils et suggestions que vous m'avez donnés au cours de ces recherches. Je voudrais vous exprimer tous mes remerciements pour votre soutien et votre patience pendant ces années de thèse. A vous deux, merci beaucoup pour votre intérêt et votre professionnalisme, et spécialement pour la confiance constante que vous m'avez accordée.

Je tiens ensuite à remercier aussi toute l'Équipe de Statistique et Probabilités de l'Université Paul Sabatier de Toulouse pour son soutien et son attention, spécialement à Fabrice Gamboa, Jérémie Bigot, Thierry Klein et Jean-Michel Loubes. Merci à tous pour votre amabilité. Merci beaucoup en particulier à Fabrice, Thierry et Li- Vang d'avoir maintenu vivante la collaboration entre l'Université de La Havane et l'Université Paul Sabatier, qui a rendu ce possible.

Je voudrais remercier Égide et l'Équipe de Statistique et Probabilités de l'Université Paul Sabatier de Toulouse par lesquels cette recherche a été financée. Une reconnaissance particulière pour Françoise Michel, Agnès Requis et Marie-Laure Ausset pour leur aide sur tous les détails administratifs qui sont apparus durant la thèse.

Je voudrais remercier vivement mes rapporteurs Juan Antonio Cuesta-Albertos et Bernard Bercu, pour vous la plus sincère reconnaissance pour avoir accepté d'être les rapporteurs de la thèse. Merci pour le soin apporté à la lecture et à l'évaluation de ce manuscrit. Merci beaucoup aussi à Arazoza pour tout et pour le soutien moral que cela m'a donné. Je voudrais remercier spécialement Cuesta pour son intérêt et ses conseils au-delà du but de cette thèse.

Je veux aussi remercier tous mes collègues de la Faculté de Mathématiques et Calcul de l'Université de La Havane, spécialement tous les enseignants du département de Mathématiques Appliquées (Jorge, Sira, Marta Lourdes, Angela, Valdés, Nicolás, enfin tous), pour m'avoir beaucoup soutenu durant la réalisation de cette thèse. Ma reconnaissance la plus profonde à Vivian et Aymée dont sur lesquels j'ai toujours pu compter à chaque instant.

Je veux remercier d'une manière très spéciale Lilian Muñiz, pour une amitié que j'apprécie beaucoup, pour tous les moments que nous avons partagés à Toulouse. Son soutien a beaucoup signifié pour moi. A Michael Solís Chacón et Laura Murillo, pour toute l'affection et l'aide qu'ils m'ont offerte dans cette dernier étape; à Ricardo, pour être mon ami et par ses conseils et soutien jusqu'au dernière moment; à Ariadna par tous ces souvenirs rafraîchissants et agréables; à Mary Ana et Tony qui m'ont accompagné durant tout ce trajet; à Santiago et Karol pour l'aide inconditionnelle qu'ils m'ont toujours

donnée; à Alexander qui m'a montré Toulouse pour la première fois, *merci pour le café*; à Ignacio par son aide. Remerciements à Ariel, Pipo et Fanny pour son amitié et soutien depuis mon premier séjour à Toulouse; à Iván, Edmée et Lola pour m'accepter dans sa famille et pour toute l'affection qu'ils m'ont donné; à Thibaut pour son amitié et spécialement pour les discussions mathématiques, merci pour toute l'aide pour la rédaction en français, à toi et à Michael merci pour les précieux conseils que vous m'avez donnés pour la présentation de ce travail; je voudrais remercier, encore une fois Marie-Laure pour son attention et son soutien.

Simplement, je voudrais remercier tous les nouveaux amis que j'ai trouvés durant mon séjour en France, merci à tous pour les moments inoubliables que nous avons partagés à Toulouse.

À Christiane Fontas pour être l'une des personnes plus belles que j'ai connu en France et à celle qui je garde une affection très spéciale.

Je souhaite aussi saluer et remercier mes amis depuis des années, à mon groupe éternel de la Havana Vieja: German, Sergio, Carlitos, Wilfredo, Alejandro, Raupa, Marjorie et Maylen, qui plus que des amis sont la famille avec qui j'ai partagé les meilleures années de ma vie; à mes amis de l'université : Lissette, Yadira, Celia et Adriana qui ont vécu avec moi cette expérience; à mes amis Freddy, Alejandro, William, Gretel, Zaira, Leo, Lisi y Gladis qui ont toujours été intéressé par les processus stochastiques.

Je voudrais remercier toutes ces personnes qui se font maintenant partie de ma famille: à Esther pour son soutien éternel, et l'affection qu'elle m'a toujours donné; à Maria, Liset, Migue et Mariam pour être toujours à mon côté dans les moments difficiles; à Iris, Tania et Erick pour tous les détails et préoccupations qu'ils ont eues pour moi; à ma belle sœur Joanna et ma belle mère Victoria pour son soutien et amour.

Une reconnaissance très spéciale à ma famille: à ma mère pour son soutien inconditionnel; à mon père, pour ce qu'il signifie pour moi; à ma sœur Leipzig, qui a toujours été à mon côté, à ma sœur Anays, pour son intérêt et son soutien pendant ces années; à mon cousin Edu, sur qui j'ai toujours pu compter; à ma tante Delia, pour l'amour et l'affection qu'elle m'a toujours montrée; à mon frère Victor pour son intérêt et son affection, à mes nièces Alejandra et Victoria.

De tout mon cœur, je veux remercier spécialement mon épouse Arlette, pour être la personne qui a été à mon côté dans les moments les plus difficiles, pour son accompagnement, son amour et son soutien, qui m'a aidé à poursuivre pendant toutes ces années et sans laquelle cette thèse n'aurait pas pu aboutir.

Merci beaucoup à tous ceux qui m'ont aidé d'une façon ou d'une autre durant la réalisation de cette thèse!

En fin, el mar.

Agradecimientos

En primer lugar quisiera expresar mi más sincero agradecimiento y gratitud a mis tutores Jean-Claude Fort y Li-Vang Lozada Chang por todos estos años de trabajo. Estoy muy contento por haber compartido todos los consejos y sugerencias que me brindaron en el transcurso de la investigación. Quisiera expresarles mis agradecimientos por su apoyo y paciencia. A ambos, muchas gracias por su interés y profesionalidad y en especial por la confianza constante que depositaron en mí.

Muchas gracias a todo el Equipo de Estadística y Probabilidades de Toulouse por su apoyo y preocupación, en especial a Fabrice Gamboa, Jérémie Bigot, Thierry Klein y Jean-Michel Loubes. Gracias a todos por su amabilidad. Muchas gracias en particular a Fabrice, Thierry y Li-Vang por mantener viva la colaboración entre la Universidad de La Habana y la Universidad Paul Sabatier, gracias a los cuales ha sido posible la realización de esta tesis en cotutela entre ambas universidades.

Quisiera agradecer en especial a Égide y al Equipo de Estadística y Probabilidades de Toulouse gracias a los cuales esta investigación ha podido financiarse. Un agradecimiento de manera muy especial para Françoise Michel , Agnès Requis y Marie-Laure Ausset por su ayuda con todos los detalles administrativos que surgieron durante la tesis.

Quisiera agradecer a mis oponentes Juan Antonio Cuesta-Albertos y Bernard Bercu, mis más sinceros agradecimientos por aceptar ser los oponentes de la tesis. Gracias por la dedicación aportada a la lectura y a la evaluación de este manuscrito. Muchas gracias también a Arazoza por aceptar ser parte del tribunal y el apoyo moral que me brindó. En especial quisiera agradecer a Cuesta por su preocupación y todos los consejos más allá del alcance de esta tesis.

Muchísimas gracias también a todos mis colegas de la Facultad de Matemática y Computación de la Universidad de La Habana, en especial a todos los profesores del departamento de Matemática Aplicada (Jorge, Sira, Marta Lourdes, Angela, Valdés, Nicolás, en fin, a todos), por haberme apoyado mucho durante la realización de esta tesis. Mi más profundo agradecimiento a Vivian y Aymée con las que siempre he podido contar en todo momento.

Quiero agradecer de manera muy especial a Lilian Muñiz, por una amistad que valoro mucho y por todos los momentos que compartimos juntos en Toulouse. Su apoyo ha significado mucho para mí. A Michael Solís Chacón y Laura Murillo, por todo el cariño y la ayuda que me brindaron en esta última etapa; a Ricardo, por su amistad y los consejos y apoyo que me brindó hasta el último momento; a Ariadna por todos esos recuerdos refrescantes y agradables; a Mary Ana y Tony que me han acompañando durante todo este trayecto; a Santiago y Karol por la ayuda incondicional que me han dado siempre; a Alexander que me adentró en el mundo de Toulouse, *gracias por el café*; a Ignacio por su

ayuda. Gracias a Ariel, Pipo y Fanny por su amistad y apoyo desde mi primera estancia en Toulouse; a Iván, Edmée y Lola por aceptarme en su familia y por todo el cariño que me brindaron; a Thibaut por su amistad y en especial por los excelentes debates de matemática, gracias por toda la ayuda con la redacción en frances, a ti y a Michael gracias por los útiles consejos que me dieron para la presentación de este trabajo; a Marie-Laure quisiera agradecerle una vez más por su preocupación y apoyo.

Simplemente, quisiera agradecer a todos los nuevos amigos que he encontrado durante mi estancia en Francia, gracias a todos por los inolvidables momentos que hemos compartido en Toulouse.

A Christiane Fontas por ser una de las personas más bellas que he conocido en Francia y a la que le guardo un cariño muy especial.

Deseo saludar y agradecer a mis amigos de los años, a mi eterno grupo de la Havana Vieja: German, Sergio, Carlitos, Wilfredo, Alejandro, Raupa, Marjorie y Maylen, que más que amigos son la familia con la que he compartido los mejores años de mi vida; a mis amigos de la universidad: Lissette, Yadira, Adriana y Celia que han vivido conmigo esta experiencia; a mis amigos Freddy, Alejandro, William, Gretel, Zaira, Leo, Lisi y Gladis que siempre les han interesado los procesos estocásticos.

Quisiera agradecer a todas esas personas que se han convertido en parte de mi familia: a Esther por su eterno apoyo, y los cuidados que siempre me ha brindado; a Maria, Liset, Migue y Mariam por estar siempre a mi lado en los momentos difíciles; a Iris, Tania y Erick por todos los detalles y preocupaciones que han tenido conmigo; a mi familia política, en especial a Joanna y Victoria por todo su amor y apoyo.

Un agradecimiento muy especial para mi familia: a mi madre por todo el apoyo incondicional; a mi padre, por lo que significa para mí; a mis hermana Leipzig, que siempre ha estado a mi lado, a mi hermana Anays, por su interés y apoyo durante estos años; a mi primo Edu, con el cual siempre he podido contar; a mi tía Delia, por el amor y el cariño que siempre me ha mostrado; a mi *hermano* Victor por su interés y cariño, a mis sobrinas Alejandra y Victoria.

Quiero agradecer con todo mi corazón a mi esposa Arlette, por ser la persona que ha estado a mi lado en los momentos más difíciles, por su compañía y por todo ese amor y cariño que me ha ayudado a seguir adelante durante todos estos años y sin la cual esta tesis no se habría podido terminar.

Muchas gracias a todos los que me han ayudado de una forma u otra durante la realización de esta tesis!

En fin, el mar.

Contents

Résumé en français

Cette thèse a pour objectif principal l'étude de résultats asymptotiques autour de la quantification fonctionnelle. Après les résultats obtenus dans Sagna [2008] sur le rayon maximal du quantifieur optimal dans $\mathbb{R}^d$ nous cherchons l'asymptotique du rayon maximal en dimension infinie, spécifiquement pour le mouvement brownien. Nous présentons aussi un nouvel algorithme stochastique en dimension finie. Nous proposons une nouvelle méthode d'estimation pour le paramètre de Hurst dans des processus gaussiens fractionnaires plus robuste pour le calcul numérique que le maximum de vraisemblance en utilisant la décomposition de Karhunen-Loève des processus gaussiens.

La théorie de la quantification en dimension finie des vecteurs aléatoires commence au début des années 50. Elle est apparue pour la première fois en théorie du traitement du signal (pour discrétiser des émissions de signaux stationnaires). Depuis, elle s'applique en théorie de l'information. Par la suite, elle a trouvé d'autres domaines d'application comme la finance (l'estimation des prix de certaines options financières), l'intégration numérique (l'estimation d'une espérance conditionnelle) et plus récemment les probabilités numériques. Voir Graf and Luschgy [2000], Gersho and Gray [1992], Tarpey [1996], Bally et al. [2003], Tou and Gonzalez [1974], Diday and Simon [1976], Pagès et al. [2004] et Pagès [1998].

La quantification fonctionnelle des processus gaussiens permet de choisir un nombre fixé de trajectoires dans l'espace d'état du processus. Cette quantification en dimension infinie est étroitement liée à la décomposition de Karhunen-Loève du processus qui permet de distinguer les dépendances en t et en ω du processus.

Le premier travail sur la quantification fonctionnelle remonte à Luschgy and Pagès [2002]. Depuis on trouve une littérature abondante sur le sujet, le cas le plus étudié étant celui des processus gaussiens: Luschgy and Pagès [2004a], Dereich et al. [2003], Luschgy and Pagès [2002], Graf et al. [2003], Dereich [2005]. Pour les processus de Lévy et les processus de diffusion voir Luschgy and Pagès [2006] Pagès and Printems [2005], Wilbertz [2005], Dereich and Scheutzow [2006], Dereich and Lifshits [2005] et pour la première application au clustering de données fonctionnelles voir Tarpey and Kinateder [2003].

La première approche de l'asymptotique de l'erreur de quantification vient du travail de Luschgy and Pagès [2002]. Des bornes supérieures et inférieures de l'erreur de quantification sont données en utilisant le comportement des valeurs propres de l'opérateur de covariance et la ϵ-entropie de Shannon-Kolmogorov. Une approche différente basée sur les fonctions à variation régulière, les probabilités de petites boules et leur liaison avec l'erreur de quantification est proposée par Graf et al. [2003] et Dereich [2003]. La plupart des résultats importants pour l'asymptotique précise de l'erreur de quantification est détaillée dans Luschgy and Pagès [2004a,b].

Les algorithmes d'approximation stochastique ont été présentés au début des années 50 par Robbins et Monro. Ils ont élaboré et étudié une procédure récursive de recherche des zéros d'une fonction à valeurs réelles. Parmi les applications de cette théorie, nous pouvons mentionner les problèmes d'estimation de paramètres inconnus basés sur des données d'observations contenant l'information ou la recherche de maxima de certaines fonctions (voir Duflo [1996, 1997] par exemple). L'utilisation d'algorithmes stochastiques est la meilleure méthode pour obtenir numériquement des quantifieurs optimaux en dimension finie et infinie, dans ce dernier cas c'est presque l'unique façon de les calculer.

Nous avons organisé notre travail de la façon suivante: une introduction générale, une première partie de trois chapitres où on rappelle diverses conventions, définitions et les résultats classiques les plus importants qui seront utilisés par la suite. Les trois derniers chapitres sont consacrés à détailler les nouveaux résultats et applications dans le cadre de cette thèse.

Le chapitre 4 est consacré à l'étude de l'algorithme stochastique que nous proposons: le *Average Competitive Learning Vector Quantization* (ACLVQ). Cette nouvelle méthode utilise la méthode classique *Competitive Learning Vector Quantization* (CLVQ) avec une légère modification dans la *competitive phase* qui diminue le nombre d'itérations de la méthode sans sacrifier la précision des résultats. La convergence de cette nouvelle méthode est discutée. Nous effectuons une étude de simulation qui montre le comportement de ce nouvel algorithme.

Dans le chapitre 5, nous développons une nouvelle méthode d'estimation de paramètre du Hurst pour des processus gaussiens fractionnaires basée sur la décomposition de Karhunen-Loève du processus. Nous proposons un estimateur du paramètre de Hurst qui se comporte presque comme l'estimateur du maximum de vraisemblance mais qui est plus robuste pour le calcul numérique. Nous étudions les cas du mouvement brownien fractionnaire (fBm), la famille des processus fractionnaires d'Ornstein-Uhlenbeck (fOU) et le pont brownien fractionnaire (fBb). Nous comparons numériquement nos résultats à ceux obtenus par la méthode du maximum de vraisemblance.

Le dernier chapitre contient les résultats asymptotiques autour de l'erreur de quantification dans le cas fonctionnel. Dans la premiere partie nous fournissons l'asymptotique de l'erreur de quantification d'un processus gaussien avec la fonction de covariance donnée par $\Gamma(t) = e^{-\theta t^2}$. Notre idée repose sur le comportement asymptotique des valeurs propres associées à Γ obtenu en employant les travaux de Widom [1964] et Luschgy and Pagès [2002, 2004a]. Une application au krigeage est également discutée. Dans la deuxième partie nous limitons notre attention au comportement asymptotique du rayon maximal du quantifieur n-optimal pour processus de Wiener. La discussion est focalisée sur quelques quantifieurs stationnaires, l'interpolation linéaire du mouvement brownien et le processus lui-même.

Notions fondamentales

Processus gaussiens. Représentation de Karhunen Loève

Le cadre de cette thèse est celui des processus stochastiques. Nous travaillons avec le mouvement brownien et quelques processus fractionnaires (brownien fractionnaire,

Ornstein-Uhlenbeck, le pont brownien).

La plupart des résultats qu'on utilise dans le cadre de la quantification fonctionnelle dépendent de la représentation du processus dans la base de Karhunen-Loève. La décomposition de Karhunen-Loève permet de représenter une fonction aléatoire par une combinaison linéaire de fonctions déterministes dont les poids sont aléatoires.

Cette représentation est la plus importante dans le cadre de la quantification fonctionnelle. Elle est liée à la fonction d'auto-covariance du processus. Par la suite, et pour simplifier les notations on considère un processus X centré ($\mathbb{E}X = 0$).

Pour un processus gaussien centré $X \in \mathcal{L}^2$ sa fonction de covariance C_X est définie par

$$C_X(s,t) = cov(X(s), X(t)) = \mathbb{E}(X(t)X(s)),$$

et satisfait

$$C_X(s,t) = \sum_{k\geq 1} \lambda_k \varphi_k(t)\varphi_k(s),$$

où $\{\varphi_k\}_{k\geq 1}$ est une base orthonormale du $\mathcal{L}^2$ et $\lambda_k = \mathbb{E}\rho_k^2$ avec

$$\rho_k = \langle X(t), \varphi_k(t)\rangle = \int_0^1 X(t)\varphi_k(t)dt.$$

Le système $(\lambda_k, \varphi_k)_{k\geq 1}$ est la base de Karhunen-Loève du processus. L'unique décomposition de X dans cette base est

$$X(t) \overset{\mathcal{L}^2}{=} \sum_{k\geq 1} \sqrt{\lambda_k}\xi_k\varphi_k(t),$$

où $(\xi_k)_{k\geq 1}$ est une suite des variables aléatoires indépendants de loi $N(0,1)$. La suite $(\lambda_k)_{k\geq 1}$ est la suite décroissante des valeurs propres de l'opérateur de covariance de X définie par $\lambda_k = V(\langle X, \varphi_k\rangle)$ et qui satisfait

$$\sum_{k\geq 1} \lambda_k = \mathbb{E}||X||^2.$$

Quantification vectorielle

La théorie de la quantification générale est une méthode pour discrétiser l'espace des trajectoires d'un phénomène aléatoire:

- Dimension finie, des vecteurs aléatoires sur $\mathbb{R}^d$ (quantification vectorielle).
- Dimension infinie, des processus stochastique dans une espace de Hilbert comme par exemple $\mathcal{L}^2([0,1])$ (quantification fonctionnelle).

Autrement dit la quantification consiste à représenter un ensemble aléatoire continu, disons X, par un ensemble $\widehat{X}$ fini. La quantification vectorielle d'une variable aléatoire ou d'un processus consiste à leur associer une autre variable aléatoire ou processus à valeurs dans le même espace mais ne prenant qu'un nombre fini de valeurs: sa version quantifiée.

En général, le quantifieur est construit à partir du vecteur aléatoire ou processus original par sa projection au plus proche voisin. On dit que la quantification est optimale

si pour un nombre de points fixé, la disposition des points du quantifieur minimise l'erreur quadratique, d'ordre r plus généralement.

Du point de vue mathématique, on se place dans un espace $(\Omega, \mathcal{F}, \mathbb{P})$ et on se donne un vecteur aléatoire X de loi $\mathbb{P}_X$ à valeurs dans $\mathbb{R}^d$. On suppose que X admet un moment d'ordre fini avec $r > 1$. Étant donné un entier $n \geq 1$, un n-quantifieur $\widehat{X}$ est défini par une application borélienne $f : \mathbb{R}^d \to \alpha \subset \mathbb{R}^d$ où $|\alpha| \leq n$. La meilleure approximation possible de X est associée à la fonction borélienne qui minimise l'erreur de quantification définie par

$$\begin{aligned} e_{n,r}(X) &= \inf_{f \in \mathcal{F}_n} (\mathbb{E}||X - f(X)||^r)^{1/r} = \inf_{\substack{\alpha \subset \mathbb{R}^d \\ |\alpha| \leq n}} (\mathbb{E} \min_{a \in \alpha} ||X - a||^r)^{1/r} \\ &= \inf \left\{ (\mathbb{E}||X - \widehat{X}^\alpha||^r)^{1/r} : \alpha \subset \mathbb{R}^d,\ |\alpha| \leq n \right\}. \end{aligned}$$

où le vecteur aléatoire

$$\widehat{X}^\alpha = \sum_{i=1}^{n} x_i \mathbf{1}_{C_i}(X),$$

est la projection au plus proche voisin sur $\alpha = \{x_1, \cdots, x_n\}$ et est souvent appelée la quantification de X ou un n-quantifieur optimal associé à X (en l'absence d'ambiguïté on omettra parfois l'exposant α). L'ensemble $(C_i)_{1 \leq i \leq n}$ est une partition de Voronoï associée à la grille α, c'est-à-dire une partition borélienne de $\mathbb{R}^d$ vérifiant pour tout $x_i \in \alpha$

$$C_i \subset \{x \in \mathbb{R}^d : ||x - x_i|| = \min_{b \in \alpha} ||x - b||\}.$$

L'erreur de quantification est une suite décroissante qui converge vers zéro quand la taille du quantifieur n tend vers $+\infty$. La vitesse de convergence est fournie par le Théorème de Zador énoncé comme suit

***Théorème 1** Taux asymptotique*
Soit $r > 0$ et on suppose que $\int_{\mathbb{R}^d} |\xi|^{\xi+\eta} \mathbb{P}(d\xi) < +\infty$ pour $\eta > 0$. Soit $f \stackrel{\triangle}{=} d\mathbb{P}/d\lambda_d$ (λ_d la mesure de Lebesgue), alors

$$\lim_n \left(n^{\frac{r}{d}} \min_{(\mathbb{R}^d)^n} e^r_{n,r}(X) \right) = J_{r,d} ||f||_{\frac{d}{d+r}} < +\infty,$$

où

$$||g||_p \stackrel{\triangle}{=} \left(\int_{\mathbb{R}^d} |g(\xi)|^p \mathbb{P}(d\xi) \right)^{1/p} \quad \textit{pour tout } p \in (0, +\infty).$$

La constante $J_{r,d}$ correspond à la limite pour la loi uniforme sur $[0,1]^d$. On sait que $J_{r,1} = 1/(2^r(r+1))$, $J_{2,2} = 5/(18\sqrt{3})$. Dans le cas $d \geq 3$ on ne connaît pas la valeur de cette constante $J_{r,d}$. On sait qu'il existe des asymptotiques: $J_{r,d} = d/(2\pi e)^{r/2} + o(d)$ quand $d \to +\infty$.

Pour plus de détails voir Graf and Luschgy [2000].

Quantification fonctionnelle

En général la quantification fonctionnelle étudie la meilleure approximation d'un processus stochastique X dans l'espace d'état du processus. Le plus souvent on étudie le cas d'un espace de Hilbert $H = \mathcal{L}^2$, mais d'autres choix sont possibles comme par exemple $\mathcal{L}^p([0,1], dt)$ et $\mathcal{C}([0,1], \mathbb{R})$.

Formellement, la présentation est la même que dans le cadre fini. Du point de vue mathématique, on se place dans un espace $(H, \langle \cdot, \cdot \rangle_H)$ qui satisfait que $\mathbb{E}||X||^2 < +\infty$. Dans le cas fonctionnel, on travaille dans l'espace $H = \mathcal{L}^2([0,1], dt)$ avec la norme usuelle $||X||^2_{\mathcal{L}^2} = \int_0^1 X^2(t)dt$.

Soit X une vecteur aléatoire dans l'espace $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans H

$$X : (\Omega, \mathcal{A}, \mathbb{P}) \longmapsto (H, \langle \cdot, \cdot \rangle).$$

Pour $n \in \mathbb{N}$, le problème de quantification pour X est liée à la minimisation de

$$\left(\mathbb{E} \min_{a \in \alpha} ||X - a||^r \right)^{1/r},$$

sur tous les ensembles $\alpha \subset H$ avec $|\alpha| \leq n$. L'ensemble α s'appelle n-quantifieur (n-*codebook* dans la littérature en compression d'information).

L'erreur de quantification $e_n^r(X)$ est définie par

$$\begin{aligned} e_n^r(X) &= \inf_{\substack{\alpha \subset H \\ |\alpha| \leq n}} \mathbb{E} \min_{a \in \alpha} ||X - a||^r \\ &= \inf \left\{ \mathbb{E} \min_{a \in \alpha} ||X - a||^r : \alpha \subset H,\ |\alpha| \leq n \right\}. \end{aligned}$$

Quand

$$e_n^r(X) = \mathbb{E} \min_{a \in \alpha} ||X - a||^r,$$

on dit que α est un quantifieur n-optimal de X.

La projection au plus proche voisin $\widehat{X}^\alpha : H \to \alpha \subset H$ pour X est

$$\widehat{X}^\alpha = \sum_{i=1}^{n} x_i \mathbf{1}_{C_i}(X),$$

où l'ensemble $C_i \stackrel{\triangle}{=} C_\alpha(x_i)$ est une partition de Voronoï associée à α.

$$C_i(X) \subset V_\alpha(x_i) \stackrel{\triangle}{=} \{x \in H : ||x - x_i|| = \min_{b \in \alpha} ||x - b||\}.$$

L'erreur de quantification peut s'écrire

$$e_n^r(X) = \inf \left\{ \mathbb{E}||X - \widehat{X}^\alpha||^r : \widehat{X}^\alpha : \Omega \to H, \text{ vecteur aléatoire}, |\widehat{X}^\alpha(\Omega)| \leq n \right\}.$$

On va se placer maintenant dans le cas $r = 2$. Dans le cadre de la quantification fonctionnelle il faut mentioner quelques résultats importants.

Soit $\mathfrak{C}_n(X)$ l'ensemble de tous les quantifieurs n-optimaux et C_a^0 la region de Voronoï ouverte associée à $a \in \alpha$. Pour $\alpha \in \mathfrak{C}_n(X)$ et $|\operatorname{supp}(\mathbb{P})| \geq n$, on a que $|\alpha| = n$, $\min_{a\in\alpha} \mathbb{P}(C_a^0) > 0$ et

1. Pour tout $x_i \in \alpha$,
$$x_i = \mathbb{E}(X|X \in C_i). \tag{1}$$
2. Pour tout $x_i, x_j \in \alpha$, $x_i \neq x_j$
$$\mathbb{P}(C_i \cap C_j) = 0. \tag{2}$$

La notion de *quantifieur stationnaire* est très importante: c'est un ensemble $\alpha \subset H$ qui satisfait que $|\alpha| = n$, $\min_{a\in\alpha} \mathbb{P}(C_a^0) > 0$, (1) et (2). Alors

***Lemme* 1**
Si $\alpha \subset H$ est un quantifieur n-stationnaire pour X alors

1. $\alpha \subset \operatorname{cl}\operatorname{conv}(\operatorname{supp}\mathbb{P})$.
2. *Si $\mathbb{E}(X) = 0$ alors α est dans l'espace de Cameron-Martin K_X.*

L'espace de Cameron-Martin ou *reproducing kernel Hilbert space* est un espace vectoriel de fonctions défini par

$$K_X \stackrel{\triangle}{=} \left\{\mathbb{E}(ZX) : Z \in \operatorname{cl}_{L^2(\mathbb{P})}\{\langle y, X\rangle : y \in H\}\right\}.$$

Il est clair que tout quantifieur n-optimal est n-stationnaire.

Un résultat très important est le suivant, qui permet de réduire un problème de quantification en dimension infinie à un problème de dimension fini.

***Théorème* 2**
Soit U un sous-espace vectoriel fini de H, soit Π_U la projection orthogonale de H sur U et $\alpha \subset U$, alors

1. $\alpha \in \mathfrak{C}_n(X)$.
2. $\alpha \in \mathfrak{C}_n(\Pi_U(X))$ *et* $e_n^2(X) = \mathbb{E}||X - \Pi_U(X)||^2 + e_n^2(\Pi_U(X))$.

Soit $\mathbf{u} = \{u_1, \cdots, u_m\}$ un sous-ensemble orthonormal de H, $U = \operatorname{span}\{u_1, \cdots, u_m\}$ (où $\operatorname{span}(\mathbf{u})$ dénote le sous-espace de H engendré par $\mathbf{u}$ et $Z = (\langle u_1, X\rangle, \cdots, \langle u_m, X\rangle)$. Soit $T : U \to \mathbb{R}^m$ une isométrie linéaire bijective telle que $T(u_j) = v_j$ pour $j = 1, \cdots, m$ où $\{v_1, \cdots, v_m\}$ est la base standard de $\mathbb{R}^m$, alors:

$$e_n(\Pi_U(X)) = e_n(Z). \tag{3}$$

Pour des processus gaussiens, on a le

***Théorème* 3**
Soit $\alpha \subset H$ un quantifieur n-stationnaire de X, soit Γ_X l'opérateur de covariance du processus et $U = \operatorname{span}(\alpha)$ alors $\Pi_U(X)$ et $X - \Pi_U(X)$ sont indépendants, $\Gamma_X(U) = U$ et $\alpha \subset \Gamma_X(H) \subset K_X$.

Si nous organisons les valeurs propres de l'opérateur de covariance Γ_X en ordre décroissant et nous utilisons le fait que $\sum_{k\geq 1} \lambda_k = \mathbb{E}||X||^2$ alors il vient:

***Théorème* 4**
Soit $\alpha \in \mathfrak{C}_n(X)$, $U = \text{span}(\alpha)$ *et* $m = \dim(U)$. *Alors* $\Gamma_X(U) = U$ *et*

$$\mathbb{E}||X - \Pi_U(X)||^2 = \sum_{k \geq m+1} \lambda_k.$$

Ce théorème assure que les sous-espaces m-dimensionnels de H engendrés par des ensembles de quantifieurs m-optimaux sont engendrés par les vecteurs propres de Γ_X des m plus grandes valeurs propres. Notons également que

$$\sum_{k \geq m+1} \lambda_k = \inf \left\{ \mathbb{E}||X - \Pi_V(X)||^2 : \ V \subset H \text{ sous-espace vectoriel }, \dim(V) = m \right\}.$$

Puisque le problème infini dimensionnel de quantification n'est pas soluble par un algorithme fini d'ordinateur, il est nécessaire de réduire la dimension du problème. Si α est un quantifieur n-optimal pour $\bigotimes_{j=1}^{\infty} \mathcal{N}(0, \lambda_j)$ alors $U = \text{span}(\alpha)$ est un sous-espace de l^2. Par le Théorème 3 et (3) on a

$$e_n^2 \left(\bigotimes_{k=1}^{\infty} \mathcal{N}(0, \lambda_k) \right) = e_n^2 \left(\bigotimes_{k=1}^{d_n} \mathcal{N}(0, \lambda_k) \right) + \sum_{k \geq d_n + 1} \lambda_k,$$

où d_n est la dimension de U. En fait

$$\begin{aligned} e_n^2(X) &= \sum_{j \geq m+1} \lambda_j + e_n^2 \left(\bigotimes_{j=1}^{m} \mathcal{N}(0, \lambda_j) \right) \quad \forall\, m \geq\, d_n \\ e_n^2(X) &< \sum_{j \geq m+1} \lambda_j + e_n^2 \left(\bigotimes_{j=1}^{m} \mathcal{N}(0, \lambda_j) \right) \quad \forall\, m <\, d_n. \end{aligned}$$

Le nombre entier (d_n) est défini comme $d_n = \min\{\dim \text{span}(\alpha) : \alpha \in \mathfrak{C}_n\}$ et il satisfait

$$\frac{1}{b^{1/(b-1)}} \frac{2 \ln n}{b} \lesssim d_n \leq n - 1.$$

La borne supérieure vient de

$$\begin{aligned} d_n \leq \overline{d}_n &\triangleq \max\{\dim \text{span}(\alpha) : \alpha \in \mathfrak{C}_n\} \\ &\leq \max\{\dim \text{span}(\alpha) : \alpha \ n\text{-stationnaire}\} \\ &\leq n - 1. \end{aligned}$$

La borne inférieure a été obtenue par Luschgy et al. [2010] où b est définie comme dans le Théorème 2.2 dans Luschgy and Pagès [2004a] si on suppose les mêmes conditions pour le processus gaussien X. Par exemple on pourra se référer aux articles Luschgy and Pagès [2002], Luschgy et al. [2010]. Jusqu'à présent le taux de décroissance de d_n est inconnu.

Schémas pour les quantifieurs des processus gaussiens

Le problème de quantification fonctionnelle optimale quadratique des processus gaussiens centrés est réduit à un problème optimal fini dimensionnel de quantification pour une distribution gaussienne avec une structure diagonale de covariance (voir Luschgy et al. [2010]). Dans cet article les auteurs décrivent quatre schémas de calcul basés sur la décomposition de Karhunen-Loève des processus gaussiens. Nous les rappelons brièvement.

Premier schéma de quantification (optimal)

Les résultats précédents pour d_n nous conduisent à décomposer la recherche d'un ensemble de quantifieurs optimaux en deux étapes (si possible). D'abord évaluer théoriquement ou numériquement les valeurs propres et les vecteurs propres dans la décomposition de Karhunen-Loève de l'opérateur de covariance. En second lieu, trouver d tel que $\sum_{k\geq d+1}\lambda_k$ soit petit par rapport à $e_n^2\left(\bigotimes_{k=1}^{\infty}\mathcal{N}(0,\lambda_k)\right)$ et calculer le quantifieur $\bigotimes_{k=1}^{d}\mathcal{N}(0,\lambda_k)$-optimal $\alpha^d\subset\mathbb{R}^d$ avec $|\alpha^d|\leq n$. Ce problème d'optimisation peut être lu comme

$$(\mathcal{O}_n) \;\triangleq\; \left\{ e_n^2\left(\bigotimes_{k=1}^{d_n}\mathcal{N}(0,\lambda_k)\right)+\sum_{k\geq d_n+1}\lambda_k\,.\right.$$

Évidemment $(\mathcal{O}_n)$ est un problème d'optimisation fini dimensionnel. Il dépend fortement du comportement des valeurs propres du processus X. En raison de leur forte décroissance pour presque tous les processus gaussiens classiques, ce problème d'optimisation est mal conditionné. On peut éviter cet écueil si $(\mathcal{O}_n)$ est écrit comme un problème d'optimisation pour la distribution normale standard $N(0,I_{d_n})$, mais en changeant la norme.

$$(\mathcal{O}_n) \;\Leftrightarrow\; \left\{\begin{array}{l} \text{quantification } n\text{-optimal de } \bigotimes_{k=1}^{d_n}\mathcal{N}(0,1) \\ \\ \text{pour la norme de covariance } |z_1,\cdots,z_{d_n}|^2=\sum_{k=1}^{d_n}\lambda_k z_k^2. \end{array}\right.$$

Cette approche est traitée dans Pagès and Printems [2005]. Les trois schémas suivants sont basés sur des quantifieurs produit qui sont stationnaires. Ces schémas ont également l'avantage d'être moins exigeants en mémoire.

Deuxième schéma de quantification

Quantification $\bigotimes_{j=k_i+1}^{k_{i+1}}\mathcal{N}(0,\lambda_j)$-optimal $\alpha^{(i)}\subset\mathbb{R}^{l_i}$ avec $|\alpha^{(i)}|\leq n_i$ pour des entiers $m\in\mathbb{N}$, $l_1,\cdots,l_m\leq l$, $n_1,\cdots,n_m>1$ et $\prod_{i=1}^m n_i\leq n$.

Troisième schéma de quantification

Quantification $\bigotimes_{j=k_i+1}^{k_{i+1}}\mathcal{N}(0,1)$-optimal $\alpha^{(i)}\subset\mathbb{R}^{l_i}$ avec $|\alpha^{(i)}|\leq n_i$ pour des entiers $m\in\mathbb{N}$, $l_1,\cdots,l_m\leq l$, $n_1,\cdots,n_m>1$ et $\prod_{i=1}^m n_i\leq n$.

Quatrième schéma de quantification

Quantification $\mathcal{N}(0,1)$-optimal $\alpha_i \subset \mathbb{R}$ avec $|\alpha_i| \leq n_i$ pour des entiers $n_1, \cdots, n_m > 1$, $m \in \mathbb{N}$ et $\prod_{i=1}^m n_i \leq n$.

Une discussion approfondie de ces schémas se trouve dans Luschgy et al. [2010]. Ils fournissent quelques propriétés du comportement de ces schémas pour le mouvement brownien.

Taux asymptotique

Le taux asymptotique de l'erreur de quantification de e_n^2 a été étudié par plusieurs auteurs, la référence est Luschgy and Pagès [2002]. Pour plus de détails voir également Luschgy and Pagès [2004a], Dereich [2003] et leurs références. Dans le cadre infini, pour des processus gaussiens, on dispose d'une généralisation du Théorème 1.

Théorème 5 Taux asymptotique
Soit X un processus gaussien centré à valeurs dans un espace de Hilbert de loi $\mathbb{P}$, avec opérateur de covariance Γ_C, et $(u_j^X)_{j\geq 1}$ un sous-ensemble orthonormal de H, tel que $K_X \subset \operatorname{cl}\operatorname{span}\{u_j^X, j \in \mathbb{N}\}$. On définit

$$\mu_j = \operatorname{Var}\langle u_j^X, X\rangle = \langle u_j^X, \Gamma_C(u_j^X)\rangle \quad et \quad \Sigma_m = (\langle u_j^X, \Gamma_C(u_k^X)\rangle)_{0\leq j,k\leq m},$$

et pour $n \in \mathbb{N}$, on pose

$$g_n(m) = e_n(\mathcal{N}(0, \Sigma_m)).$$

Supposons maintenant que $\{u_j^X :\subset \operatorname{cl}(K_X)\}$. Alors $\det \Sigma_m > 0$ et

$$\lim_{n\to\infty} n^{1/m} g_n(m) = Q(m) \quad pour\ tout \quad m \geq 1,$$

où $Q(m) \in (0, +\infty)$ et

$$Q(m) \sim \left(m(\det \Sigma_m)^{1/m}\right)^{1/2} \quad p.s \quad m \to \infty.$$

En particulier

$$\lim_{m\to\infty} Q(m) = 0.$$

Le théorème précédent est tout à fait général. Des asymptotiques plus spécifiques pour l'erreur de quantification existent pour une grande variété de processus gaussiens.

La première approche du calcul de l'asymptotique de l'erreur de quantification est due à Luschgy and Pagès [2002]. Ils donnent des bornes supérieures et inférieures pour e_n en utilisant le comportement des valeurs propres de l'opérateur de covariance et de la ϵ-entropie de Shannon-Kolmogorov (voir Théorèmes 4.6 et 4.12 dans Luschgy and Pagès [2002]).

Dans Graf et al. [2003] et Dereich [2003] on peut trouver une approche différente utilisant des fonctions à variation régulière, les probabilités des petites boules et leur liaison avec l'erreur de quantification, voir par exemple Théorème 1.2 dans Graf et al.

[2003] et Théorème 2.3 dans Dereich [2003]. Des bornes plus précises sont données pour une grande classe de processus gaussiens (voir aussi Dereich et al. [2003]).

La plupart des résultats importants pour l'asymptotique précise de l'erreur de quantification sont détaillés dans Luschgy and Pagès [2004a,b]. L'idée générale est principalement basée sur les fonctions à variation régulières et la ϵ-entropie de Shannon-Kolmogorov. Nous rappelons ici le Théorème 2.2 (a) de Luschgy and Pagès [2004a] qui fournit le taux précis de convergence de l'erreur de quantification (distorsion en compression d'information).

***Théorème* 6**
Soit X un processus gaussien avec valeurs propres $\lambda_j \sim \varphi(j)$ quand $j \to \infty$, où $\varphi : (s, \infty) \to (0, \infty)$ est une fonction décroissante à variation régulière à l'infini d'index $b > 1$ avec $s \geq 0$. Posons pour chaque $x > s$,

$$\psi(x) \triangleq \frac{1}{x\varphi(x)},$$

alors

$$e_n(X) \sim \left(\left(\frac{b}{2} \right)^{b-1} \frac{b}{b-1} \right)^{1/2} \psi(\ln n)^{-1/2}. \tag{4}$$

La forme la plus répandue pour φ est

$$\varphi(x) = cx^{-b}(\ln x)^{-a}, \quad b > 1,\ a \in \mathbb{R},\ x > \max\{1, e^{-a/b}\},$$

et (4) devient

$$e_n(X) \sim \sqrt{c} \left(\left(\frac{b}{2} \right)^{b-1} \frac{b}{b-1} \right)^{1/2} (\ln n)^{-(b-1)/2} (\ln \ln n)^{-a/2}.$$

Pour le cas particulier des diffusions, la ϵ-entropie de Shannon-Kolmogorov joue un rôle fondamental, voir par exemple Luschgy and Pagès [2006], Dereich and Scheutzow [2006] et Dereich [2008]. Également des asymptotique sont données pour le mouvement brownien d-dimensionnel. Le théorème précédent montre le rôle des valeurs propres de l'opérateur de covariance. Les théorèmes qui fournissent l'asymptotique des valeurs propres proviennent des travaux de Widom et de Rosenblatt (voir Théorème 1 dans Widom [1964] et Théorème 3 dans Rosenblatt [1963]).

Algorithmes stochastiques

Les origines des algorithmes stochastiques récursifs remontent au début des années 50 dans le travail de Robbins et Monro, et également dans les travaux de Kiefer et Wolfowitz. Ils ont élaboré les procédures récursives pour trouver les zéros d'une fonction à valeurs réelles, fonction inconnue mais dont on observe des valeurs bruitées.

Une des premières applications des algorithmes stochastiques est celle d'un problème de dosage en chimie, liée à l'idée de trouver la dose x^* qui produit, à un niveau donné α, un

effet moyen $f(x) = \mathbb{E}(F(x, \epsilon))$ où ϵ est une variable aléatoire et f une fonction inconnue. Le problème est de résoudre l'équation $f(x^*) = \alpha$. L'approche de Kiefer Wolfowitz est analogue à la méthode proposée par Robbins et Monro, avec emploi de différences finies pour localiser des extrema d'une fonction inconnue (voir par exemple Duflo [1997] pour plus de détails).

La méthode du gradient stochastique (i.e. une méthode qui utilise une version bruitée du gradient pour rechercher un extremum) est basée sur la représentation intégrale du critère à optimiser et peut être vue comme cas particulier de l'algorithme de Robbins-Monro lorsque la fonction a un unique minimum, et que la fonction et son espace de définition sont convexes.

La formulation classique pour les algorithmes stochastiques est donnée par

$$X_{k+1} = X_k + \gamma_{k+1} F(X_k, \omega_{k+1}), \tag{5}$$

où $(\omega_k)_{k\geq 1}$ est une suite de variables aléatoires à valeurs réelles, indépendant et identiquement distribuées avec variance finie, $(X_k)_{k\geq 1}$ est une suite d'un espace euclidien, $(\gamma_k)_{k\geq 1}$ est une suite qui tend vers zéro quand k tend vers l'infini, (dans des problèmes plus spécifiques quelques autres propriétés sont exigées pour cette suite) et F est une fonction qui prend une forme différente pour chaque problème particulier.

Le schéma de base des algorithmes stochastiques est une version bruité du schéma de convergence suivant. Soit $f \subset \mathcal{O} \subset \mathbb{R}^d \rightsquigarrow \mathbb{R}^d$ une fonction continue sur l'ensemble ouvert $\mathcal{O}$. Supposons qu'il existe un point unique $X^* \in \mathcal{O}$ tel que $f(X^*) = a$, où a est une valeur connue. Si X^* satisfait $\langle X - X^*, f(X) - a\rangle < 0$ pour tout $X \neq X^*$ alors

$$X_{k+1} = X_k + \gamma_{k+1}(f(X_k) - a),$$

où $\sum \gamma_k = +\infty$ et $\gamma_k \searrow 0$. Quand X^* satisfait $\langle X - X^*, f(X) - a\rangle > 0$ pour tout $X \neq X^*$, alors si on pose

$$X_{k+1} = X_k - \gamma_{k+1}(f(X_k) - a),$$

X_k converge vers X^*.

Nous nous intéressons spécialement au comportement asymptotique des algorithmes du gradient stochastique (5) qui s'écrivent:

$$X_{k+1} = X_k - \gamma_{k+1} dG(X_k, \xi_{k+1}),$$

On dispose d'un théorème qui assure la convergence de ces algorithmes :

Théorème 7 Convergence p.s.
Soit $g : E \to \mathbb{R}_+$ une fonction continuement différentiable dont le différentiel dg admet une représentation intégrale sur E par rapport à $\mathbb{P}$

$$dg(x) = \int_{\mathbb{R}^d} dG(x, \xi)\mathbb{P}(d\xi).$$

On suppose que dg et dG satisfont

$$\lim_{|x|\to+\infty} g(x) = +\infty \text{ et } dg \text{ est Lipschitz continu},$$

$$dg(x) = O(g(x)),\ quand\ |x| \to +\infty.$$

Soit $((X_k)_{k\geq 0}, (\xi_k)_{k\geq 1}, (\gamma_k)_{k\geq 1})$ *un algorithme stochastique de gradient*

$$X_{k+1} = X_k - \gamma_{k+1} dG(X_k, \xi_{k+1}),$$

où le pas *satisfait*

$$\sum_{k\geq 1} \gamma_k = +\infty, \qquad \sum_{k\geq 1} \gamma_k^2 < +\infty.$$

Alors $g(X_k)$ *converge p.s. vers une variable aléatoire* $g_\infty \in \mathbb{R}_+$ *non négative et* X_k *converge p.s vers une composante connexe* χ^* *de* $\{dg = 0\} \cap \{g = g_\infty\}$. *En particulier, si* $\{dg = 0\} = \{x^*\}$, *alors*

$$X_k \to x^*\ p.s.\ quand\ k \to +\infty.$$

Plus généralement le théorème de Kushner-Clark a été employé pour décrire le comportement des algorithmes stochastiques. C'est la *méthode de l'équation différentielle ordinaire* présentée dans Kushner and Clark [1978]. Les résultats les plus importants de convergence *p.s.* des algorithmes stochastiques en découlent. L'idée principale consiste à comparer le comportement de la solution de l'EDO ($\dot{x} = -h(x)$) et le comportement asymptotique de chaque trajectoire de l'algorithme stochastique.

Le théorème de Kushner-Clark est un théorème de convergence conditionnelle. Il identifie les limites possibles en tant que points d'équilibre attractifs de l'EDO associée (voir Kushner and Clark [1978]).

Donc si on réécrit l'équation générale (5) comme

$$X_{k+1} = X_k - \gamma_{k+1} h(X_k) + \gamma_{k+1}(\Delta M_{k+1} + \eta_{k+1}), \tag{6}$$

où $h(X_k) = \mathbb{E}[H(X_k, \xi_{k+1})]$ avec $H(X_k, \xi_{k+1}) = -F(X_k, \xi_{k+1}) + \eta_{k+1}$, $\Delta M_{k+1} = h(X_k) - H(X_k, \xi_{k+1})$ et η_k est un *petit* bruit, on a:

***Théorème* 8**
Soit x^* *un point d'équilibre stable de l'EDO* $\dot{x} = -h(x)$, *où* h *est une fonction continue. Soit* G_{x^*} *un voisinage de* x^* *inclus dans son bassin d'attraction. Si* $(X_k)_{k\geq 1}$ *dans (6) est une suite bornée,* $\sum_{k\geq 1} \gamma_k = +\infty$, $\sum_{k\geq 1} \gamma_k^2 < +\infty$ *et*

$$\begin{cases} (i.) \sum_{k=1}^{+\infty} \gamma_k \Delta M_k < +\infty \\ \\ (ii.) \lim_{k\to\infty} \eta_k = 0 \end{cases}$$

alors

$$X_k \to x^* \qquad quand \quad k \to +\infty,$$

sur l'événement $A_K^{x^*} \stackrel{\triangle}{=} \{X_k \in K \subset G_{x^*}\ infiniment\ souvent\}$, K *compact.*

Algorithme stochastique: ACLVQ

Nous proposons un nouvel algorithme stochastique *Average Competitive Learning Vector Quantization* (ACLVQ), simple modification de la méthode classique *Competitive Learning Vector Quantization* (CLVQ). Nous décrivons le schéma et discutons la convergence de notre méthode. Une étude par simulation a été faite pour évaluer le comportement de la méthode.

Schéma pour le ACLVQ

L'inconvénient du CLVQ est le temps de calcul. Cette méthode modifie un seul quantifieur à chaque itération, donc il faut un temps de calcul important pour atteindre de bons résultats

Nous proposons d'utiliser un ensemble de vecteurs aléatoires ξ au lieu d'un seul comme le fait la méthode CLVQ. Dans la *competitive phase*, le CLVQ utilise un vecteur aléatoire et modifie le quantifieur qui est le plus proche ce vecteur aléatoire. Notre version transforme cette phase. On utilise simultanément N vecteurs aléatoires au lieu d'un seul. La position courante du n-quantifieur étant $x^1, \cdots, x^n$, où $x^i \in \mathbb{R}^d$, $d \geq 2$, nous identifions les vecteurs aléatoire les plus proches de chaque quantifieur et dans la *learning phase* nous prenons la moyenne de tous les groupes constitués dans l'étape précédente pour modifier le n-quantifieur.

Par conséquent nous obtenons pour chaque élément du n-quantifieur (s'il existe au moins un vecteur aléatoire dans sa cellule de Voronoï) un nouveau vecteur aléatoire qui est la moyenne dans sa cellule. Autrement dit l'ACLVQ change au moins un élément dans le n-quantifieur chaque fois. En utilisant suffisamment de vecteurs aléatoires à chaque itération nous pouvons modifier tous les éléments du quantifieur à chaque étape. Évidemment le choix du nombre N de vecteurs aléatoires dépend fortement de n.

Étapes pour le ACLVQ

1. Créer un n-quantifieur initial $x_0^1, \cdots, x_0^n$.
2. Assigner $x^i = x_0^i$ pour tout i.
3. Tirer un ensemble de vecteurs aléatoires *i.i.d.* $\xi = (\xi_1, \cdots, \xi_N)$ avec la même loi $\mathbb{P}$.
4. Calculer la matrice de distance entre le n-quantifieur et les vecteurs aléatoires (***competitive phase***).

   Nous proposons une nouvelle méthode qui calcule la matrice de distance entre deux ensembles de vecteurs de dimension plus grande que deux. La procédure est décrite dans l'annexe.
5. Trouver les vecteurs aléatoires proches de chaque élément du quantifieur à l'étape k, i.e.:

$$I_k^j = \{i : \|\xi_i - x_k^j\| < \|\xi_i - x_k^m\|, \ \forall\, m \neq j\}, \quad j = 1, \cdots, n, \ i = 1, \cdots, N.$$

   On utilise la suite d'índices définis précédemment pour trouver les groupes pour chaque élément du quantifieur. Il se peut que pour un j, $I_k^j = \emptyset$.

6. Calculer la moyenne dans chaque groupe.

$$\begin{cases} \widetilde{\xi}_j = \frac{\sum_{i \in I_k^j} \xi_i}{|I_k^j|} & I_k^j \neq \emptyset \\ \\ \widetilde{\xi}_j = 0 & I_k^j = \emptyset, \end{cases}$$

où $|A|$ désigne le cardinal de l'ensemble A.

7. Actualiser le quantifieur (***learning phase***)

$$\begin{cases} x_{k+1}^i = x_k^i - \gamma_{k+1}(x_i^k - \widetilde{\xi}_i), & I_k^i \neq \emptyset \\ \\ x_{k+1}^i = x_k^i, & I_k^i = \emptyset \end{cases}$$

où $\gamma_k > 0$, appelé le *pas* de l'algorithme, satisfait

$$\sum \gamma_k = +\infty, \text{ et } \sum \gamma_k^2 < +\infty.$$

8. Répéter les étapes 2-7 jusqu'à vérifier un critère de convergence.

Pour obtenir le comportement asymptotique et les résultats de convergence de la méthode ACLVQ il faut écrire l'algorithme dans le cadre de Kushner-Clark. Si on définit $\mathbf{X} = (X^i)_{1 \leq i \leq n}$ où $X^i \in \mathbb{R}^d$ pour tout $i = 1, \cdots, n$ et si γ_k est le *pas* de l'algorithme alors l'ACLVQ peut se réécrire comme suit

$$X_{k+1}^i = X_k^i - \gamma_{k+1} H_i(\mathbf{X}_k, \boldsymbol{\omega}_{k+1}), \tag{7}$$

où $H_i(\mathbf{X}_k, \boldsymbol{\omega}_{k+1}) = X_k^i - \Theta_i(\mathbf{X}_k, \boldsymbol{\omega}_{k+1})$, $(\boldsymbol{\omega}_k)_{k \geq 1}$, est une suite de vecteurs aléatoires à valeurs dans $(\mathbb{R}^d)^N$; $\boldsymbol{\omega}_k = (\omega_k^1, \cdots, \omega_k^N)$, avec $\omega_k^j \overset{i.i.d}{\sim} \mathbb{P}$ pour tout k et $j = 1, \cdots, N$, $N \in \mathbb{N}$. Par conséquent $\boldsymbol{\omega}_k \sim \mathbb{P}_N = \bigotimes_{j=1}^N \mathbb{P}$. La fonction Θ_i est définie par

$$\Theta_i(\mathbf{X}_k, \boldsymbol{\omega}_{k+1}) = \begin{cases} \frac{\sum_{j=1}^N \mathbf{1}_{C_i(\mathbf{X}_k)}(\omega_{k+1}^j)\omega_{k+1}^j}{\sum_{j=1}^N \mathbf{1}_{C_i(\mathbf{X}_k)}(\omega_{k+1}^j)} & \text{si } \sum_{j=1}^N \mathbf{1}_{C_i(\mathbf{X}_k)}(\omega_{k+1}^j) > 0 \\ \\ 0 & \text{si } \sum_{j=1}^N \mathbf{1}_{C_i(\mathbf{X}_k)}(\omega_{k+1}^j) = 0, \end{cases}$$

où $C_i(\mathbf{X}_k)$ est la cellule de Voronoï pour X_k^i. Alors le modèle peut s'écrire comme

$$X_{k+1}^i = \begin{cases} X_k^i - \gamma_{k+1}\left(X_k^i - \frac{\sum_{j=1}^N \mathbf{1}_{C_i(\mathbf{X}_k)}(\omega_{k+1}^j)\omega_{k+1}^j}{\sum_{j=1}^N \mathbf{1}_{C_i(\mathbf{X}_k)}(\omega_{k+1}^j)}\right) & \text{si } \sum_{j=1}^N \mathbf{1}_{C_i(\mathbf{X}_k)}(\omega_{k+1}^j) > 0 \\ \\ X_k^i & \text{si } \sum_{j=1}^N \mathbf{1}_{C_i(\mathbf{X}_k)}(\omega_{k+1}^j) = 0 \end{cases}. \tag{8}$$

Pour le cas de $N = 1$ on retrouve le CLVQ classique.

Notre but est de montrer la convergence asymptotique de la solution de (8) vers un quantifieur stationnaire de la loi $\mathbb{P}$. Nous employons le Théorème de Kushner-Clark pour cela. Afin de l'appliquer nous devons montrer une propriété de continuité de la fonction $h_i(\mathbf{X}_k) = \mathbb{E}\left[H_i(\mathbf{X}, \boldsymbol{\omega}_{k+1})/\mathbf{X} = \mathbf{X}_k\right]$. Le lemme suivant assure la continuité de $\mathbf{h} = (h_1, \cdots, h_n)$.

***Lemme* 2**
La fonction $\mathbf{h} = (h_1, \cdots, h_n)$ *pour* $\mathbf{x} = (x^1, \cdots, x^n)$, $x^i \in \mathbb{R}^d$ *pour tout* $i = 1, \cdots, n$ *est* $\mathbb{P}$*-p.s. continue.*

Le résultat principal est

Théorème* 9 *Convergence pour l'ACLVQ
Si la suite $\mathbf{X}_k = (X_k^1, \cdots, X_k^n)$ *définie par l'algorithme stochastique du (7) vit dans un ensemble compact de* $\mathcal{O} = \{x \in \mathbb{R}^d, i \neq j \longrightarrow x_i \neq x_j\}$, *alors elle converge* $\mathbb{P}$*-p.s vers un zéro* $x^\star = (x^{\star,1}, \cdots, x^{\star,n})$ *de la fonction* $\mathbf{h}$, *qui est un quantifieur stationnaire.*

Estimation du paramètre de Hurst dans des processus gaussiens fractionnaires

Dans cette section, nous présentons une méthode basée sur la décomposition de Karhunen-Loève des processus gaussiens pour estimer le paramètre de Hurst de processus fractionnaires, spécifiquement le mouvement brownien fractionnaire (fBm), la famille des processus fractionnaires d'Ornstein-Uhlenbeck (fOU) et le pont brownien fractionnaire (fBb). Nous comparons nos résultats à ceux obtenus par la méthode du maximum de vraisemblance pour montrer la validité de notre proposition. Notre objectif est de proposer un estimateur du paramètre de Hurst d'un processus gaussien fractionnaire qui se comporte aussi bien que l'estimateur du maximum de vraisemblance mais plus robuste relativement aux problèmes numériques.

Mouvement brownien fractionnaire

Le fBm est un processus gaussien centré $\{W_t^\theta,\, t \geq 0\}$ nul en zéro avec paramètre de Hurst $\theta \in (0, 1)$ qui satisfait

1. La fonction de covariance pour W^θ est défini par

$$\mathbb{E}(W_t^\theta, W_s^\theta) = \frac{1}{2}\left(|t|^{2\theta} + |s|^{2\theta} - |t-s|^{2\theta}\right).$$

2. Le processus W^θ a des accroissements stationnaires et des trajectoires $p.s.$ continues.
3. Pour tout $a > 0$: $a^{-\theta} W_{at}^\theta \overset{\mathcal{D}}{=} W_t^\theta$, où $\overset{\mathcal{D}}{=}$ signifie même loi (θ-autosimilarité).

Pont brownien fractionnaire

Le fBb $\{B_t^\theta,\, t \geq 0\}$ avec paramètre de Hurst $\theta \in (0, 1)$ est défini sur $[0,\, 1]$ par

$$B_t^\theta = W_t^\theta - \frac{|t|^{2\theta} + 1 - |t-1|^{2\theta}}{2} W_1^\theta,$$

où $\{W_t^\theta,\, t \geq 0\}$ est un fBm défini sur $[0,\, 1]$.

Les processus fractionnaires d'Ornstein-Uhlenbeck

On peut dire qu'il y a trois type de fOU: les processus qui s'obtiennent comme solution de l'équation de Langevin avec le fBm, les processus stationnaires construits par la transformation de Lamperti du fBm et les processus avec une fonction de covariance "stretched". Voir Cheridito et al. [2003] et Kaarakka and Salminen [2007] pour plus de détails par rapport au deux premiers processus et voir Bogachev [1998] pour le troisième.

fOU(1) Le fOU(1) $X^{\theta,1}$ est la solution de l'équation différentielle stochastique avec le mouvement brownien fractionnaire ($\alpha > 0$)

$$dX_t^{\theta,1} = -\alpha(X_t^{\theta,1} - \mu)dt + \sigma dW_s^\theta,$$

et pour $\mu = 0$ on a

$$X_t^{\theta,1} = e^{-\alpha t}\left(\sigma \int_{-\infty}^{t} e^{\alpha s} dW_s^\theta\right).$$

fOU(2) Le fOU(2) $X^{H,2}$ est défini par la transformation de Lamperti comme

$$X_t^{\theta,2} = e^{-\alpha t} W^\theta_{a_{t,\theta}},$$

où $a_{t,\theta} = \theta e^{\alpha t/\theta}/\alpha$.

fOU(3) Le fOU(3) $X^{\theta,3}$ a une fonction de covariance défini par

$$\mathbb{E}(X_t^{\theta,3}, X_s^{\theta,3}) = e^{-\alpha|t-s|^{2\theta}}.$$

Notre cadre est tout à fait classique: nous disposons d'observations d'un processus aléatoire qui sont équidistantes ou aléatoirement distribuées, $X(t_1), \cdots, X(t_n)$.

Supposant que le modèle pour ces observations est un processus fractionnaire gaussien (mouvement brownien, processus d'Ornstein-Uhlenbeck, pont brownien) nous estimons le paramètre de Hurst noté θ, $\theta \in (0, 1)$. Nous sommes strictement dans le cadre d'estimation paramétrique.

Évidemment le meilleur estimateur s'obtient en maximisant la vraisemblance, mais cette procédure implique le calcul numérique de l'inverse de la matrice de covariance et du logarithme de son déterminant. La première partie est encore faisable quand n est de l'ordre 500 ou plus en utilisant la décomposition LU, mais la deuxième ne l'est plus.

Notre méthode alternative est très simple: elle se fonde sur un résultat simple, à l'aide de la décomposition de Karhunen-Loève. Dans la suite, pour tout le processus fractionnaire, il y a un paramètre d'intérêt θ, le paramètre de Hurst qui décrit la *rugosité* des trajectoires. Pour la famille de fOU nous supposons que $\mu = 0$, α et σ sont connus.

Pour le fBm, il y a plusieurs méthodes pour l'estimation du paramètre de Hurst. Dans Dieker [2002] un traitement général est fourni. Dans Jennane et al. [2001] les auteurs récapitulent une partie de ces méthodes et ils donnent des résultats numériques. Le lecteur peut trouver plus de détails dans Coeurjolly [2001] et Cohen [2004].

La littérature sur l'estimation du paramètre de Hurst pour le fOU(1) est encore réduite. Les résultats sont plus riches pour l'estimation des paramètres impliqués dans l'équation de Langevin. Cependant la méthode du maximum de vraisemblance (MLE) peut toujours être employée de la même manière. Il existe d'autres méthodes d'estimation issues de la théorie des fractals pour le fOU(3) (voir Gneiting et al. [2011]).

Cadre général

Nous considérerons dans toute cette partie que X appartient à un élément de la famille (X^θ) des processus gaussiens fractionnaires centrés avec paramètre de Hurst $\theta \in (0,1)$. Nous supposons que l'intervalle de temps d'observation pour le processus est $[0,1]$. On note θ_0 le vrai paramètre inconnu du processus observé, X^{θ_0}. Un processus gaussien possède une représentation unique dans la base de Karhunen-Loève, donnée par la diagonalisation de l'opérateur de covariance du processus X.

Pour chaque valeur de θ nous notons $(\lambda_k^\theta, \varphi_k^\theta)_{k\geq 1}$ les valeurs propres et les fonctions propres normales associées à l'opérateur de covariance par ordre décroissant. Alors il existe une suite des variables aléatoires normales standards ξ_i^θ *i.i.d.*, telle que

$$X^\theta(t) = \sum_{k\geq 1} \sqrt{\lambda_k^\theta}\xi_k^\theta \varphi_k^\theta(t).$$

Nous observons les trajectoires du processus X^{θ_0} aux temps $t_1, \cdots, t_n$, qui peuvent être régulièrement espacés ou aléatoirement distribués. Si $\mathbf{t} = (t_1, \cdots, t_n)'$ nous écrivons:

$$X^\theta(\mathbf{t}) = \begin{pmatrix} X^\theta(t_1) \\ \vdots \\ X^\theta(t_n) \end{pmatrix}.$$

Alors on écrit

$$X^{\theta_0}(t) = \sum_{k\geq 1} \rho_k^{\theta,\theta_0} \varphi_k^\theta(t),$$

où

$$\rho_k^{\theta,\theta_0} = \int_0^1 X^{\theta_0}(t)\varphi_k^\theta(t)dt \quad \stackrel{\theta=\theta_0}{\Longrightarrow} \quad \rho_k^{\theta_0,\theta_0} = \sqrt{\lambda_k^{\theta_0}}\xi_k^{\theta_0}.$$

Maintenant on note

$$Y^{\theta,\theta_0}(\mathbf{t}) = (\Sigma^\theta)^{-\frac{1}{2}} X^{\theta_0}(\mathbf{t}),$$

où Σ^θ est la matrice de covariance définie par

$$\Sigma^\theta = \left(\sum_{k\geq 1} \lambda_k^\theta \varphi_k^\theta(t_i)\varphi_k^\theta(t_j) \right)_{\substack{1\leq i\leq n \\ 1\leq j\leq n}}.$$

Évidement $Y^{\theta_0,\theta_0}(\mathbf{t}) \sim N(0, I_n)$ et $\|Y^{\theta_0,\theta_0}(\mathbf{t})\|^2/n$ *p.s.* converge vers 1.

La fonction de covariance de X^{θ_0}, en utilisant la base $\{\varphi_k^\theta\}_{k\geq 1}$, est calculée comme

$$C^{\theta,\theta_0}(t,s) = \sum_{\substack{l\geq 1 \\ k\geq 1}} \left(\int_0^1 \int_0^1 C^{\theta_0}(u,v)\varphi_l^\theta(u)\varphi_k^\theta(v)dudv \right) \varphi_l^\theta(t)\varphi_k^\theta(s),$$

où $C^{\theta_0}(u,v)$ est la fonction de covariance de X^{θ_0} avec la base $\{\varphi_k^{\theta_0}\}_{k\geq 1}$ correspondante. Si on note

$$A_{l,k}^{\theta,\theta_0} = \int_0^1 \int_0^1 C^{\theta_0}(u,v)\varphi_l^{\theta}(u)\varphi_k^{\theta}(v)dudv,$$

l'expression finale pour la fonction de covariance du processus X^{θ_0} est

$$C^{\theta,\theta_0}(t,s) = \sum_{\substack{l\geq 1\\ k\geq 1}} A_{l,k}^{\theta,\theta_0}\varphi_l^{\theta}(t)\varphi_k^{\theta}(s).$$

Par conséquent la matrice de covariance de $X^{\theta_0}(\mathbf{t})$ est

$$\Sigma^{\theta,\theta_0} = \left(\sum_{\substack{l\geq 1\\ k\geq 1}} A_{l,k}^{\theta,\theta_0}\varphi_l^{\theta}(t_i)\varphi_k^{\theta}(t_j)\right)_{\substack{1\leq i\leq n\\ 1\leq j\leq n}}.$$

Pour des valeurs générales de θ, on calcule la fonction de covariance de Y^{θ,θ_0} comme suit

$$cov(Y^{\theta,\theta_0}(\mathbf{t})) = (\Sigma^{\theta})^{-1/2}\Sigma^{\theta_0}(\Sigma^{\theta})^{-1},$$

et on a

$$\mathbb{E}\|Y^{\theta,\theta_0}(\mathbf{t})\|^2 \;=\; \mathbb{E}\left(X^{\theta_0}(\mathbf{t})^{'}(\Sigma^{\theta})^{-1}X^{\theta_0}(\mathbf{t})\right) = \mathrm{tr}(\Sigma^{\theta,\theta_0}(\Sigma^{\theta})^{-1}).$$

Nous définissons la fonction de contraste h comme:

$$h(\theta;\theta_0) = \lim_{n\to+\infty}\frac{\mathrm{tr}(\Sigma^{\theta,\theta_0}(\Sigma^{\theta})^{-1})}{n} = \lim_{n\to+\infty}\frac{\mathbb{E}\|Y^{\theta,\theta_0}(\mathbf{t})\|^2}{n}.$$

Donc si $\theta = \theta_0$, alors $\Sigma^{\theta_0,\theta_0} = \Sigma^{\theta_0}$ et

$$h(\theta_0;\theta_0) = \lim_{n\to+\infty}\frac{\mathrm{tr}(\Sigma^{\theta_0,\theta_0}(\Sigma^{\theta_0})^{-1})}{n} = \lim_{n\to+\infty}\frac{\mathrm{tr}(I_n)}{n} = \lim_{n\to+\infty}\frac{n}{n} = 1.$$

La fonction h prend donc la valeur 1 quand $\theta = \theta_0$. Nous avons montré pour quelques processus le résultat suivant:

***Théorème* 10**
Soit X^{θ} un processus stochastique fractionnaire sur $[0,1]$ avec paramètre de Hurst $\theta \in (0,1)$ et soit $X^{\theta_0}(\mathbf{t})$ pour $\mathbf{t} = (t_1,\cdots,t_n)$ l'observation partielle d'une trajectoire du processus X^{θ_0}. Supposons que les temps $t_1,\cdots,t_n$ sont régulièrement espacés ou aléatoirement distribués (i.i.d) avec densité positive. La fonction

$$h(\theta;\theta_0) = \lim_{n\to+\infty}\frac{\mathrm{tr}(\Sigma^{\theta,\theta_0}(\Sigma^{\theta})^{-1})}{n},$$

satisfait

- ✓ *Si $\theta = \theta_0$ alors $h(\theta;\theta_0) = 1$.*
- ✓ *Si $\theta > \theta_0$ alors $h(\theta;\theta_0) = \infty$.*
- ✓ *Si $\theta < \theta_0$ alors $h(\theta;\theta_0) = 0$.*

Une preuve complète est fournie dans quelques cas particuliers. De plus, nous avons fait une étude par simulation qui vérifie la validité de notre proposition dans d'autres cas.

Résultats asymptotiques autour de la quantification fonctionnelle

Taux asymptotique de l'erreur de quantification

Le comportement asymptotique de l'erreur de quantification d'un processus gaussien X avec fonction de covariance exponentielle carrée

$$\Gamma(t,s) = e^{-\theta|t-s|^2},$$

n'a pas été étudié, alors que cette forme de covariance est fréquemment utilisée en krigeage.

Cadre général

Ici nous présentons quelques résultats préliminaires pour obtenir le taux asymptotique. Soit X un processus stochastique centré, $(\lambda_k)_{k\geq 1}$ la suite ordonnée des valeurs propres associées à l'opérateur de covariance du processus et $e_n^2(X)$ l'erreur de quantification de X. Nous rappelons quelques résultats relatifs à la borne supérieure de $e_n^2(X)$ (voir la section 4.1 dans Luschgy and Pagès [2002]).

$$e_n^2(X) \;\leq\; \sum_{j\geq m+1} \lambda_j + \inf\left\{\sum_{j=1}^{m} \lambda_j n_j^{-2} :\; n_1,\cdots,n_m \in \mathbb{N} : \prod_{j=1}^{m} n_j \leq n\right\},$$

et

$$\inf\left\{\sum_{j=1}^{m} \lambda_j n_j^{-2} :\; n_1,\cdots,n_m \in \mathbb{N} : \prod_{j=1}^{m} n_j \leq n\right\} \;=\; \sum_{j=1}^{m} \lambda_j z_j^{-2} = n^{-\frac{2}{m}} m \left(\prod_{j=1}^{m} \lambda_j\right)^{1/m},$$

où

$$z_j = n^{\frac{1}{m}} \lambda_j^{1/2} \left(\prod_{j=1}^{m} \lambda_j\right)^{-1/2m}.$$

Si on note

$$m_n^\star = \max\left\{m \geq 1 : n^{\frac{2}{m}} \lambda_m \left(\prod_{j=1}^{n} \lambda_j\right)^{-1/m} \geq 1\right\},$$

et

$$a_k = \frac{1}{2}\ln\left(\frac{\prod_{j=1}^{k}\lambda_j}{\lambda_k^k}\right) = \frac{k}{2}\ln\left(\frac{(\prod_{j=1}^{k}\lambda_j)^{1/k}}{\lambda_k}\right) = \frac{1}{2}\sum_{j=1}^{k}\ln\left(\frac{\lambda_j}{\lambda_k}\right).$$

Évidement $a_k \to +\infty$ quand $k \to +\infty$ et on peut définir

$$m_n^\star = \max\left\{m \geq 1 : a_m \leq \ln n\right\}.$$

De plus,

$$a_{m_n^\star} \leq \ln n < a_{m_n^\star+1}.$$

Dans les conditions indiquées dans le Corollaire 2 de Widom [1964] le comportement asymptotique des valeurs propres $(\lambda_k)_{k\geq 1}$ est $\ln \lambda_k \sim -k \ln k$. Nous pouvons supposer sans perte de généralité que $\ln \lambda_k = -k \ln k$ parce que nous allons démontrer un résultat du type $\approx$ pour $e^2(X)$ et qu'on peut utiliser la même idée que dans la démonstration du Corollaire 4.13 (c) dans Luschgy and Pagès [2002] relativement au comportement des valeurs propres de l'opérateur de covariance, résultat basé sur le Lemme 4.11 de Luschgy and Pagès [2002] et par conséquent

$$a_m \sim \frac{m^2 \ln m}{4}. \tag{9}$$

Par suite de (9) et la définition de $m_n^\star$ il vient que

$$m_n^\star \sim 2\sqrt{2}(\ln n)^{\frac{1}{2}}(\ln \ln n)^{-\frac{1}{2}}.$$

Si on note

$$R_n = \sum_{k\geq n+1} \lambda_k + n\lambda_n = \sum_{k\geq n+1} k^{-k} + n^{1-n}.$$

On peut montrer que

$$R_n \approx n^{1-n}.$$

Le théorème principal s'énonce ainsi

***Théorème* 11**
Si X est un processus stochastique de fonction de covariance $\Gamma(t,s) = e^{-\theta|t-s|^2}$ alors les valeurs propres du processus $(\lambda_k)_{k\geq 1}$ satisfont $\ln \lambda_k \sim -k \ln k$, on a:

$$e_n(X) = O\left(\psi\left((2\sqrt{2})^{1-\delta}(\ln n)^{\frac{1}{2}}(\ln \ln n)^{-\frac{1}{2}}\right)^{-1}\right)$$

et

$$e_n(X) = \Omega\left(\psi\left((2\sqrt{2})^{1+\delta}(\ln n)^{\frac{1}{2}}(\ln \ln n)^{-\frac{1}{2}}\right)^{-1}\right),$$

où $\psi(n) = n^{n-1}$ pour $\delta > 0$.

Rayon maximal pour les processus gaussiens

Nous présentons ici quelques résultats asymptotiques sur le rayon maximal du processus de Wiener. Nous suivons le travail présenté dans la thèse de Sagna [2008]. L'auteur fournit des bornes supérieures et inférieures pour le rayon maximal dans le cadre fini dimensionnel. La prolongation au cas infini apporte des difficultés supplémentaires parce que certaines bonnes propriétés des quantifieurs optimaux de $\mathbb{R}^d$ ne sont pas connues en dimension infinie. Notre but est d'obtenir des résultats similaires à ceux de Sagna pour le rayon maximal de la suite $(\alpha_n)_{n\geq 1}$ de quantifieurs n-optimal du processus de Wiener, le rayon maximal étant défini pour $n \geq 1$ par:

$$\rho_n = \rho(\alpha_n) = \max\{\|a\|_{\mathcal{L}^2}, a \in \alpha_n\}.$$

Rayon maximal pour les quantifieurs produits

Dans la suite nous travaillons avec une version modifiée des quantifieurs produits définis dans Pagès and Printems [2005]. L'idée principale est d'employer encore une fois la décomposition de Karhunen-Loève d'un processus gaussien X pour produire une quantification fonctionnelle pour X en utilisant les quantifieurs élémentaires de la forme

$$\widehat{X}_t = \sum_{j\geq 1} \sqrt{\lambda_j}\widehat{\xi}_j \varphi_j(t), \tag{10}$$

où $\{\varphi_j\}_{j\geq 1}$ est la base de Karhunen-Loève de X et $(\lambda_j)_{j\geq 1}$ est la suite ordonné des valeurs propres de $\{\varphi_j\}_{j\geq 1}$, $\widehat{\xi}_j$ est le quantifieur n_j-optimal de la $N(0,1)$ et $n_1 \times \cdots \times n_k \leq n$, $n_1, n_2, \cdots, n_k \geq 1$ pour tout k. Pour k assez grand on a que $n_k = 1$ et cela implique que $\widehat{\xi}_j = 0$. Par conséquent la série ci-dessus devient une somme finie quand n est fixe pour k assez grand. Le $n_1 \times \cdots \times n_k$-quantifieur produit χ (10) est de la forme

$$\chi_{\underline{i}}(t) = \sum_{j\geq 1} \sqrt{\lambda_j} x_{i_j}^{n_j} \varphi_j(t), \tag{11}$$

où $x^{n_j} = (x_1^{n_j}, \cdots, x_{n_j}^{n_j})$ est le n_j-quantifieur unique de la normale sur $\mathbb{R}$, $\underline{i} = (i_1, i_2, \cdots, i_k)$ et $i_j \in \{1, \cdots, n_j\}$ pour tout $j = 1, 2, \cdots, k$.

Si on note par $\mathcal{O}(X, n)$ l'ensemble

$$\mathcal{O}(X,n) = \{\chi : \chi \text{ est le quantifieur produit de taille au plus } n \text{ défini par (11)}\}.$$

Pour chaque n on pose

$$D_n = \max\{l : n_1 \times \cdots \times n_l \leq n,\ n_1, \cdots, n_l \geq 2\}.$$

Il est facile de vérifier pour n fixé que $D_n \leq k$. Évidemment D_n n'est pas borné quand $n \to +\infty$, donc la somme finie de (11) se transforme en une somme infinie. Afin d'éviter cet écueil nous pouvons travailler avec un sous-ensemble approprié de $\mathcal{O}(X, n)$. Fixant $k_0 \in \mathbb{N}$ pour tout n nous pouvons écrire

$$\chi_{\underline{i}}^{k_0}(t) = \sum_{j\geq 1}^{k_0} \sqrt{\lambda_j} x_{i_j}^{n_j} \varphi_j(t), \tag{12}$$

où $x^{n_j} = (x_1^{n_j}, \cdots, x_{n_j}^{n_j})$ est le n_j-quantifieur unique de la normal sur $\mathbb{R}$, $\underline{i} = (i_1, i_2, \cdots, i_{k_0})$ et $i_j \in \{1, \cdots, n_j\}$ pour tout $j = 1, 2, \cdots, k_0$. Alors on pose:

$$\mathcal{O}_{k_0}(X,n) \ = \ \left\{\chi^{k_0} : \chi^{k_0} \text{ est le quantifieur produit de taille au plus } n \text{ défini par (12)}\right\}.$$

De la définition ci-dessus, il est facile de vérifier que

$$\mathcal{O}_{k_0}(X,n) \subset \mathcal{O}(X,n),$$

pour tous n et k_0 fixés.

***Théorème* 12**
Si on note

$$\rho_{n,k_0} = \max\{\|\chi^{k_0}_{\underline{i}}\|_{\mathcal{L}^2}, \chi^{k_0}_{\underline{i}} \in \mathcal{O}_{k_0}(X,n)\},$$

le rayon maximal de $\chi^{k_0} \in \mathcal{O}_{k_0}(X,n)$ *alors*

$$(6c_{j_0})^{1/2} \leq \liminf_{n\to+\infty} \frac{\rho_{n,k_0}}{\sqrt{\ln n}} \leq \limsup_{n\to+\infty} \frac{\rho_{n,k_0}}{\sqrt{\ln n}} \leq (6C_{j_0})^{1/2},$$

pour certaines constantes positives c_{j_0}, C_{j_0}.

Rayon maximal pour les quantifieurs linéaires

Ici on travaille directement avec une approximation linéaire du processus stochastique X sur $[0,1]$. On note π une discrétisation de l'intervalle de temps $[0,1]$:

$$\pi = \{0 = t_0 < t_1 < \cdots < t_{k-1} < t_k = 1\}.$$

L'interpolation linéaire de X par π est

$$X^\pi_L(t) = \sum_{i=0}^{k-1} \left[\frac{t_{i+1}-t}{\Delta_i} X(t_i) + \frac{t-t_i}{\Delta_i} X(t_{i+1})\right] \mathbf{1}_{]t_i,t_{i+1}]}(t),$$

où $\Delta_i = t_{i+1} - t_i$, $\forall\, i = 1, 2, \cdots, k-1$. L'exposant π dans $X^\pi_L(t)$ sera omis en l'absence d'ambiguïté. La longueur du plus long sous-intervalle de π est

$$\delta_\pi = \sup_{0\leq i\leq k-1} |\Delta_i|.$$

Soit H^π l'ensemble des fonctions X_L linéaires par morceaux de $[0,1]$ sur les points t_i, $i = 0, 1, \cdots, k$. Évidemment, $H^\pi \subset H = \mathcal{L}^2([0,1], dt)$. Soit T^π l'opérateur linéaire borné défini par

$$\begin{array}{rcl} T^\pi : (\mathcal{L}^2, \|\cdot\|_{\mathcal{L}^2}) & \longrightarrow & (H^\pi, \|\cdot\|_{\mathcal{L}^2}) \\ X & \overset{\pi}{\rightsquigarrow} & T^\pi(X) = X^\pi_L. \end{array}$$

Clairement T^π est surjectif. Soit $A^\pi = (T^\pi)^{-1}(f^\pi)$ avec $f^\pi \in H^\pi$, alors $A^\pi \neq \emptyset$ car f^π appartient à A^π. La norme définie sur H^π est la même que la norme sur H.

Pour $X_L \in H^\pi$ on a

$$\|X_L\|^2_{\mathcal{L}^2} = \frac{1}{3} \sum_{i=0}^{k-1} \Delta_i (X^2(t_{i+1}) + X^2(t_i) + X(t_i)X(t_{i+1})).$$

La formule précédente nous permet de calculer la norme de X_L en utilisant exactement les valeurs de X^π avec $X^\pi = (X(t_0), \cdots, X(t_k))$. De la définition de X^π, on voit clairement que $X^\pi \in \mathbb{R}^{k+1}$. Dans cette espace on peut définir la norme suivante

$$\|X^\pi\|^2_{\mathbb{R}^{k+1}} \overset{\triangle}{=} \|X_L\|^2_{\mathcal{L}^2}. \tag{13}$$

Cette norme $\|\cdot\|^2_{\mathbb{R}^{k+1}}$ est équivalente à la norme euclidienne sur $\mathbb{R}^{k+1}$.

Soit W un processus de Wiener sur [0, 1]. L'interpolation linéaire W_L de W sur les points t_i, $\forall\, i = 0, 1, \cdots, k$ est

$$T^\pi(W(t)) = W_L(t) = \sum_{i=0}^{k-1} \left[\frac{t_{i+1} - t}{\Delta_i} W(t_i) + \frac{t - t_i}{\Delta_i} W(t_{i+1}) \right] 1_{]t_i, t_{i+1}]}(t), \tag{14}$$

et satisfait

$$\lim_{\delta_\pi \to 0} \mathbb{E}\|W - W_L^\pi\|_{\mathcal{L}^2} = 0.$$

Pour les processus gaussiens, on a de plus la convergence en espérance maximale carrée

$$\lim_{\delta_\pi \to 0} \max_{t \in [0,1]} \mathbb{E}(W(t) - W_L^\pi(t))^2 = 0.$$

Voir par exemple Hüsler et al. [2003] et leurs références. Le taux d'approximation pour l'interpolation linéaire en espérance carrée est également optimal dans un certain sens pour certain processus (voir Seleznjev [1996]). Les propriétés optimales de l'interpolation linéaire pour les trajectoires des processus aléatoires sont étudiées dans Su and Cambanis [1993].

***Théorème* 13**
Pour toute discrétisation $\pi = \{0 = t_0 < t_1 < \cdots < t_{k-1} < t_k = 1\}$ on note W_L^π l'interpolation linéaire définie dans (14) pour W sur H^π et $W^\pi = (W(t_0), \cdots, W(t_k))$. Alors

1. *L'erreur de quantification $e_n^2(W_L^\pi)$ satisfait*

$$\lim_{\delta_\pi \to 0} e_n^2(W_L^\pi) = e_n^2(W),$$

   et

2. *si $\widetilde{e}_n^2(W^\pi)$ est l'erreur de quantification pour le vecteur gaussien W^π sur $\mathbb{R}^{k+1}$ avec la norme définie par (13), alors*

$$e_n^2(W_L^\pi) = \widetilde{e}_n^2(W^\pi).$$

***Remarques* 1**
Une conséquence du Théorème 13 est que si $\alpha = (a_1, \cdots, a_n)$ est un quantifieur n-optimal pour W et $\beta^\pi = \{\beta_1, \cdots \beta_n\}$ est un quantifieur asymptotiquement optimal pour W^π, alors $a_j(t_i) = \beta_j^i$, $\forall\, a_j \in \alpha$, $i = 0, 1, 2, \cdots, k-1$, et

$$a_{j,L}^\pi(t) = \sum_{i=0}^{k-1} \left[\frac{t_{i+1} - t}{\Delta_i} \beta_j^i + \frac{t - t_i}{\Delta_i} \beta_j^{i+1} \right] \mathbf{1}_{]t_i, t_{i+1}]}(t), \qquad \forall\, j = 1, \cdots, n.$$

Le théorème suivant, concernant le rayon maximal de l'interpolation linéaire du mouvement brownien, est une conclusion immédiate du théorème précédent et des résultats de Sagna [2008].

Nous rappelons brièvement les hypothèses de base employée par Sagna pour obtenir l'asymptotique du rayon maximal d'un quantifieur optimal d'un vecteur aléatoire X à valeurs dans $\mathbb{R}^d$.

- Le vecteur aléatoire X a une loi $\mathbb{P}$ avec support non bornée et un moment d'ordre r fini.
- La loi $\mathbb{P}$ pour X satisfait

$$\mathbb{P}(dx) \geq \varepsilon_0 \mathbf{1}_{x \in \overline{B}(x_0, r_0)} \lambda_d(dx), \qquad \varepsilon_0,\, r_0 > 0,\; x_0 \in \mathbb{R}^d,$$

  où λ_d est la mesure de Lebesgue sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ et $\overline{B}(x_0, r_0)$ est la boule fermée en $\mathbb{R}^d$. Cette hypothèse est vérifiée par les distributions habituelles, en particulier toutes les gaussiennes.
- Dans ce cadre la norme de base utilisée par Sagna [2008] est la norme euclidienne. Cette dernière hypothèse a pu être relâchée dans certains cas.

Le théorème principal est le suivant

***Théorème* 14**
Le rayon maximal $\rho_{n,L}$ pour l'interpolation linéaire W_L par π défini comme

$$\rho_{n,L} = \rho(\alpha_L) = \max\{\|a\|_{\mathcal{L}^2}, a \in \alpha_L\}, \qquad \forall\, n \in \mathbb{N},$$

satisfait

$$c_k \sqrt{\frac{4+2k}{k}} < \liminf_{n \to +\infty} \frac{\rho_{n,L}}{\sqrt{\ln n}} \leq \limsup_{n \to +\infty} \frac{\rho_{n,L}}{\sqrt{\ln n}} < 2C_k \sqrt{\frac{4+2k}{k}},$$

où c_k, C_k sont des constantes positives.

Introduction

The purpose of the present thesis is to study the theory of functional quantization for some Gaussian process. Our goal is to investigate some general asymptotic properties of the quantization error and concepts related as the maximal radius of the optimal quantizer. We also develop a new method based on the Karhunen-Loève expansion of fractional Gaussian process to estimate the Hurst parameter associated to this processes. We derive a new stochastic algorithm mainly based on the *Competitive Learning Vector Quantization* (CLVQ). We examine the convergence of this method and present some numerical results of it behaviour.

Motivation

The term "quantization" originates in the theory of signal processing in electrical engineering in the early 1950's. The main idea was to use a finite number of n codes (or quantizers) to transmit efficiently a continuous stationary signal.

Rigourously speaking the quantization for probability distributions is related with the best approximation of a d-dimensional probability distribution $\mathbb{P}$ by a discrete probability distribution with finite support. The quantization searches the best approximation of a d-dimensional random vector with probability distribution $\mathbb{P}$ by a random vector Y with at most n values in its image. The mathematical aspects of quantization are treated extensively in Graf and Luschgy [2000]. This theory has been applied in several areas such as cluster analysis, pattern recognition, finance and numerical probability. For a more detailed exposition on the application of the vector quantization we refer the reader to Gersho and Gray [1992], Tarpey [1996], Bally et al. [2003] and Pagès et al. [2004] and references therein.

A quick understanding on the fields of application of vector quantization leads us to the methods needed to obtain optimal quantizers. The theory of stochastic algorithms is the answer to that question.

The basic stochastic approximation algorithms was introduced in the 1950s by the works of Robbins and Monro. They developed a recursive procedure for finding the root of a real-valued function. The idea is to take an observation at the current estimator of the root and use it to make a small correction in the estimate, then takes an observation at the new value of the estimator, and so forth. Among the wide range of application of this theory we can mention the problems of estimating unknown parameters based on observation data containing information (pattern classification, adaptative control, etc) or the problem of searching the maximum of certain functions (see Duflo [1996, 1997] for

instance).

The stochastic gradient method is based on the integral representation of the criterion to be optimized and can be seen as a particular case of the Robbins-Monro algorithm when the function has an unique point of minima, and both, the function and its definition space are convex. The CLVQ is an on-line algorithm that can be seen as a particular case of the "stochastic gradient method with decreasing step".

The application of the CLVQ on the searching of optimal quantizers has the drawback that an excessive amount of calculations is required to achieve good numerical results. The *competitive phase* in the algorithm seems to be too slow: at each step only one quantizer is modified. For that reason we propose a new variant of the method.

Our main interest is in the infinite dimensional version of the vector quantization: the functional one. The aim of the so-called "functional quantization" of stochastic processes is to quantize some processes viewed as random vectors taking values in their path spaces. In general, functional quantization consists in studying the best approximation of stochastic process X in a Hilbert space H by H-valued vectors taking at most n values. For the Gaussian case the general properties has been well studied, see for instance Luschgy and Pagès [2002], Dereich et al. [2003], Graf et al. [2003], Dereich [2005]), Pagès and Printems [2005], Dereich and Scheutzow [2006], Dereich and Lifshits [2005] and references therein.

We specifically focus our attention on the asymptotic behaviour of the quantization error. The methods used to establish the rate of convergence relies in the behaviour of the eigenvalues of the covariance operator, regularly varying functions, small ball probabilities and Shannon-Kolmogorov ϵ-entropy.

The first approach to the asymptotics of quantization error is due to Luschgy and Pagès [2002]. Upper and lower bounds for the quantization error are given using the eigenvalues behaviour of the covariance operator and Shannon-Kolmogorov's ϵ-entropy respectively. A different approach based on regularly varying functions, small ball probabilities and their link with the quantization error is proposed by Graf et al. [2003] and Dereich [2003]. The most importants results for sharp asymptotics of quantization error are detailed in Luschgy and Pagès [2004a,b]. Using the general idea proposed in Luschgy and Pagès [2002] and Widom [1964] we obtain some asymptotics for the quantization error of a Gaussian process with covariance function $\Gamma(t) = e^{-\theta t^2}$, which was not handled before whilst arising in several kriging applications.

Following the results obtained in Sagna [2008] for the maximal radius in $\mathbb{R}^d$ of the optimal quantizer we will intend to achieve similar asymptotics for the maximal radius in the infinite dimensional case, specifically for the Brownian motion.

The study of the Karhunen-Loève expansion of Gaussian process was the initial point of the study of parametric estimation in fractional processes. A new method more robust by numerical computation that the maximum likelihood is proposed.

Classical framework

Quantization settings

Let H be Hilbert space. The Voronoï partition of H is an important concept for optimal quantization. In general the Voronoï region generated by $a \in \alpha$, where α is a finite set is defined by

$$W(a|\alpha) \stackrel{\triangle}{=} \{x \in H : ||x - a|| = \min_{b \in \alpha} ||x - b||\}.$$

The set $\{W(a|\alpha) : a \in \alpha\}$ is called the Voronoï diagram of α, and is a local covering of H. The Borel measurable partition $\{C_a : a \in \alpha\}$ of H with respect to α is called Voronoï partition if $C_a \subset W(a|\alpha)$. This concept allow us to obtain the optimal quantizer.

In general the quantization problem consists in approximating a vector X by a random vector taking finitely many values in H: in vector quantization ($H = \mathbb{R}^d$) this can be seen as the problem of finding a discrete probability distribution which is a good approximation (in some sense) of the probability distribution of the random vector X and the functional quantization, usually $H = \mathcal{L}^2([0,1])$, consists in studying the best approximation of a stochastic process X in a Hilbert space H by H-valued vectors taking at most n values. The best references for both are Graf and Luschgy [2000] and Luschgy and Pagès [2002] respectively.

Let us consider a separable Hilbert space $(H, \langle \cdot, \cdot \rangle_H)$ with its natural σ-algebra and satisfying $\mathbb{E}||X||^2 < +\infty$. One considers a random vector X defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking its values in H. The quantization problem for X of level n consists in minimizing

$$\left(\mathbb{E} \min_{a \in \alpha} ||X - a||^r\right)^{1/r},$$

over all the sets $\alpha \subset H$ and $|\alpha| \leq n$. The set α is also called *n-codebook* or n-quantizer.

The n-th quantization error for X of order r, usually called "distortion", is defined by

$$\begin{aligned} e_{n,r}(X) &= \inf_{\substack{\alpha \subset H \\ |\alpha| \leq n}} (\mathbb{E} \min_{a \in \alpha} ||X - a||^r)^{1/r} \\ &= \inf \left\{ (\mathbb{E} \min_{a \in \alpha} ||X - a||^r)^{1/r} : \alpha \subset H,\ |\alpha| \leq n \right\}. \end{aligned}$$

Thorough the thesis we work with the quadratic case ($r = 2$). The usual norm in vector quantization is the Euclidean one and the $\mathcal{L}^2$-norm for the infinite case.

The best approximation $\widehat{X}^\alpha : H \to \alpha \subset H$ of X by the rule of closest neighbour is

$$\widehat{X}^\alpha = \sum_{i=1}^{n} x_i \mathbf{1}_{C_i}(X),$$

where $\alpha \subset H$ with $|\alpha| \leq n$ is one n-optimal quantizer and C_i is the Voronoï partition induced by each x_i.

Stochastic algorithms

We work here with a version of the *Competitive Learning Vector Quantization*: an online algorithm which is a particular case of the "stochastic gradient method with decreasing step" (see Duflo [1996]).

The classical formulation for stochastic algorithms is given by

$$X_{k+1} = X_k + \gamma_k F(X_k, \omega_{k+1}),$$

where $(\omega_k)_{k\geq 1}$ is a sequence of real-valued, mutually independent, and identically distributed, random variables with finite variance, $(X_k)_{k\geq 1}$ is a sequence in some Euclidean space, $(\gamma_k)_{k\geq 1}$ is a sequence that tends to zero when k goes to infinity, (in more specific problems, some others properties are required for this sequence) and F is a function that take different form for each particular problem.

The behaviour of some classical algorithms (Robbins-Monro, Kiefer-Wolfowitz) are discussed in Duflo [1996, 1997]. We also refer the reader to Fort and Pagès [1996] for a different approach based on the Kushner-Clark theorem. It is also called *method of ordinary differential equation* and was introduced in Kushner and Clark [1978]. The most important results on *a.s.* convergence of stochastic algorithms is due to the mentioned theorem.

Contribution of the thesis

This thesis is divided in six chapters and anexes. The first three chapters are a review of the state of art and the last three ones present our new results. We present here a brief summary of our work and we discuss some important aspects of the thesis.

Chapter presentation

Chapter 1 concerns the study of stochastic process. We briefly summarize the definitions and properties of the Gaussian process that we consider in our work. The principal aspects to note in this chapter are the simulation methods and the expansion of Gaussian process in the Karhunen-Loève basis.

In **Chapter 2** we present the most important results of the quantization theory. The first part of the chapter provides the basis of vector quantization. We give a detailed exposition of the general formulation and the asymptotic behaviour of the quantization error. On the second part of the chapter we discuss the functional case. A deeper treatment of the asymptotic behaviour is exposed.

Chapter 3 contains a brief summary on stochastic algorithms. We touch a few aspects of the theory: the general formulation and the most important convergence theorems. We also describe the scheme of two methods: the CLVQ and the Lloyd's method due to the importance of both in our work.

Chapter 4 is devoted to the study of our stochastic algorithm: the *Average Competitive Learning Vector Quantization* (ACLVQ). This new method use the basis of the CLVQ and introduce a slight modification in the *competitive phase* like a short Lloyd to

improve the results. The convergence of this new method is discussed. We carry out a simulation study which shows its numerical behaviour.

In **Chapter 5** we develop a new estimation method for the Hurst parameter of fractional Gaussian process based on its Karhunen-Loève expansion. The estimator proposed performs almost as well as the maximum likelihood estimator but is a bit more robust to computational problems. We investigate the cases of the fractional Brownian motion (fBm), the fractional Ornstein-Uhlenbeck (fOU) family and the fractional Brownian bridge (fBb). We numerically compare our results with the ones obtained by the maximum likelihood method.

Finally **Chapter 6** contains the asymptotics results related with the quantization error in the functional case. We first provide some asymptotics for e_n of a Gaussian process with covariance function given by $\Gamma(t) = e^{-\theta t^2}$. Our idea relies in the asymptotic behaviour of the eigenvalues associated to Γ obtained by using the work of Widom [1964] and Luschgy and Pagès [2002, 2004a]. One application to kriging is also discussed. In the second part we restrict our attention in the asymptotic behaviour of the maximal radius of the quantizer n-optimal of the Wiener process.

Some conclusions and considerations are provided at the end of this work.

Average Competitive Learning Vector Quantization

The CLVQ is a stochastic gradient algorithm with decreasing step. The use of this method in the quantization setting (specifically in $\mathbb{R}^d$ for convenience, but it is straightforwardly extend to a Hilbert space) has a fundamental drawback: the amount of calculations that requieres the method to achieve good numerical results (see for instance Pagés and Printems [2003]).

Usually the probability distribution to quantize is denoted by $\mathbb{P}$ and we assume that $\mathbb{P}$ is diffuse. Let $\mathcal{O} = \{x \in (\mathbb{R}^d)^n, \forall i \neq j,\ x_i \neq x_j\}$. This assumption ensures that $\mathbb{P}(x \in \mathcal{O}) = 1$.

One of the drawback of the CLVQ comes from the *competitive phase* (most time consuming). The new method proposed generate a set of N random vectors ξ instead of one as the CLVQ method do to find the "winning index". Then we identify the nearest random vector to each quantizer using some distance criterium. In the *learning phase* we take the mean of all groups formed in the previous step. Therefore we obtain for each element of the n-quantizer (in the case that exists at least one random vector in its Voronoï tessel) a new random vector that is the mean of the group drawn in its tessel.

The discussion which follows is focused in the convergence of this new method and is based on the Kushner-Clark theorem. The simulation study that we carry out also shows the validity of our proposal.

Estimation of the Hurst parameter in fractional Gaussian process

The general estimation method used for the estimation of the Hurst parameter in fractional Gaussian processes is based on the maximum likelihood estimator. For the fractional Brownian motion there are several other methods to estimate the Hurst parameter, see for instance Jennane et al. [2001], Coeurjolly [2001] and Cohen [2004]. For the case

of the fractional Ornstein-Uhlenbeck family the maximum likelihood method (MLE) can be always used. The drawback of the MLE is that it involves the numerical computation of the inverse of the covariance matrix and the logarithm of its determinant. The first is still tractable when n is of magnitude larger than 500 using LU decomposition, but the second is not.

We consider that X belongs to one of the family (X^θ) of centered fractional Gaussian process in $[0,1]$ with Hurst parameter θ in $(0,1)$ and we denote by θ_0 the actual and unknown parameter of the process actually observed, X^{θ_0}. We use the unique representation of X^θ in the basis of Karhunen-Loève

$$X^\theta(t) = \sum_{k\geq 1} \sqrt{\lambda_k^\theta}\xi_k^\theta \varphi_k^\theta(t),$$

where $(\lambda_k^\theta, \varphi_k^\theta)_{k\geq 1}$ are the eigenvalues and associated normalized eigenfunctions of the covariance operator of X^θ in decreasing order and $(\xi_k^\theta)_{k\geq 1}$ are an $i.i.d.$ sequence of standard normal random variables.

Based on the observations we construct an *asymptotic* contrast function h such that:
✓ If $\theta = \theta_0$ then $h(\theta;\theta_0) = 1$.
✓ If $\theta > \theta_0$ then $h(\theta;\theta_0) = \infty$.
✓ If $\theta < \theta_0$ then $h(\theta;\theta_0) = 0$.

This statement was proved for the fractional Brownian motion and the fractional Ornstein-Uhlenbeck obtained from the solution of a Langevin equation. For other fractional Gaussian process the previous statement is still to be proven. However the simulation study carried out shows that it seems to be true.

Asymptotics results for Gaussian process

The asymptotic behaviour of the quantization error for Gaussian process has been well studied by Luschgy and Pagès [2004a] among others. However for the Gaussian process X with covariance function $\Gamma(t) = e^{-\theta t^2}$ it does not exist any asymptotic result. Yet this covariance shape is often used by practitioners for kriging. In general if the eigenvalues of a Gaussian process satisfies that $\lambda_j \sim \varphi(j)$ as $j \to \infty$, where $\varphi : (s,\infty) \to (0,\infty)$ is a decreasing, regularly varying function at infinity of index $-b < -1$ for some $s \geq 0$. Set, for every $x > s$,

$$\psi(x) \stackrel{\triangle}{=} \frac{1}{x\varphi(x)},$$

then

$$e_n(X) \sim \left(\left(\frac{b}{2}\right)^{b-1} \frac{b}{b-1}\right)^{1/2} \psi(\ln n)^{-1/2}.$$

But this result can not be used for X. Using the work of Widom [1964] it is easy to show that the eigenvalues $(\lambda_k)_{k\geq 1}$ associated to the covariance operator of X satisfies that $\ln \lambda_k \sim -k \ln k$.

Working with the specific form of the eigenvalues of the process X we provides some asymptotics for the quantization error. The general idea relies in the work of Luschgy and Pagès [2002].

As we mention before the covariance function $\Gamma(t) = e^{-\theta t^2}$ also appears in the framework of kriging. With the aim to provide more information about the kriging estimation we investigate the relation between optimal quantization and kriging. We construct the 5000-optimal quantizer of X for $\theta = 20$ and $\theta = 100$ and we use it to "approximate" some data recollection. The results are not better than those obtained by the kriging version. However they could be used for other purposes.

The study of the maximal radius in finite dimension due to Sagna [2008] was the initial point to the study of the infinite dimensional case (specifically for the Wiener process). If $(\alpha_n)_{n\geq 1}$ is the optimal sequence of n-quantizers for the Wiener process then the maximal radius sequence is defined for $n \geq 1$ by

$$\rho_n = \rho(\alpha_n) = \max\{\|a\|_{\mathcal{L}^2}, a \in \alpha_n\}.$$

In finite dimension some lower and upper bounds are provided in Sagna [2008].

We obtain some bounds for a kind of stationary quantizers (designs II, III and IV in Luschgy et al. [2010]) using the result of Sagna and the decomposition of these quantizers on the Karhunen-Loève basis. We also work with a linear interpolation of W in $[0,1]$ using a time partition $\pi = \{0 = t_0 < t_1 < \cdots < t_{k-1} < t_k = 1\}$

$$W_L^{\pi}(t) = \sum_{i=0}^{k-1} \left[\frac{t_{i+1} - t}{\Delta_i} W(t_i) + \frac{t - t_i}{\Delta_i} W(t_{i+1}) \right] \mathbf{1}_{]t_i, t_{i+1}]}(t),$$

where $\Delta_i = t_{i+1} - t_i$, $\forall\, i = 1, 2, \cdots, k-1$.

For this process some bounds are provided for the maximal radius.

Using two different approaches we present some ideas about the asymptotics for the maximal radius associated to the optimal n-quantizers of the Wiener process.

Chapter 1

Gaussian process. Karhunen-Loève decomposition

In this chapter we summarize without proof the elementary properties of classical Gaussian processes including some basic facts about simulation methods. We emphasize on the expansion of Gaussian processes in the Karhunen-Loève basis. The Cameron-Martin space and the covariance operator are also discussed.

1.1 Short overview on Gaussian stochastic process

This section is a short review of some classical Gaussian processes where we recall their general properties and definitions.

1.1.1 Wiener process

The Wiener process (also called Brownian motion) is one of the best known stochastic process. It was named after the English biologist Robert Brown in the 1820s. It was originated with the study of the irregular movement of small pollen particles suspended in a liquid observed by Brown. It is often called Wiener process motion after Norbert Wiener (1923), who began developing the mathematical theory of it. This process is a martingale, a strong Markov process, a process with independent and stationary increments and a Gaussian process.

The Brownian motion plays an important role in pure and applied mathematics. This process is the starting point of stochastic calculus (integration with respect to Brownian motion, diffusion processes, etc.) and it is crucial to describe more complicated stochastic processes. It is considered as a very good approximation to many real-life phenomena: stochastic modeling for a wide variety of processes in physics (statistical mechanics, filtering and control theory), biology (e.g. population dynamics, migration, disease spreading). In finance it is an useful tool for common stock prices, in particular the Black-Scholes option pricing model.

Definition* 1.1 *Wiener process (Brownian motion)
A Gaussian stochastic process $\{W_t,\, t \geq 0\}$ *is a Wiener process if*

1. $W_{t_0} = W_0 = 0$.
2. *The process* W *has "stationary increments", i.e. the law of* $W_{t+h} - W_t$ *does not depend on* t.
3. *The process* W *has "independent increments" with* $W_t - W_s \overset{\mathcal{D}}{=} W_{t-s} \sim N(0, t-s)$, *where* $\overset{\mathcal{D}}{=}$ *means same distribution.*
4. *The process* W *has almost surely continuous path.*

Next proposition summarize some of the more important properties for this process.

***Proposition* 1.1**
Let W *be a Wiener process*

1. $\mathbb{E}(W_t) = 0$ *and* $\operatorname{cov}(W_t, W_s) = t \wedge s$.
2. *The process* W *is a Markov process and a continuous martingale.*
3. *The process* W *is nowhere differentiable in* $\mathbb{R}$.
4. *It has the following marginal distribution*

$$\mathbb{P}(W_t < s) = \int_{-\infty}^{s} \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} dx.$$

5. *The following processes are also Wiener processes*

   Self-similarity property: *If* $a > 0$ *then* $a^{-1} W_{a^2 t} \overset{\mathcal{D}}{=} W_t$.

   Inversion principle: $W_{\frac{1}{t}} \mathbf{1}_{\{t>0\}} + W_0 \mathbf{1}_{\{t=0\}} \overset{\mathcal{D}}{=} W_t$.

   Reflecting principle: $-W_t \overset{\mathcal{D}}{=} W_t$.

   Differential property: $W_{t+t_0} - W_{t_0} \overset{\mathcal{D}}{=} W_t$, *with* t_0 *fixed.*

6. *Variations:*
   If $a = t_0 < t_1 < \cdots < t_n = b$ *is a time partition of* $[a, b]$ *and* $\triangle t = \max\{t_{i+1} - t_i, i = 0, 1, \cdots, n-1\}$, *then*
   - *Variation:*
   $$\lim_{\triangle t \to 0} \sum_{i=0}^{n-1} |W_{t_{i+1}} - W_{t_i}| = +\infty \quad (a.s.).$$
   - *Quadratic variation:*
   $$\lim_{\triangle t \to 0} \sum_{i=0}^{n-1} |W_{t_{i+1}} - W_{t_i}|^2 = b - a \quad (\mathbb{P}).$$

1.1.2 Brownian bridge

Here we present some preliminaries for the Gaussian process called Brownian bridge. It is a continuous-time stochastic process that is equal in distribution to a Brownian motion on $[0, 1]$ that is restricted to hit 0 at time 1. An important application of this process is in the nonparametric Kolmogorov-Smirnov statistical hypothesis test. In particular, for large samples, the normalized difference between the empirical distribution and the true distribution is approximately the maximum of a Brownian bridge.

The process B is called a Brownian bridge if it satisfies the following stochastic differential equation:

$$\begin{aligned} dB_t &= -\frac{B_t}{1-t}dt + dW_t, \qquad \forall\, t \in [0,1) \\ B_1 &= 0, \end{aligned}$$

where W is a Wiener process. The equation above has an explicit stochastic integral solution:

$$B_t = (1-t)\int_0^t \frac{1}{1-s} dW_s. \tag{1.1}$$

Some important properties of the Brownian bridge B are presented in the following proposition

***Proposition* 1.2**
Let B be a Brownian bridge in $[0, 1]$

1. *$E(B_t) = 0$.*
2. *$cov(B_t, B_s) = s(1-t)$, for $0 \leq s \leq t$.*
3. *The Brownian bridge is not a self-similar process.*
4. *The following processes are also Brownian bridge.*
   - $X_t = W_t - tW_1 \overset{\mathcal{D}}{=} B_t$.
   - $X_t = W_{1-t} - (1-t)W_1 \overset{\mathcal{D}}{=} B_t$.

1.1.3 Fractional Brownian motion

The fractional Brownian motion (fBm) is a continuous-time Gaussian process which has an important place among self-similar processes (see for instance Samorodnitsky and Taqqu [1994]). It was first introduced by Kolmogorov [1940] as a way to generate Gaussian spirals in a Hilbert space. It is the only self-similar Gaussian processes with stationary increments. This process is characterized by a single parameter: the Hurst index H. It is worth pointing out that H is linked to different properties of the fBm as are the smoothness of the sample paths, the long-range dependence of its increments and the self-similarity.

The fractional Brownian motion is mostly used for modeling long-range dependent stochastic processes with continuous path. There are a wide class of fields where we can find a fBm, for example in finance, more precisely in the long-range dependence of

stock returns on the Black-Sholes formula, arbitrage and many others (see for instance Cheridito [2003] and Doukhan et al. [2003] for more details). The fBm is also related with applications on real life like image processing, internet traffic, hydrology, etc. The formal definition is

Definition* 1.2 *Fractional Brownian motion
A stochastic process $\{W_t^H,\, t \geq 0\}$ *is a fractional Brownian motion with Hurst parameter* $H \in (0,1)$ *if*

1. $W_0^H = 0$.
2. *It is a Gaussian process with* $\mathbb{E}(W_t^H) = 0$ *and* $\mathbb{E}(W_t^H)^2 = t^{2H}$, *for all* $t \geq 0$.
3. *The process* W^H *has stationary increments.*
4. *The process* W^H *has almost surely continuous path.*

The most importants properties associated to the fractional Brownian motion are resumed in the next proposition. For a deeper discussion we refer the reader to Biagini et al. [2008a,b].

***Proposition* 1.3**
Let W^H *be a fractional Brownian motion.*

1. *The covariance function of* W^H *is defined by*

$$\Gamma_{fBm}(t,s) = \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right).$$

2. *The fractional Brownian motion is* H*-selfsimilar for all* $a > 0$*:* $a^{-H} W_{at}^H \overset{\mathcal{D}}{=} W_t^H$.
3. *When* $H = 0.5$ *the fBm coincides with the standard Wiener process.*
4. *The fBm is not a semimartingale for* $H \neq 0.5$.
5. *The fractional Brownian motion is long range dependent (the covariance function has a power decay) if* $H \in (1/2, 1]$ *and short range dependent (the covariance function has an exponential decay) if* $H \in [0, 1/2)$.

1.1.4 Ornstein-Uhlenbeck

The Ornstein-Uhlenbeck (OU) process was introduced by Leonard Ornstein and George Eugene Uhlenbeck in 1930 in a physical modeling context. It is widely used for modeling a mean reverting process (the process tends to drift towards its long-term mean). The OU process is defined as the unique strong solution of the following stochastic differential equation (SDE)

$$dX_t \;=\; -\theta(X_t - \mu)dt + \sigma dW_t, \tag{1.2}$$

where θ, μ, σ are the parameters of the SDE and W is a Wiener process. The previous SDE also corresponds to a particular case of the Langevin equation for the Brownian motion. The behaviour of the solution depends on the initial value of the system. If $X_0 = c$ is constant, then we are in the presence of the classical Ornstein-Uhlenbeck process. In the

other hand if $X_0 \sim N(\mu, \sigma^2/2\theta)$ then X is a stationary Gaussian process. For more detail we refer the reader to Karatzas and Shreve [1991].

The solution of the equation (1.2) is explicitly given by the stochastic integral

$$X_t = e^{-\theta t}\left(X_0 - \mu(1-e^{\theta t}) + \sigma\int_0^t e^{\theta s} dW_s\right). \tag{1.3}$$

Some basic properties of X are easily derived from (1.2) and (1.3)

***Proposition* 1.4**
If X is the solution of the equation (1.2) then

1. *X is a stationary Gaussian process if $X_0 \sim N(\mu, \sigma^2/2\theta)$ with*
   - $E(X_t) = \mu$.
   - $cov(X_t, X_s) = \frac{\sigma^2}{2\theta} e^{-\theta|t-s|}$.
2. *If $X_0 = c$, constant*
   - $E(X_t|X_0 = c) = \mu + (c-\mu)e^{-\theta t}$.
   - $cov(X_t, X_s) = \frac{\sigma^2}{2\theta}(e^{-\theta|t-s|} - e^{-\theta|t+s|}) = \frac{\sigma^2}{2\theta} e^{-\theta|t+s|}(e^{2\theta(t\wedge s)} - 1)$.

From now on we focus on the stationary Ornstein-Uhlenbeck with $\mu = 0$. The expression in (1.3) is easily written following Kaarakka and Salminen [2007]

$$X_t = e^{-\theta t}\left(\sigma\int_{-\infty}^t e^{\theta s} dW_s\right).$$

The stationary OU with $\mu = 0$ can be also obtained by a Lamperti transformation of the Brownian motion

$$X_t = e^{-\theta t} W_{\alpha e^{2\theta t}}, \tag{1.4}$$

where $\alpha = \frac{\sigma^2}{2\theta}$.

1.1.5 Fractional Ornstein-Uhlenbeck processes

There are a wide extensions of Ornstein-Uhlenbeck process. We will restrict the discussion to three types of generalized OU, the fractional ones (fOU): the process obtained from the solution of a Langevin equation, the stationary process by applying the Lamperti transformation to fBm and the process with stretched exponential covariance (or powered exponential covariance). All these processes depend on the Hurst parameter $H \in (0, 1)$.

The best general reference here can be found in Kaarakka and Salminen [2007] and Cheridito et al. [2003]. For further details see for instance Lim and Muniandy [2003] (spectral representation) and Yan et al. [2008].

fOU of first kind (fOU(1))

In the sequel we follow Kaarakka and Salminen [2007] and Cheridito et al. [2003]. The fractional Ornstein Uhlenbeck of first kind ($X^{H,1}$) is an extension of the Ornstein-Uhlenbeck process with fractional Brownian motion driving term. We consider the fractional analogue of the Ornstein-Uhlenbeck process, i.e. the solution of a one-dimensional

homogeneous linear stochastic differential equation driven by a fractional Brownian motion in place of the usual Brownian motion. In Cheridito et al. [2003] they show that the solution of a Langevin equation with fractional Brownian motion noise is stationary, unique and continuous.

This kind of process have applications mainly in the field of finance due to the long memory property of the process (one-factor short-term interest rate model, arbitrage free pricing formulas for European options, among others). In finance it is also known as the fractional Vasicek model. We refer the reader to the work of Høg and Frederiksen [2006].

The solution of a stochastic differential equation driven by a fractional Brownian motion

$$dX_t^{H,1} = -\theta(X_t^{H,1} - \mu)dt + \sigma dW_s^H,$$

is called fOU of first kind $X_t^{H,1}$ with $\mu = 0$ (Salminen, 2007) and is given by

$$X_t^{H,1} = e^{-\theta t}\left(\sigma \int_{-\infty}^{t} e^{\theta s} dW_s^H\right). \tag{1.5}$$

The covariance function associated satisfies the following

***Lemma* 1.1**
Let $H \in (0, \frac{1}{2}) \cup (\frac{1}{2}, 1]$ and $N = 1, 2, \cdots,$. Then for a fixed $t \in \mathbb{R}$ and $s \to +\infty$

$$\mathbb{E}(X_t^{H,1}, X_{t+s}^{H,1}) = \frac{\sigma^2}{2}\sum_{n=1}^{N}\theta^{-2n}\left(\prod_{k=0}^{2n-1}(2H-k)\right)s^{2H-2n} + O(s^{2H-2N-2}).$$

It is straightforward that the decay of the covariance function is like that of a power function (for details see Cheridito et al. [2003]). In Yan et al. [2008] the following bounds are provided. In particular they are useful in the proof of Theorem 5.1.

***Lemma* 1.2**
Let $H \in (0, 1)$, $\theta = 1$ and $\mu = 0$ then

$$c_H\sigma^2\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right) \leq \mathbb{E}(X_t^{H,1}, X_s^{H,1}) \leq C_H\sigma^2\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right), \tag{1.6}$$

for certain constants $c_H, C_H > 0$ and all $0 < s < t < 1$.

The following lemma concerns to the spectral density of the process

***Lemma* 1.3**
For $H \in (0, 1)$ and $\mu = 0$ the spectral density of $X^{H,1}$ satisfies

$$f_H(\lambda) = \theta^2(\theta^2 + \lambda^2)^{-(H+1/2)}.$$

The process obtained as the solution of the Langevin SDE with fBm as the driving process does not coincide with the process obtained by Lamperti transformation of the fractional Brownian motion due to Doob (see for instance Doob [1942]). For more details we refer the reader to Cheridito et al. [2003] and Kaarakka and Salminen [2007] and references therein. The most significative difference comes from the decay of the covariance function. It exhibits a power decay for the fOU(1) and an exponential decay for the fOU of second kind (fOU(2)).

fOU of second kind (fOU(2))

The second fractional Ornstein-Uhlenbeck $X^{H,2}$ is defined by a Lamperti transformation as

$$X_t^{H,2} = e^{-\theta t} W^H_{a_{t,H}}, \tag{1.7}$$

where $a_{t,H} = He^{\theta t/H}/\theta$. It holds

***Lemma* 1.4**
The covariance function for $X^{H,2}$ when $t > s$ satisfies

$$\begin{aligned}\mathbb{E}(X_t^{H,2}, X_s^{H,2}) &= \frac{1}{2}\left(\frac{H}{\theta}\right)^{2H}\left(e^{\theta(t-s)} + e^{-\theta(t-s)} - e^{\theta(t-s)}\left(1 - e^{-\frac{\theta(t-s)}{2H}}\right)^{2H}\right)\\ &= \left(\frac{H}{\theta}\right)^{2H}\left(\cosh(\theta(t-s)) - 2^{2H-1}\sinh^{2H}\left(\frac{\theta(t-s)}{2H}\right)\right)\\ &= \frac{1}{2}\left(\frac{H}{\theta}\right)^{2H}\left(e^{-\theta(t-s)} + \sum_{n=1}^{\infty}(-1)^{n-1}\binom{2H}{n}e^{-\frac{\theta n(t-s)}{2H}}\right).\end{aligned}$$

Concerning the spectral density we have

***Lemma* 1.5**
For $H \in (0,1)$, $\sigma = 1$ and $\mu = 0$ the spectral density of $X^{H,2}$ satisfies

$$f_H(\lambda) \sim \lambda^{-(2H+1)}.$$

fOU of third kind (fOU(3))

The next formulation of the fOU(3) comes from Bogachev [1998].

***Definition* 1.3**
The process $X^{H,3}$ is a fractional Ornstein-Uhlenbeck process of third kind if its covariance function is defined by

$$\mathbb{E}(X_t^{H,3}, X_s^{H,3}) = e^{-\theta|t-s|^{2H}}.$$

Using the Fourier transform and the statement on page 1584 in Luschgy and Pagès [2004a] it follows that

***Lemma* 1.6**
For $H \in (0,1)$ and $\mu = 0$ the spectral density of $X^{H,3}$ satisfies

$$f_H(\lambda) \sim \rho(H)\lambda^{-(2H+1)},$$

where $\rho(H) = \pi^{-1}\Gamma(2H+1)\sin(H\pi)$ and $f_H(\lambda)$ is known for $H = 1/2$ and $H = 1$ and the expressions are

$$f_{1/2}(\lambda) = \theta^2(\theta^2 + \lambda^2)^{-(H+1/2)}, \qquad f_1(\lambda) = \frac{1}{2}\sqrt{\frac{\pi}{\theta}}e^{-\frac{\lambda^2}{4\theta}}.$$

The fOU(3) $X^{H,3}$ is a stationary Gaussian process and the covariance function associated is known as stretched exponential or "powered exponential". It has applications in the theory of kriging, geostatistics and fractals. (see Diggle et al. [1998], Gelfand et al. [2010] and Gneiting et al. [2011]).

1.2 Karhunen-Loève decomposition

Throughout the section $(H, \langle \cdot, \cdot \rangle)$ is a separable Hilbert space. Typically H is $\mathcal{L}^2 \stackrel{\triangle}{=} \mathcal{L}^2([0,1], dt)$ with the usual norm $||X||^2_{\mathcal{L}^2} = \int_0^1 X^2(t)dt$. This is the space where the processes of Section 1.1 live. Here we consider Gaussian processes with values in $\mathcal{L}^2$.

A Hilbert space has always an orthonormal basis. Let $\{\psi_k\}_{k\geq 1}$ be an orthonormal basis for $\mathcal{L}^2$. It is well-known that any $X \in \mathcal{L}^2$ (centered) can be represented into an infinite expansion on $\{\psi_k\}_{k\geq 1}$, i.e.:

$$X(t) = \sum_{k\geq 1} \rho_k \psi_k(t), \tag{1.8}$$

where

$$\rho_k = \langle X(t), \psi_k(t) \rangle = \int_0^1 X(t)\psi_k(t)dt,$$

with the following property

$$\mathbb{E}(\rho_i \rho_j) = \int_0^1 \int_0^1 \psi_i(s) C_X(s,t) \psi_j(t) ds dt,$$

and $C_X(s,t)$ the covariance function, $C_X(s,t) = EX(s)X(t)$.

The equality in (1.8) holds in the mean-square sense, i.e.,

$$\lim_{n\to+\infty} \mathbb{E}\left\| X(t) - \sum_{k=1}^{n} \rho_k \psi_k(t) \right\|^2 = 0,$$

and the convergence is uniform in t.

More general classes of series expansions of stochastic process can be found by instance in Masry et al. [1968]. Among all these representations that can be chosen in a Hilbert space we will focus our attention on the so-called "Karhunen-Loève"(KL) expansion which can be seen in some sense as some infinite dimensional *Principal Component Analysis* of the process itself.

Let us recall the formal definition of the covariance function as well some important properties.

Definition* 1.4 *Covariance function
For a zero mean Gaussian process X defined in H the covariance function is a positive semidefinite Hermitian kernel and

$$C_X(s,t) = cov(X(s), X(t)) = \mathbb{E}(X(t)X(s)).$$

For the stationary stochastic process the correlation function is an useful tool. It can be represented as a Fourier or Fourier-Stieltjes integral.

Definition* 1.5 *Spectral representation
Let $C_X(s,t)$ be the covariance function of a stationary stochastic process X, then

$$C_X(s,t) = C_X(t-s),$$

and the correlation function can be expressed by

$$\gamma(t) = \frac{C_X(t)}{C_X(0)} = \int_{\mathbb{R}} e^{2\pi i\lambda t} d\mu(\lambda),$$

where $\mu(\lambda)$ is the spectral measure associated to X. When $\mu(\lambda)$ admits a density f the previous formula is rewritten as

$$\gamma(t) = \int_{\mathbb{R}} e^{2\pi it\lambda} f(\lambda) d\lambda.$$

The function $f(\lambda)$ (called spectral density) is the Fourier transform of $\gamma(t)$, i.e.

$$f(\lambda) = \int_{\mathbb{R}} e^{-2\pi i\lambda t} \gamma(t) dt.$$

Let us mention some properties of the covariance function.

***Lemma* 1.7**
If X is a Gaussian process defined in the Hilbert $H = \mathcal{L}^2([0,1], dt)$ with covariance function C_X then

1. $C_X(s,t) = C_X(t,s)$.
2. $\int_0^1 \int_0^1 |C_X(s,t)|^2 dsdt < +\infty$.
3. $C_X(s,t)$ *is positive semi definite, i.e.:*

$$\int_0^1 \int_0^1 f(t) C_X(s,t) f(s) dsdt \geq 0.$$

4. *There exists an orthonormal basis $\{\varphi_k\}_{k\geq 1}$ of H for which $C_X(s,t)$ has the following decomposition*

$$C_X(s,t) = \sum_{k\geq 1} \lambda_k \varphi_k(t) \varphi_k(s),$$

where $\lambda_k = \mathbb{E}\rho_k^2$.

Proof.

1. Is straightforward using definition.
2. $\int_0^1 \int_0^1 |C_X(s,t)|^2 dsdt \leq \mathbb{E}||X(t)||^2 \mathbb{E}||X(s)||^2 < +\infty$.
3. $\int_0^1 \int_0^1 f(t) C_X(s,t) f(s) dsdt = \mathbb{E}\left(\left| \int_0^1 X(t) f(t) dt \right|^2 \right) \geq 0$.
4. It follows as a straightforward application of Mercer's theorem.

□

1.2.1 Karhunen-Loève expansion

The Karhunen-Loève expansion plays a fundamental role in the functional quantization framework. This representation is related with the covariance function of the process.

Definition* 1.6 *Karhunen-Loève basis
For a zero mean Gaussian process X defined in H the orthonormal basis $\{\varphi_k\}_{k\geq 1}$ for which

$$C_X(s,t) = \sum_{k\geq 1} \lambda_k \varphi_k(t)\varphi_k(s),$$

with $\lambda_k = \mathbb{E}\rho_k^2$ is called the Karhunen-Loève basis for the process X and the unique representation of X in that system is

$$X(t) \overset{\mathcal{L}^2}{=} \sum_{k\geq 1} \sqrt{\lambda_k}\xi_k\varphi_k(t),$$

where the sequence of $(\xi_k)_{k\geq 1}$ are independent standard normal random variables, and the sequence of $(\lambda_k)_{k\geq 1}$ is the sequence (in decreasing order) of the eigenvalues associated to an orthonormal basis of eigenvectors, also $\lambda_k = V(\langle X, \varphi_k\rangle)$ and

$$\sum_{k\geq 1} \lambda_k = \mathbb{E}||X||^2.$$

The sequence $\{\varphi_k\}_{k\geq 1}$ is a complete orthonormal basis of $\mathcal{L}^2$. The system $(\varphi_k(t), \lambda_k)_{k\geq 1}$ is crucial for the quantization designs on functional quantization (see for instance Luschgy et al. [2010]).

For all stochastic Gaussian zero mean process X with $\mathbb{E}||X||^2_{\mathcal{L}^2} < +\infty$ there exists a Karhunen-Loève basis. It is a hard task to find the exact expression or some approximations for the eigenfunctions and eigenvalues of the covariance operator. Only in a few cases (if we center our attention on the classical Gaussian processes) these exact expressions are know. This is the case of the Wiener process and Brownian bridge. Asymptotic formulae have been established for the fractional Brownian motion. In the case of the Ornstein-Uhlenbeck process one can find a semi closed form. We briefly sketch here these results

Wiener process

If W is a Wiener process in $[0,T]$ then

$$\varphi_k^W(t) = \sqrt{\frac{2}{T}}\sin\left(\frac{(2k-1)t}{T}\frac{\pi}{2}\right), \qquad \lambda_k^W = \left(\frac{2T}{\pi(2k-1)}\right)^2, \qquad k = 1, 2, \cdots, .$$

For the Wiener process the Karhunen-Loève expansion converges in a much stronger sense:

$$\sup_{[0,T]}\left|W_t - \sum_{k=1}^{n}\sqrt{\lambda_k^W}\xi_k\varphi_k^W(t)\right| \;\to\; 0 \quad \mathbb{P}-a.s.,$$

and

$$\left\| \sup_{[0,T]} \left| W_t - \sum_{k=1}^{n} \sqrt{\lambda_k^W} \xi_k \varphi_k^W(t) \right| \right\|_2 = O(\sqrt{\log n / n}),$$

where $||X||_2 = (\mathbb{E}|X|_H^2)^{\frac{1}{2}}$.

Brownian bridge

If B is a Brownian bridge in $[0,T]$ then

$$\varphi_k^B(t) = \frac{2}{T} \sin\left(\frac{t}{T}\pi k\right), \qquad \lambda_k^B = \left(\frac{T}{\pi k}\right)^2, \qquad k = 1, 2, \cdots, .$$

Ornstein-Uhlenbeck

If we take $\mu = 0$ in (1.2) on $[0,T]$

$$\varphi_k^X(t) = \left(\frac{T}{2} - \frac{\sin(2\omega_{\lambda_k^X} T)}{4\omega_{\lambda_k^X}}\right)^{-1/2} \sin(\omega_{\lambda_k^X} t), \qquad \lambda_k^X = \frac{\sigma^2}{\omega_{\lambda_k^X}^2 + \theta^2}, \qquad k = 1, 2, \cdots,$$

where $(\omega_{\lambda_k^X})_{k \geq 1}$ are the strictly positive solutions of the equation

$$\theta \sin(\omega_{\lambda_k^X} T) + \omega_{\lambda_k^X} \cos(\omega_{\lambda_k^X} T) = 0.$$

For details we refer the reader to Corlay and Pagès [2010].

1.2.2 Covariance operator

The covariance operator of a stochastic process X is diagonalized in the Karhunen-Loève expansion of X. This concept with the Cameron-Martin space defines the "spatial localization" of stationary quantizers and are crucial in the numerical design of stationary quantizers.

Let us define the covariance operator $\Gamma_C : H \to H$ associated to C_X by

$$\Gamma_C f(t) = \int_0^1 C_X(s,t) f(s) ds = \mathbb{E}[\langle X, f \rangle X(t)], \tag{1.9}$$

where $f \in H$.

The basic properties of the covariance operator can be resumed in the following lemma.

***Lemma* 1.8**
If X is a zero mean Gaussian process with covariance function C_X and also

$$\mathbb{E}||X||_{\mathcal{L}^2}^2 < +\infty,$$

then the covariance operator Γ_C shares the following properties:

1. *Γ_C is compact self-adjoint operator and positive semidefinite.*
2. *Γ_C is a Hilbert-Schmidt integral operator.*
3. *Spectral representation,*
   - *The covariance operator has a spectral representation in the basis $\{\varphi_k\}_{k\geq 1}$, i.e.:*

$$\Gamma_C(f) = \sum_{k\geq 1} \lambda_k \langle f, \varphi_k(t)\rangle \varphi_k(t). \tag{1.10}$$

   - *The basis $\{\varphi_k\}_{k\geq 1}$ are the eigenfunctions of Γ_C with the corresponding eigenvalues $(\lambda_k)_{k\geq 1}$, i.e.,*

$$\Gamma_C \varphi_k(t) = \int_0^1 C_X(s,t)\varphi_k(s)ds = \lambda_k \varphi_k(t).$$

Proof.

1. To prove that Γ_C is positive semidefinite we need to show that $\langle \Gamma_C X, X\rangle \geq 0,\ \forall\, X \in H$ and that is a direct conclusion using properties of C_X, specifically it follows by 3 in Lemma 1.7 that

$$\begin{aligned} \langle \Gamma_C X, X\rangle &= \int_0^1 \Gamma_C X(s) X(s) ds \\ &= \int_0^1 \int_0^1 C_X(s,t) X(s) X(t) ds dt \geq 0. \end{aligned}$$

   Using a similar argument, this time 1 in Lemma 1.7, it is straightforward that Γ_C is self-adjoint.

   The demonstration of compactness can be found in Gohberg and Goldberg [1981], after that it follows that Γ_C is a linear bounded operator.
2. This assertion is a straightforward application of Theorem 6.11 in Weidmann [1980].
3. The proof is derived from the general properties of Hilbert-Schmidt operators:
   - Γ_C is a Hilbert-Schmidt integral operator, therefore using a convenient spectral theorem, (see for instance Corollary 4.10.2 in Debnath and Mikusiński [2005]) we can assure that there exists $\{\varphi_k\}_{k\geq 1}$ an orthonormal basis of H, (eigenvectors of Γ_C) with the correspond eigenvalues $(\lambda_k)_{k\geq 1}$ that satisfies (1.10).
   - Taking $f = \varphi_k$ for any fix $k \in \mathbb{N}$ in (1.10) it is easy to check that

$$\Gamma_C \varphi_k(t) = \lambda_k \varphi_k(t).$$

□

1.2.3 Cameron-Martin space

The Cameron-Martin space or *the reproducing kernel Hilbert space* can be defined for any Gaussian process X with values in a Hilbert H. This space is denoted by K_X and is a linear subspace of H. In the functional quantization framework plays a crucial role: it is the space where optimal and stationary quantizers of Gaussian process live (see for

instance Luschgy and Pagès [2002]). The Cameron-Martin space is defined through the covariance operator of a Gaussian process and it is closely related to their Karhunen-Loève expansion.

Definition* 1.7 *Cameron-Martin space
The Cameron-Martin space K_X is a linear space of functions defined by:

$$K_X \stackrel{\triangle}{=} \left\{\mathbb{E}(ZX) : Z \in \mathrm{cl}_{L^2(\mathbb{P})}\{\langle y, X\rangle : y \in H\}\right\},$$

where cl_A is the closure of A.

This space can always be defined for any stochastic Gaussian process. It also has the reproducing property.

***Lemma* 1.9**
The Cameron-Martin space K_X for the stochastic process X equipped with the inner product

$$\langle k_1, k_2\rangle_{K_X} = \langle Z_1, Z_2\rangle_H = \mathbb{E}(Z_1 Z_2) \quad \textit{where} \quad k_i = \mathbb{E}(Z_i X) \in K_X, \quad i = 1, 2,$$

is a reproducing kernel Hilbert space of the covariance operator of X_t.

Proof.

Using the definition of Γ_C in (1.9) it follows that

$$\begin{aligned}\langle k_1, k_2\rangle_{K_X} &= \langle \mathbb{E}(\langle X, z_1\rangle X), \mathbb{E}(\langle X, z_2\rangle X)\rangle_{K_X} = \mathbb{E}(\langle X, z_1\rangle\langle X, z_2\rangle)\\ &= \langle \mathbb{E}(\langle X, z_1\rangle X), z_2\rangle.\end{aligned}$$

The last assertion comes from Fubini's theorem and it yields to the *reproducing property*:

$$\langle k, \Gamma_C(y)\rangle_{K_X} = \langle k, y\rangle \qquad \text{where} \quad k \in K_X \quad \text{and} \quad y \in H.$$

□

For more information about general properties of the Cameron-Martin space we refer the reader to Janson [2008].

1.3 Simulation of Gaussian process

This section is devoted to the simulation of Gaussian process. In the sequel, π is a time partition of the interval $[0, T]$ namely

$$\pi = \{0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = T\},$$

where $\Delta_i = t_{i+1} - t_i$, $\forall\, i = 0, 1, 2, \cdots, n-1$. In general we take $\Delta_i = n^{-1}$ for all i and $T = 1$. Throughout this section, for any Gaussian process X, we generate the random variables $X(t_1), \cdots, X(t_n)$. Using a linear interpolation we construct the process X_L^π as follows

$$X_L^\pi(t) = \sum_{i=0}^{n-1} \left[\frac{t_{i+1} - t}{\Delta_i} X(t_i) + \frac{t - t_i}{\Delta_i} X(t_{i+1})\right] \mathbf{1}_{]t_i, t_{i+1}]}(t),$$

Brownian motion

The most simple method for generate one path of a Brownian motion relies in the definition of the process itself. More precisely in the property of stationary and independents increments (random walks construction). For the Wiener process W we write

$$W_{t_i} = \sum_{j=1}^{i}(W_{t_j} - W_{t_{j-1}}) = \sum_{j=1}^{i} Z_j, \qquad i = 1, 2, \cdots, n, \quad W_{t_0} = 0,$$

where $Z_j \sim N(0, t_j - t_{j-1})$ for all $j = 1, 2, \cdots, n$. It is straightforward that to obtain a Brownian motion in the points of π it suffices to generate n random variables $N(0, 1)$. This method is exact in the sense that the simulated values coincides with the join distribution of W in π. The drawback of this method is that if we change the partition a new generation of normal random variables is needed. Others simulation methods for the Wiener process can be seen for instance in Glasserman [2004] and Asmussen and Glynn [2007].

We emphasize that the simulation of the Wiener process is crucial to obtain simulated versions for the Brownian bridge and the Ornstein-Uhlenbeck process.

Brownian bridge

The simulation of a Brownian bridge is straightforward if we know how to generate a Wiener process.

1. By Lemma 4 it follows

$$B_{t_i} = W_{t_i} - t_i W_{t_n}, \qquad \forall i = 1, \cdots, n.$$

2. Using the expression of the solution in the SDE on (1.1) we obtain

$$\begin{aligned} B_{t_i} &= (1 - t_i)\int_0^{t_i} \frac{1}{1-s} dW_s = (1 - t_i)\sum_{j=1}^{i}\int_{t_{j-1}}^{t_j} \frac{1}{1-s} dW_s \\ &\approx (1 - t_i)\sum_{j=1}^{i} \frac{1}{1-t_j}\Delta W_{t_j} = (1 - t_i)\sum_{j=1}^{i} \frac{W_{t_j} - W_{t_{j-1}}}{1-t_j} \qquad \forall i = 1, \cdots, n. \end{aligned}$$

The Brownian bridge can be also constructed using the previous development but in a compact autoregressive form

$$B_{t_i} \approx \frac{1 - t_i}{1 - t_{i-1}} B_{t_{i-1}} + \Delta W_{t_i}, \qquad \forall i = 0, 1, \cdots, n.$$

The first method is the most used for its simplicity.

Ornstein-Uhlenbeck

Several methods can be used to simulate the Ornstein-Uhlenbeck process. We focus on methods for the stationary case with $\mu = 0$. By (1.3) it is easy to check that

$$X_{t_i} \approx e^{-\theta t_i}\left(X_0 + \sigma \sum_{j=1}^{i} e^{\theta t_j}\Delta W_{t_j}\right), \qquad \forall i = 0, 1, \cdots, n,$$

where $\Delta W_{t_j} = W_{t_j} - W_{t_{j-1}}$.

A method based on the Lamperti transformation of the Wiener process (see (1.4)) can be used

$$X_{t_i} = \sigma \frac{e^{-\theta t_i}}{\sqrt{2\theta}} W_{e^{2\theta t_i}}, \qquad \forall i = 0, 1, \cdots, n.$$

Fractional Brownian motion

There are a wide class of methods to simulate the fractional Brownian motion. The existing methods can be classified as exact or approximative.

The first group is based on the knowledge of the covariance function. There is the one based on Cholesky decomposition of the covariance matrix. This is an exact method which simulate the discrete fBm. This becomes untractable when the partition becomes very large (several thousands points) due to the amount of storage and CPU time. The method proposed by Hosking (also known as the Durbin or Levinson method) is an algorithm to simulate a general stationary Gaussian process. The "fractional Gaussian noise" (fGn) is used with that purpose (see for details Dieker [2002]). The fGn $\{X_k : k = 0, 1, 2 \cdots, \}$ is defined by

$$X_k = W_{k+1}^H - W_k^H.$$

Another method was proposed by Davies and Harte [1987] (recent modifications of it can be found in Dieker [2002]). The drawback for the first two methods is that they are slow and they demand a considerable amount of storage. The third is the faster one.

In general the approximative methods provide good results and are less time consuming. These methods can be classified, in general, under *aggregation methods*. That include spectral techniques based on the analysis of a stochastic process in the so-called spectral or frequency domain. The integral representation of the fBm with finite sums; a wavelet-based simulation; the *random midpoint displacement* method, among many others. For a deeper discussion on approximative simulation methods for the fractional Brownian motion we refer the reader to Dieker [2002] and references therein. See also the work of Bardet et al. [2003].

We use an approximative method based on the representation of the fBm in the orthonormal basis of the path space. This approach is based on the covariance matrix of the fBm. Any Gaussian processes with values in a separable Hilbert space $\mathcal{L}^2([0,1], dt)$ has a unique representation in the Karhunen-Loève basis (see Section 1.2 for details).

$$W_t^H = \sum_{k \geq 1} \sqrt{\lambda_k^H} \xi_k \varphi_k^H(t),$$

where $\{\varphi_k^H\}_{k \geq 1}$ is the Karhunen-Loève basis of $\mathcal{L}^2$, the sequence of $(\xi_k)_{k \geq 1}$ are independent standard normal random variable, and $(\lambda_k^H)_{k \geq 1}$ is the sequence (in decreasing order) of the eigenvalues associated to this basis. For numerical purposes, we truncate the infinite expansion of the process given in the the expression before to obtain an approximation of it

$$W_{t_i}^H \approx \sum_{k\geq 1}^{N^\star} \sqrt{\widehat{\lambda_k^H}}\xi_k \widehat{\varphi_k^H}(t_i), \qquad \forall i = 0, 1, \cdots, n.$$

where $N^\star$ is a fixed number, $\widehat{\lambda_k^H}$ and $\widehat{\varphi_k^H}(t_i)$ are estimated by the covariance matrix for all $i = 1, 2, \cdots, n$ using the Cholesky factorization in the points of π. For any other point $t' \notin \pi$, $t' \in [0, 1]$ an interpolate value of $\widehat{\varphi_k^H}(t')$ can be calculated using a suitable interpolation function (linear or parabolic depending on n). If $t' \notin \pi$ and $t > 1$ the interpolation can not be used. The estimation of $\widehat{\varphi_k^H}$ has to be done in a new partition $\pi' \ni t'$. If $t' \gg 1$ then the size of π' need to be considerable large to obtain a good estimation of $\widehat{\varphi_k^H}$.

Fractional Ornstein-Uhlenbeck processes

For the fOU(1) and fOU(2) the simulation methods relies in the generation of a fractional Brownian motion. For the fOU(3) we use the Karhunen-Loève representation of the process to obtain a truncated version as for the fBm.

fOU(1) By (1.5) we get

$$X_{t_i}^{H,1} \approx e^{-\theta t_i}\left(X_{t_0}^{H,1} + \sigma \sum_{j=1}^{i} e^{\theta t_j} \Delta W_{t_j}^H\right), \qquad \forall i = 0, 1, \cdots, n,$$

where $\Delta W_{t_j}^H = W_{t_j}^H - W_{t_{j-1}}^H$.

fOU(2) Using the Lamperti transformation of the fBm in (1.7)

$$X_{t_i}^{H,2} = \left(\frac{H}{\theta}\right)^H e^{-\theta t_i} W_{e^{\theta t_i/H}}^H, \qquad \forall i = 0, 1, \cdots, n.$$

The drawback of this simulation method is that we need to generate a fBm in the interval $[1, e^{\theta/H}]$ instead of $[0, 1]$. As we mention before the simulation of the fBm could be a hard task. Using the self-similarity property of the fractional Brownian motion this issue can be avoided.

Let assume that we have a simulation of the fBm $(W_{t_0}^H, W_{t_1}^H, \cdots, W_{t_n}^H)$. By the self-similarity property it follows for $t_1, t_2, \cdots, t_n$ that

$$X_{t_i}^{H,2} = \left(\frac{H}{\theta}\right)^H t_i^{-H} W_{t_i}^H \qquad \forall i = 1, \cdots, n.$$

This approach can not be used for t_0. However this can be avoided by taking n large enough for which t_1 is near to zero.

fOU(3) As for the fBm we can write the Karhunen-Loève representation of the process as

$$X_t^{H,3} = \sum_{k\geq 1} \sqrt{\lambda_k^H}\xi_k\varphi_k^H(t),$$

where $\{\varphi_k\}_{k\geq 1}$ is the Karhunen-Loève basis of $\mathcal{L}^2$, the sequence of $(\xi_k)_{k\geq 1}$ are independent standard normal random variables and $(\lambda_k)_{k\geq 1}$ is the sequence (in decreasing order) of the eigenvalues associated to the KL basis. As before we use a truncated version to obtain one simulated process for the fOU(3)

$$X_{t_i}^{H,3} \approx \sum_{k\geq 1}^{N^\star} \sqrt{\widehat{\lambda_k^{X^{H,3}}}}\xi_k\widehat{\varphi_k^{X^{H,3}}}(t_i) \qquad \forall i = 0, 1, \cdots, n,$$

where $N^\star$ is a fixed number, $\widehat{\lambda_k^H}$ and $\widehat{\varphi_k^H}(t_i)$ are estimated by the covariance matrix for all $i = 1, 2, \cdots, n$.

Chapter 2

Quantization of Gaussian process

In this chapter we give a detailed exposition on the quantization topic. The first section summarizes without proof the most relevant facts on Vector Quantization theory. For a fuller approach we refer the reader to the work of Graf and Luschgy [2000]. The second section is devoted to functional quantization. We compile the general theory about optimal quantizers in infinite dimension and we focus on the quantization error. A more complete theory may be obtained in the extensive work of Luschgy and Pagès [2002], Luschgy et al. [2010] among others.

2.1 Vector quantization

"Quantization is the division of a quantity into a discrete number of small parts, often assumed to be integral multiples of a common quantity", or we can view quantization "as a partition problem of the underlying space".

The term "quantization" originates in the theory of signal processing in electrical engineering in the early 1950's. It has been conceived in order to drastically cut down the storage of signal data to be analyzed, the idea was to use a finite number of n codes (or quantizers) to transmit efficiently a continuous stationary signal. Since the late 1940's this field has been extensively investigated.

As a mathematical topic, the aim of quantization is to approximate a given probability measure by discrete probability measures with finite supports. More specifically, the quantization for probability distributions is related with the best approximation of a d-dimensional probability distribution $\mathbb{P}$ by a discrete probability distribution with finite support. The quantization search the best approximation of a d-dimensional random vector with probability distribution $\mathbb{P}$ by a random vector Y with at most n values in its image. The mathematical aspects of quantization are treated extensively in Graf and Luschgy [2000].

The scope of quantization includes different scientific fields. Beyond the classical applications in fields such as signal processing and information theory (see for instance Gersho and Gray [1992], Graf and Luschgy [2000]). There are other areas where the optimal quantization theory can be applied, such as cluster analysis (model-based clustering in statistics), pattern recognition (see Tarpey [1996], Tou and Gonzalez [1974], Diday and Simon [1976] for details).

Also in finance there are a survey of optimal quantization methods, related specifically with the multi-asset American option pricing and portafolio optimization (see Bally and Pagès [2003], Bally et al. [2003], Bally et al. [2005]) and non-linear filtering problems for stochastic volatility models (Pagès and Pham [2005], Bally et al. [2001]), for a review of this general topic we refer the reader to Pagès et al. [2004]. More recently the vector quantization theory has been applied in numerical probability for numerical integration in higher dimension (we refer the reader to Pagès [1998]).

2.1.1 The quantization problem

One concept which is crucial in optimal quantization is that of the Voronoï partition of $\mathbb{R}^d$. From now on, we make the assumption that α is a finite set.

***Definition* 2.1**

1. *The Voronoï region generated by $a \in \alpha$, where α is a finite set is defined by*

$$W(a|\alpha) \stackrel{\triangle}{=} \{x \in \mathbb{R}^d : ||x-a|| = \min_{b\in\alpha} ||x-b||\}.$$

2. *The set $\{W(a|\alpha) : a \in \alpha\}$ is called the Voronoï diagram of α, and is a local covering of $\mathbb{R}^d$.*
3. *A Borel measurable partition $\{C_a : a \in \alpha\}$ of $\mathbb{R}^d$ with respect to α is called Voronoï partition if*

$$C_a \subset W(a|\alpha).$$

4. *The open Voronoï region generated by $a \in \alpha$ is the set*

$$W_0(a|\alpha) \stackrel{\triangle}{=} \{x \in \mathbb{R}^d : ||x-a|| < \min_{b\in\alpha\setminus\{a\}} ||x-b||\},$$

and satisfy that $W_0(a|\alpha) \subset C_a$.

The vector quantization or quantization consists in approximating X by a random vector taking finitely many values in $\mathbb{R}^d$. The quantization problem for random vector can be seen as the problem of finding a discrete probability distribution which is a good approximation (in some sense) of the probability distribution of the random vector X, or we can see this like the search of a random vector $\widehat{X}^*$ approximating the random vector X taking finite number of values.

Let X be a square integrable $\mathbb{R}^d$-valued random vector with distribution $\mathbb{P}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For $n \in \mathbb{N}$, let $\mathcal{F}_n$ be the set of all Borel measurable maps $f : \mathbb{R}^d \to \mathbb{R}^d$ with $|f(\mathbb{R}^d)| \leq n$ (where $|A|$ denote the cardinal of the set A). The elements of $\mathcal{F}_n$ are called n-quantizers.

***Definition* 2.2**

1. *The n-th quantization error for X of order r is defined by*

$$e_{n,r}(X) = \inf_{f\in\mathcal{F}_n} (\mathbb{E}||X - f(X)||^r)^{1/r}.$$

2. A quantizer $f^\star \in \mathcal{F}_n$ is called n-th optimal for X of order r if

$$e_{n,r}(X) = (\mathbb{E}||X - f^\star(X)||^r)^{1/r}.$$

The n-th quantization error for X of order r is usually called "distortion". The previous definition has the following characterization

***Proposition* 2.1**

$$\begin{aligned} e_{n,r}(X) &= \inf_{\substack{\alpha\subset\mathbb{R}^d \\ |\alpha|\leq n}} (\mathbb{E}\min_{a\in\alpha}||X-a||^r)^{1/r} \\ &= \inf\left\{(\mathbb{E}\min_{a\in\alpha}||X-a||^r)^{1/r} : \alpha\subset\mathbb{R}^d,\ |\alpha|\leq n\right\}. \end{aligned}$$

The previous proposition can be used as a definition (see for instance Graf and Luschgy [2000]). The set α is also called n-quantizer and in engineering literature *n-codebook*.

A set $\alpha \subset \mathbb{R}^d$ with $|\alpha| \leq n$ is called *n-optimal set of centers for X of order r* if

$$e_{n,r}(X) = (\mathbb{E}\min_{a\in\alpha}||X-a||^r)^{1/r}.$$

The set of all n-optimal set of centers for X of order r is denoted by $\mathfrak{C}_{n,r}(X)$ and is also written as $\mathfrak{C}_{n,r}(\mathbb{P})$.

For $n \in \mathbb{N}$ fixed searching for an n-optimal quantizer is equivalent to the n-centers problem. If f is an n-optimal quantizer, then $f(\mathbb{R}^d)$ is an n-optimal set of centers. Conversely, if $\alpha \in \mathbb{R}^d$ is an n-optimal set of centers and $\{C_a : a \in \alpha\}$ is a Voronoï partition of $\mathbb{R}^d$ with respect to α, then $\widehat{X}^\star$ is an n-optimal quantizer.

The n-optimal set of centers for X of order r corresponds to the global minima of the function

$$\psi^X_{n,r} : \left(\mathbb{R}^d\right)^n \to \mathbb{R}_+, \quad \psi^X_{n,r}(x_1,\cdots,x_n) = \mathbb{E}\min_{1\leq i\leq n}||X-x_i||^r.$$

In what follows we recall some well-known results about quantization. We define the diameter of a nonempty bounded subset A of $\mathbb{R}^d$ by

$$\operatorname{diam}(A) = \sup\{||a-b|| : a,\, b \in A\}.$$

The following proposition state two properties about the set of all n-optimal set of centers for X of order r.

***Proposition* 2.2**

i. (Euclidean norms) Let $||x|| = \langle x, x\rangle^{1/2}$ for some scalar product on $\mathbb{R}^d$, then the set of all centers of $\mathbb{P}$ of order r satisfies

$$\mathfrak{C}_{n,r}(\mathbb{P}) \subset \operatorname{cl}\operatorname{conv}(\operatorname{supp}\mathbb{P}),$$

where $\operatorname{cl}\operatorname{conv}(A)$ is the closure of convex hull of A and $\operatorname{supp}\mathbb{P}$ is the support of $\mathbb{P}$.

ii. If we suppose that $\operatorname{supp}\mathbb{P}$ is compact, then

$$\sup_{\alpha\in\mathfrak{C}_{n,r}(X)}\ \min_{x\in\operatorname{supp}\mathbb{P}}||x-a|| \leq \operatorname{diam}(\operatorname{supp}\mathbb{P}).$$

The proof can be found in Graf and Luschgy [2000]. From now on we asume the following moment condition for $1 \leq r \leq \infty$,

$$\mathbb{E}||X||^r < \infty.$$

Under that hypothesis the quantization error is finite.

For the n-quantizer we define the closest neighbour projection by

$$\pi_\alpha \stackrel{\triangle}{=} \sum_{x_i \in \alpha} x_i \mathbf{1}_{C_\alpha(x_i)},$$

where $\alpha = \{x_1, \cdots, x_n\}$. The best approximation $\widehat{X}^\alpha : \mathbb{R}^d \to \alpha \subset \mathbb{R}^d$ of X by the rule of closest neighbour is

$$\widehat{X}^\alpha = \pi_\alpha(X) = \sum_{i=1}^n x_i \mathbf{1}_{C_i(X)}(X),$$

where α is the n-optimal set of centers and $C_i(X) \stackrel{\triangle}{=} C_\alpha(x_i)$ is the Voronoï partition induced by each x_i (the argument X in $C_i(X)$ will often be dropped when no confusion can arise).

It is obvious and easily checked that for any random vector $X' : \mathbb{R}^d \to \alpha \subset \mathbb{R}^d$,

$$\mathbb{E}||X - X'||^r \geq \mathbb{E}||X - \widehat{X}^\alpha||^r = \mathbb{E} \min_{a \in \alpha} ||X - a||^r.$$

If $\alpha^\star$ is an optimal set of center we denote the optimal quantizer by $\widehat{X}^\star \stackrel{\triangle}{=} \widehat{X}^{\alpha^\star}$. The quantization error for the optimal n-quantizer can be written as follows

$$\begin{aligned} e^r_{n,r}(X) &= \mathbb{E}||X - \widehat{X}^\star||^r = \sum_{i=1}^n \mathbb{E}(1_{C_i}(X)||X - x_i||^r) \\ &= \sum_{i=1}^n \int_{C_i} ||\xi - x_i||^r \mathbb{P}(d\xi) \\ &= \int_{\mathbb{R}^d} \min_{1 \leq i \leq n} ||\xi - x_i||^r \mathbb{P}(d\xi). \end{aligned}$$

2.1.2 Properties of optimal quantizers

In this section we recall some of the main properties associated to optimal quantizers. The following theorems are proved in Graf and Luschgy [2000].

Theorem* 2.1 *Existence
For all $n \in \mathbb{N}$ the function $\psi^X_{n,r}$ is continuous in $(\mathbb{R}^d)^n$ and $e_{n+1,r}(X) < e_{n,r}(X)$. The level set $\{\psi^X_{n+1,r} \leq c\}$ is compact for every $0 < c \leq e^r_{n,r}(X)$. In particular $\mathfrak{C}_{n,r}(\mathbb{P})$ is not empty.

Theorem 2.2 Necessary conditions for optimality
Let $\alpha \in \mathfrak{C}_{n,r}(\mathbb{P})$ and let $\{C_a : a \in \alpha\}$ be a Voronoï partition of $\mathbb{R}^d$. Then

$$|\alpha| = n, \quad \mathbb{P}(C_a) > 0 \text{ for every } a \in \alpha,$$

$$\beta \in \mathfrak{C}_{m,r}\Big(\mathbb{P}\big(\,\cdot\,\big|\bigcup_{a\in\beta} C_a\big)\Big) \text{ for every } \beta \subset \alpha \text{ with } |\beta| = m.$$

In particular

$$\mathbb{P}(W(a|\alpha)) > 0,\ a \in \mathfrak{C}_{1,r}\big(\mathbb{P}(\cdot|W(a|\alpha))\big) \text{ for every } a \in \alpha.$$

Theorem 2.3 Necessary conditions for optimality
Let $\alpha \in \mathfrak{C}_{n,r}(\mathbb{P})$ in $\mathbb{R}^d$ and $r > 1$ or $\mathbb{P}(\alpha) = 0$. Suppose the underlying norm is strictly convex and smooth. Then the Voronoï diagram $\{W(a|\alpha) : a \in \alpha\}$ is a $\mathbb{P}$-tessellation of $\mathbb{R}^d$, i.e.,

$$\mathbb{P}\Big(\mathbb{R}^d \Big\backslash \bigcup_{a\in\alpha} W_0(a|\alpha)\Big) = 0.$$

Finding an optimal quantizer could be a hard task for some distributions. We can define another kind of quantizers with suitable properties: the stationary quantizers. These quantizers are a useful tool in the quantization theory. Among many applications it has been used in numerical integration, see for instance Pagés and Printems [2003].

Definition 2.3 Stationary quantizers
A n-quantizer $\widehat{X}^\alpha$ is called stationary if any of the next three conditions are satisfied:

1. *If the set $\alpha \subset \mathbb{R}^d$ with $|\alpha| = n$ satisfies*

$$\mathbb{P}(W(a|\alpha)) > 0,\ a \in \mathfrak{C}_{1,r}(\mathbb{P}(\cdot|W(a|\alpha))) \text{ for every } a \in \alpha.$$

2. *If $\widehat{X}^\alpha$ is a nearest neighbour projection that satisfies*

$$\widehat{X}^\alpha = \mathbb{E}(X|\widehat{X}^\alpha). \tag{2.1}$$

3. *If $\widehat{X}^\alpha$ is the critical point of the distortion, i.e.:*

$$\nabla e^r_{n,r}(\widehat{X}^\alpha) = 0.$$

A set satisfying any of these conditions is called n-stationary set of center for X of order r. Let us denote by $\mathcal{S}_{n,r}(X)$ the set of all these n-stationary sets for X.

Obviously a locally optimal n-quantizer is always n-stationary when $r \geq 1$.

$$\mathfrak{C}_{n,r}(X) \subset \mathcal{S}_{n,r}(X).$$

The next theorem give us the relation between the stationary product quantizers and the quantization error for optimal quantizers.

Theorem 2.4 Product quantizers
Let $n_i \in \mathbb{N}$, $\beta_i \subset \mathbb{R}$ with $|\beta_i| \leq n_i$, $1 \leq i \leq d$ and $\alpha = \times_{i=1}^{d} \beta_i$ hence

1. *Suppose that $X_1, \cdots, X_d$ are independent and let $n = \prod_{i=1}^{d} n_i$. Then $\alpha \in \mathcal{S}_{n,r}(X)$ if and only if $\beta_i \in \mathcal{S}_{n_i,r}(X_i)$ for every i.*
2. *If $\beta_i \in \mathfrak{C}_{n_i,r}(X_i)$ for every i, then*

$$\mathbb{E} \min_{a \in \alpha} ||X - a||^r = \sum_{i=1}^{d} e^r_{n_i,r}(X_i).$$

The next result only holds in dimension one ($d = 1$) and it concerns the uniqueness of the n-stationary sets and consequently gives the uniqueness of the n-optimal sets of centers.

Theorem 2.5 Uniqueness
If $\mathbb{P}$ is strongly unimodal, then $|\mathcal{S}_{n,r}(X)| = 1$, for every $n \in \mathbb{N}$, $1 \leq r < +\infty$.

This theorem can be found in Graf and Luschgy [2000] and is due to Kieffer [1983]. Typically the uniqueness of optimal quantizers in higher dimension never occurs.

In one dimension, uniqueness of the optimal n-quantizer was extensively studied in Kieffer [1983] and Trushkin [1982] and leads to the following criterion: if the distribution of X is absolutely continuous with a log-concave density function, then, for every $n \geq 1$ there exists only one stationary quantizer of size n, which turns out to be the optimal quantizer at level n. It is posible to compute the optimal quantizer at level n in one dimension as the solution of the stationarity equation (2.1) either by a zero search method (Newton-Raphson gradient descent) or a fixed point procedure (like the Lloyd I procedure). In higher dimensions, deterministic gradient descent methods become intractable and one uses Lloyd I procedure and/or stochastic procedures to compute optimal quantizers.

2.1.3 Asymptotic results

One of the most important property of optimal quantizers is the asymptotic behaviour of the quantization error. It is easy to check that $e^r_{n,r}$ decreases to zero when the number of quantizers go to infinity. The proof is immediate. (see e.g. Graf and Luschgy [2000]). The asymptotic theorem comes from the earlier work of Zador [1982] and it was completely fulfilled in Graf and Luschgy [2000].

Theorem 2.6 Rate of decay
Let $r > 0$, assume that $\int_{\mathbb{R}^d} |\xi|^{\xi+\eta} \mathbb{P}(d\xi) < +\infty$ for some $\eta > 0$. Set $f \stackrel{\triangle}{=} d\mathbb{P}/d\lambda_d$, then

$$\lim_n \left(n^{\frac{r}{d}} \min_{(\mathbb{R}^d)^n} e^r_{n,r}(X) \right) = J_{r,d} ||f||_{\frac{d}{d+r}} < +\infty,$$

where

$$||g||_p \stackrel{\triangle}{=} \left(\int_{\mathbb{R}^d} |g(\xi)|^p \mathbb{P}(d\xi) \right)^{1/p} \quad \text{for every } p \in (0, +\infty).$$

The positive real constant $J_{r,d}$ corresponds to the uniform distribution on $[0,1]^d$. One knows that $J_{r,1} = 1/(2^r(r+1))$, $J_{2,2} = 5/(18\sqrt{3})$. When $d \geq 3$, $J_{r,d}$ is unknown. However the following asymptotic expansion holds $J_{r,d} = d/(2\pi e)^{r/2} + o(d)$ when $d \to +\infty$.

Asymptotically uniform distribution of the local distortion

Now, we focus on the asymptotic behaviour of the local distortion in one dimension. The statement is that local distortion of stationary or optimal quantizers is asymptotically uniform (see Delattre et al. [2004]). We can rewrite the quantization error as following

$$e^r_{n,r}(X) \;=\; \sum_{i=1}^{n} e^r_{n,i,r}(X),$$

where $e^r_{n,i,r} = \int\limits_{C_i} ||\xi - x_i||^r \mathbb{P}(d\xi)$ is the local distortion.

The conjeture is that for any distribution $\mathbb{P}$ having a positive probability distribution function f and a number of quantizers n

$$\begin{cases} \mathbb{P}(C_i(x^{(n)})) = \frac{1}{n} f^{\frac{r}{r+d}}(x_i^{(n)}) \left(\int_{\mathbb{R}^d} f^{\frac{r}{r+d}}(\xi) d\xi \right) + o(1/n). \\ \\ \int_{C_i(x^{(n)})} |x_i^{(n)} - \xi|^r \, \mathbb{P}(d\xi) = \frac{e^r_{n,r}(X)}{n} + o(1/n) \qquad \text{as } n \to +\infty. \end{cases}$$

The result in Delattre et al. [2004] is focused on proving this in one dimension.

***Definition* 2.4**
A sequence of $(x^{(n)})_{n\geq 1}$ of quantizers is asymptotically optimal if it achieves the rate of convergence in Theorem 2.6, i.e.:

$$\lim_n \left(n^{\frac{r}{d}} e^r_{n,r}(x^{(n)}) \right) = J_{r,d} ||f||_{\frac{d}{d+r}}.$$

The main result reads as follows

Theorem* 2.7 *Stationary asymptotically optimal quantizers
Let $\mathbb{P}$ be a distribution on the real line with a connected support. Assume that $\mathbb{P}$ has a positive and Lipschitz continuous p.d.f. on every compact set of the interior (m, M) of $\text{supp}(\mathbb{P})$. *Let $(x^{(n)})_{n\geq 1}$ be a sequence of stationary r-quantizers, $r > 0$.*

Let us define the stepwise function associated to the μ mass of the cells

$$\varphi_n(\xi) \stackrel{\triangle}{=} \sum_{i=1}^{n} \mathbf{1}_{C_i(x^{(n)})}(\xi) \mathbb{P}(C_i(x^{(n)})).$$

It holds

i. The sequence $(\varphi_n)_{n\geq 1}$ is relatively compact for the topology of the uniform convergence on compact sets of (m, M).

ii. If, furthermore, the sequence of $(x^{(n)})_{n\geq 1}$ of quantizers is asymptotically optimal, then the sequence $(\varphi_n)_{n\geq 1}$ converge uniformly on compacts sets of (m, M) toward $c_{f,1/(r+1)} f^{\frac{r}{r+1}}$, i.e., for every $[a, b] \subset (m, M)$,

$$\sup_{\{i: x_i^{(n)} \in [a,b]\}} \left| n\mathbb{P}(C_i(x^{(n)})) - c_{f,1/(r+1)} f^{\frac{r}{r+1}}(x_i^{(n)}) \right| \stackrel{n\to+\infty}{\longrightarrow} 0.$$

The local distortion is asymptotically uniformly distributed, i.e.,for every $[a, b] \subset (m, M)$

$$\sup_{\{i:x_i^{(n)}\in[a,b]\}} \left| n^{r+1} \int_{C_i(x^{(n)})} |x_i^{(n)} - \xi|^r \mathbb{P}(d\xi) - J_{r,1}||f||_{\frac{1}{1+r}} \right| \overset{n\to+\infty}{\longrightarrow} 0.$$

iii. Moreover, if $\mathbb{P}$ *has a compact support* $[m, M]$ *and* f *is Lipschitz continuous on the whole interval, then all the above convergences hold uniformly on* $[m, M]$.

This result is proved using the explicit form of the Voronoï cells in one dimension. The authors in their conclusions said that a general multidimensional method still needs to be found. One may guess that the result still holds for any dimension.

2.2 Overview on functional quantization

The optimal quantization has been rigorous extended for the infinite dimensional case. Several authors had been investigating about this subject from a theoretical point of view. The so-called "functional quantization" of stochastic processes is the most recent branch of optimal quantization, the aim is to quantize some processes viewed as random vectors taking values in their path spaces. But only for the Gaussian case exists serious extensions in this topic. (see Luschgy and Pagès [2004a], Dereich et al. [2003], Luschgy and Pagès [2002], Graf et al. [2003], Dereich [2005]). For Lévy and diffusion process we can mention the works of Luschgy and Pagès [2006] Pagès and Printems [2005], Wilbertz [2005], Dereich and Scheutzow [2006], Dereich and Lifshits [2005] and a first application of functional quantization to statistical clustering of functional data has been investigated by Tarpey and Kinateder [2003].

The first approach to functional quantization is due to Luschgy and Pagès [2002]. The main results viewed here are extensively discussed in their article.

Let us consider a separable Hilbert space $(H, \langle\cdot,\cdot\rangle_H)$ with its natural σ-algebra and satisfying $\mathbb{E}||X||^2 < +\infty$. In this framework, the usual choice for H is $\mathcal{L}^2([0,1], dt)$ with the usual norm $||X||^2_{\mathcal{L}^2} = \int_0^1 X^2(t)dt$. One considers a random vector X defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking its values in H.

In this space the stochastic process is viewed as $\mathcal{L}^2([0,1], dt)$-random vector. In this context other spaces can be chosen with other features as $\mathcal{L}^p([0,1], dt)$ and $\mathcal{C}([0,1], \mathbb{R})$ spaces.

In general, functional quantization consists in studying the best approximation of stochastic process X in a Hilbert space H by H-valued vectors taking at most n values. In this section we summarize the classical results on the topic and we focus on the quadratic case ($r = 2$). Most of the basic concepts are functional versions of those from vector quantization. The results presented in this section comes mainly from the work of Luschgy and Pagès [2002, 2004a].

2.2.1 Main properties

Set X a random vector taking its values in the Hilbert H with distribution $\mathbb{P}$ by

$$X : (\Omega, \mathcal{A}, \mathbb{P}) \longmapsto (H, \langle\cdot,\cdot\rangle).$$

For $n \in \mathbb{N}$, the $L^2(\mathbb{P})$-quantization problem for X the level n consists in minimizing

$$\left(\mathbb{E}\min_{a\in\alpha}||X-a||^r\right)^{1/r},$$

over all the sets $\alpha \subset H$ and $|\alpha| \leq n$, where the set α is also called *n-codebook* or *n*-quantizer.

As before we assume the following moment condition.

$$\mathbb{E}||X||^r < +\infty.$$

The quantization error $e_n^r(X)$ is defined by

$$\begin{aligned} e_n^r(X) &= \inf_{\substack{\alpha\subset H\\ |\alpha|\leq n}} \mathbb{E}\min_{a\in\alpha}||X-a||^r \\ &= \inf\left\{\mathbb{E}\min_{a\in\alpha}||X-a||^r : \alpha\subset H,\ |\alpha|\leq n\right\}. \end{aligned}$$

For the n-quantizer the closest neighbour projection is defined by

$$\pi_\alpha \stackrel{\triangle}{=} \sum_{x_i\in\alpha} x_i \mathbf{1}_{C_\alpha(x_i)},$$

with $x_i \in \alpha\ \forall i$. The best approximation $\widehat{X}^\alpha : H \to \alpha \subset H$ of X by the rule of closest neighbour is

$$\widehat{X}^\alpha = \pi_\alpha(X) = \sum_{i=1}^n x_i \mathbf{1}_{C_i}(X),$$

where $\alpha \subset H$ with $|\alpha| \leq n$ is the n-optimal set of centers and $C_i \stackrel{\triangle}{=} C_\alpha(x_i)$ is the Voronoï partition induced by each x_i (the argument X in $C_i(X)$ will often be dropped when no confusion can arise). In the infinite dimensional case $C_i(X)$ is a Voronoï partition as usual satisfying

$$C_i(X) \subset V_\alpha(x_i) \stackrel{\triangle}{=} \{x\in H : ||x-x_i|| = \min_{b\in\alpha}||x-b||\}.$$

Hence the quantization error can be written by

$$\begin{aligned} e_n^r(X) &= \inf\left\{\mathbb{E}||X-\widehat{X}^\alpha||^r : \widehat{X}^\alpha = f(X), f: H\to H, \text{ Borel measurable}, \quad |f(H)|\leq n\right\} \\ &= \inf\left\{\mathbb{E}||X-\widehat{X}^\alpha||^r : \widehat{X}^\alpha : \Omega\to H, \text{ random vector}, |\widehat{X}^\alpha(\Omega)|\leq n\right\}. \end{aligned}$$

In the infinite dimensional framework many of the properties on the finite case holds. The existence of optimal n-quantizer and the characteristics of the distortion can be obtained as an extension of Theorem 2.1 in the finite dimensional case. The first results for the existence of optimal quantizers are due to Cuesta and Matran [1988]. The theorem that states the existence of optimal quantizers can be found in Section 2 of Luschgy and Pagès [2002]. The proof is based on weakly sequentially lower semi-continuity.

From now on we are placed on the quadratic case ($r = 2$). In Luschgy and Pagès [2002] some necessary conditions for optimal n-quantizers are provided.

***Proposition* 2.3**
Let denote $\mathfrak{C}_n(X)$ the set of all n-optimal set of centers and C_a^0 the open Voronoï Region generated by $a \in \alpha$. Hence if $\alpha \in \mathfrak{C}_n(X)$ and $|\operatorname{supp}(\mathbb{P})| \geq n$, then $|\alpha| = n$, $\min_{a\in\alpha} \mathbb{P}(C_a^0) > 0$ and

1. *For $\widehat{X}^\alpha = \sum_{i=1}^n x_i \mathbf{1}_{C_i(X)}(X)$ we have*

$$\widehat{X}^\alpha = \mathbb{E}(X|\widehat{X}^\alpha) \quad a.s.$$

2. *In particular, for every $x_i \in \alpha$,*

$$x_i = \mathbb{E}(X|X \in C_i). \tag{2.2}$$

3. *Furthermore, for every $x_i, x_j \in \alpha$, $x_i \neq x_j$*

$$\mathbb{P}(C_i \cap C_j) = 0. \tag{2.3}$$

Now, a set $\alpha \subset H$ satisfying $|\alpha| = n$, $\min_{a\in\alpha} \mathbb{P}(C_a^0) > 0$, (2.2) and (2.3) is called n-stationary set for X and the following properties are straightforward

***Lemma* 2.1**
If $\alpha \subset H$ is a n-stationary set for X, then

1. *$\alpha \subset \operatorname{cl}\operatorname{conv}(\operatorname{supp}\mathbb{P})$.*
2. *If $\mathbb{E}(X) = 0$, then α lies in the reproducing kernel Hilbert space or Cameron-Martin space K_X.*

The infinite dimensional quantization problem can be reduced to a problem on finite linear subspaces of H. In fact,

***Theorem* 2.8**
Let U be a finite dimensional linear subspace of H, let Π_U the orthogonal projection from H onto U and $\alpha \subset U$, the following statements are equivalents:

1. *$\alpha \in \mathfrak{C}_n(X)$.*
2. *$\alpha \in \mathfrak{C}_n(\Pi_U(X))$ and $e_n^2(X) = \mathbb{E}||X - \Pi_U(X)||^2 + e_n^2(\Pi_U(X))$.*

Let $\mathbf{u} = \{u_1, \cdots, u_m\}$ be an orthonormal subset of H, $U = \operatorname{span}\{u_1, \cdots, u_m\}$ (where $\operatorname{span}(\mathbf{u})$ denotes the linear subspace of H spanned by $\mathbf{u}$) and $Z = (\langle u_1, X\rangle, \cdots, \langle u_m, X\rangle)$. Let $T : U \to \mathbb{R}^m$ be the bijective linear isometry given by $T(u_j) = e_j$ for all $j = 1, \cdots, m$ for the standard basis $\{e_1, \cdots, e_m\}$ of $\mathbb{R}^m$, then

$$e_n(\Pi_U(X)) = e_n(Z), \tag{2.4}$$

where $e_n(Z)$ denotes the n-quantization error of Z with respect to the l^2 norm.

Using a basis of a linear subspace of H an upper bound for the quantization error can be obtained.

***Proposition* 2.4**
Assume that $\mathbb{E}(X) = 0$ *for simplicity. Let* $\{u_j : j \geq 1\}$ *be an orthonormal subset of* H *such that* $\operatorname{supp}\mathbb{P} \subset \operatorname{cl}\operatorname{span}\{u_j : j \geq 1\}$. *Then for every* n *and every* $m \in \mathbb{N}$,

$$e_n^2(X) \leq \sum_{j\geq m} \operatorname{Var}\langle u_j, X\rangle + \inf\left\{\sum_{j=1}^{m} e_{n_j}^2(\langle u_j, X\rangle) : n_1, \cdots, n_m \in \mathbb{N}, \prod_{j=1}^{m} n_j \leq n\right\}.$$

We have the following results when the H-valued random vector X has a Gaussian distribution.

***Theorem* 2.9**
Let $\alpha \subset H$ *be a* n*-stationary quantizer for* X, Γ_X *its covariance operator and* $U = \operatorname{span}(\alpha)$ *then* $\Pi_U(X)$ *and* $X - \Pi_U(X)$ *are independent so that* $\Gamma_X(U) = U$ *and* $\alpha \subset \Gamma_X(H) \subset K_X$.

If we ordered the eigenvalues of the covariance operator Γ_X and we take into account that $\sum_{k\geq 1}\lambda_k = \mathbb{E}||X||^2$ then the following theorem holds

***Theorem* 2.10**
Let $\alpha \in \mathfrak{C}_n(X)$, $U = \operatorname{span}(\alpha)$ *and* $m = \dim(U)$. *Then* $\Gamma_X(U) = U$ *and*

$$\mathbb{E}||X - \Pi_U(X)||^2 = \sum_{k\geq m+1}\lambda_k.$$

This theorem ensures that the m-dimensional subspaces of H spanned by m-optimal sets are spanned by eigenvectors of Γ_X which belongs to the m largest eigenvalues. Also note that

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

Theorem 2.9 and Theorem 2.10 show the important role of the Karhunen-Loève basis associated to the decomposition of the covariance operator. This operator achieves the greatest variance of the process with the first m components of the basis (we recall here that $\lambda_1 \geq \lambda_2 \geq \cdots$). Thus the infinite dimensional space problem could be reduced to problem in a m-dimensional subspace of H with a convenient isometry, the result concerning the mapping is

***Theorem* 2.11**
Let $\alpha \subset T_1$ *an arbitrary quantizer for* X *with associated Voronoï quantization* $\widehat{X}$ *with* $\{C_\alpha(x), x \in \alpha\}$ *a Voronoï Partition induced by* α. *If*

$$I : (T_1, \langle\cdot,\cdot\rangle_{T_1}) \longrightarrow (T_2, \langle\cdot,\cdot\rangle_{T_2}),$$

is a bijective isometry from $T_1 \to T_2$, *where* T_2 *is another separable Hilbert space, then:*

1. $I(C_\alpha(a)) = C_{I(\alpha)}(I(a))$ *for every* $a \in \alpha$.

2. $I(\widehat{X}) = \widehat{I(X)}$ *is a Voronoï quantization of* $I(X)$ *induced by* $I(\alpha)$.
3. $\mathbb{E}\min_{a\in\alpha}||X-a||_{T_1}^2 = \mathbb{E}\min_{I(a)\in I(\alpha)}||I(X)-I(a)||_{T_2}^2$.

This result comes from Luschgy et al. [2010] and the proof is easy to check. Usually $T_2 = l^2$ and

$$\begin{aligned} I : T_1 &\to l^2 \\ x &\rightsquigarrow (\langle x, e_j\rangle)_{j\geq 1}. \end{aligned}$$

Therefore we focus our attention on the quantization problem associated to a Gaussian random vector

$$\eta \stackrel{\triangle}{=} I(X),$$

on l^2 with distribution

$$\eta = (\eta_j)_{j\geq 1} \sim \bigotimes_{j=1}^{\infty} \mathcal{N}(0, \lambda_j).$$

Using Theorem 2.11 it is possible to find some stationary quantizers for stochastic process using the appropriate isometry $I : H \to l^2$.

Since the infinite dimensional quantization problem is not solvable by a finite computer algorithm it is necessary to reduce the dimension of the problem. If α is an optimal n-quantizer for $\bigotimes_{j=1}^{\infty}\mathcal{N}(0,\lambda_j)$, then $U = \operatorname{span}(\alpha)$ is a subspace of l^2. By Theorem 2.9 and (2.4) it holds

$$e_n^2\left(\bigotimes_{k=1}^{\infty}\mathcal{N}(0,\lambda_k)\right) = e_n^2\left(\bigotimes_{k=1}^{d_n}\mathcal{N}(0,\lambda_k)\right) + \sum_{k\geq d_n+1}\lambda_k,$$

where d_n is the dimension of U. In fact

$$\begin{aligned} e_n^2(X) &= \sum_{j\geq m+1}\lambda_j + e_n^2\left(\bigotimes_{j=1}^{m}\mathcal{N}(0,\lambda_j)\right) \quad \forall\, m \geq d_n \\ e_n^2(X) &< \sum_{j\geq m+1}\lambda_j + e_n^2\left(\bigotimes_{j=1}^{m}\mathcal{N}(0,\lambda_j)\right) \quad \forall\, m < d_n. \end{aligned}$$

The integral number (d_n) is defined as $d_n = \min\{\dim\operatorname{span}(\alpha) : \alpha \in \mathfrak{C}_n\}$ and it satisfies that

$$\frac{1}{b^{1/(b-1)}}\frac{2\ln n}{b} \lesssim d_n \leq n-1.$$

The upper bound comes from the fact that

$$\begin{aligned} d_n \leq \overline{d}_n &\stackrel{\triangle}{=} \max\{\dim\operatorname{span}(\alpha) : \alpha \in \mathfrak{C}_n\} \\ &\leq \max\{\dim\operatorname{span}(\alpha) : \alpha \text{ } n\text{-stationary}\} \\ &\leq n-1. \end{aligned}$$

The lower bound was obtained in Luschgy et al. [2010], where b is the same as in Theorem 2.13 assuming the same conditions for the Gaussian process X. For a deeper discussion see for instance Luschgy and Pagès [2002], Luschgy et al. [2010]. Nowadays the vanishing rate of d_n is unknown.

2.2.2 Quantizer schemes for Gaussian process

The problem of quadratic optimal functional quantization of centered Gaussian process is reduced to a finite dimensional optimal quantization problem for a Gaussian distribution with a diagonal covariance structure (see Luschgy et al. [2010]). In that article the authors describe four schemes based on the KL expansion of the Gaussian process. Here we recall briefly their definitions.

Quantizer design I(Optimal quantization)

The previous results for the integral number d_n lead us to decompose the search of a set of optimal quantizers in two steps (if possible). First theoretically or numerically to evaluate the eigenvalues and eigenvectors in the Karhunen-Loève expansion of the covariance operator. Second, to find d such that $\sum_{j\geq d+1}\lambda_j$ is small with respect to $e_n^2\left(\bigotimes_{j=1}^{\infty}\mathcal{N}(0,\lambda_j)\right)$ and then compute optimal n-quantizer $\alpha^d\subset\mathbb{R}^d$ with $|\alpha^d|\leq n$ of $\bigotimes_{j=1}^{d}\mathcal{N}(0,\lambda_j)$. This optimization problem can be reads as

$$(\mathcal{O}_n) \quad \triangleq \quad \left\{ e_n^2\left(\bigotimes_{j=1}^{d_n}\mathcal{N}(0,\lambda_j)\right)+\sum_{j\geq d_n+1}\lambda_j\ .\right.$$

Moreover, if $\beta=(\beta_1,\cdots,\beta_n)$ denotes the optimal n-quantizer of $\bigotimes_{j=1}^{d_n}\mathcal{N}(0,\lambda_j)$, then the optimal n-quantizer $\alpha=(x_1,\cdots,x_n)$ for the Gaussian process X can be written as

$$x_i(t)=\sum_{j=1}^{d_n}(\beta_i)^j\varphi_j(t), \qquad i=1,\cdots,n,$$

where $\{\varphi_j\}_{j\geq1}$ is the KL basis for X and $(\lambda_j)_{j\geq1}$ the orderer eigenvalues associated to $\{\varphi_j\}_{j\geq1}$.

Obviously $(\mathcal{O}_n)$ is a finite dimensional quantization optimization problem. It depends strongly on the behaviour of the eigenvalues of the process X. Due to its fast decreasing for almost all the classical Gaussian process this optimization problem is ill conditioned. This issue can be avoided if $(\mathcal{O}_n)$ is rewritten taking the optimization problem with respect to the standard normal distribution $N(0,I_{d_n})$ by simply considering the Euclidean norm derived from a specific covariance matrix

$$(\mathcal{O}_n) \quad \Leftrightarrow \quad \left\{ \begin{array}{l} n\text{-optimal quantization of } \bigotimes_{j=1}^{d_n}\mathcal{N}(0,1) \\ \\ \text{for the covariance norm } |z_1,\cdots,z_{d_n}|^2=\sum_{j=1}^{d_n}\lambda_j z_j^2. \end{array}\right.$$

This approach is related with the work on Pagès and Printems [2005]. The next three designs are based on product quantizers and achieves more numerical accuracy and stationary. These designs have also the advantage that are less storage demanding.

Quantizer design II

Optimal $\bigotimes_{j=k_i+1}^{k_{i+1}} \mathcal{N}(0,\lambda_j)$-quantizer $\alpha^{(i)} \subset \mathbb{R}^{l_i}$ with $|\alpha^{(i)}| \leq n_i$ for some integers $m \in \mathbb{N}$, $l_1, \cdots, l_m \leq l$, $n_1, \cdots, n_m > 1$ and $\prod_{i=1}^m n_i \leq n$.

Quantizer design III

Optimal $\bigotimes_{j=k_i+1}^{k_{i+1}} \mathcal{N}(0,1)$-quantizer $\alpha^{(i)} \subset \mathbb{R}^{l_i}$ with $|\alpha^{(i)}| \leq n_i$ for some integers $m \in \mathbb{N}$, $l_1, \cdots, l_m \leq l$, $n_1, \cdots, n_m > 1$ and $\prod_{i=1}^m n_i \leq n$.

Quantizer design IV

Optimal $\mathcal{N}(0,1)$-quantizer $\alpha_i \subset \mathbb{R}$ with $|\alpha_i| \leq n_i$ for some integers $n_1, \cdots, n_m > 1$, $m \in \mathbb{N}$ and $\prod_{i=1}^m n_i \leq n$.

A deeper discussion on this topic can be see in Luschgy et al. [2010]. They provide some considerations about the behaviour of the schemes for the Brownian motion.

We propose a different scheme based on the discretization of the Gaussian process. This approach has the advantage that the eigenvalues of the Karhunen-Loève basis are not needed. The drawback is that a considerable amount of calculations is demanded.

2.2.3 Rate of convergence

The convergence of the functional quantization rate of e_n^2 has been studied by several authors. The best general reference here is due to Luschgy and Pagès [2002]. For more details see also Luschgy and Pagès [2004a] and Dereich [2003] and references therein. In the infinite dimensional Gaussian setting exits a general version of Theorem 2.6.

Theorem* 2.12 *Rate of Decay
If X is a centered H-valued random vector with a Gaussian distribution $\mathbb{P}$, covariance operator Γ_C and $\{u_j^X\}_{j\geq 1}$ is any orthonormal subset of H, such as $K_X \subset \operatorname{cl}\operatorname{span}\{u_j^X, j \in \mathbb{N}\}$, then let us define

$$\mu_j = \operatorname{Var}\langle u_j^X, X\rangle = \langle u_j^X, \Gamma_C(u_j^X)\rangle \quad \textit{and} \quad \Sigma_m = (\langle u_j^X, \Gamma_C(u_k^X)\rangle)_{0\leq j,k\leq m},$$

and for $n \in \mathbb{N}$ set

$$g_n(m) = e_n(\mathcal{N}(0,\Sigma_m)).$$

If we assume that $\{u_j^X :\subset \operatorname{cl}(K_X)\}$. Then $\det \Sigma_m > 0$ and

$$\lim_{n\to\infty} n^{1/m} g_n(m) = Q(m) \quad \textit{for every} \quad m \geq 1,$$

where $Q(m) \in (0, +\infty)$ and

$$Q(m) \sim \left(m(\det \Sigma_m)^{1/m}\right)^{1/2} \quad \textit{as} \quad m \to \infty.$$

In particular

$$\lim_{m\to\infty} Q(m) = 0.$$

The previous theorem is quite general. More specific asymptotics for the quantization error exists for a large variety of Gaussian process. The methods used to establish the rate of convergence relies in the behaviour of the eigenvalues of the covariance operator, regularly varying functions, small ball probabilities and Shannon-Kolmogorov ϵ-entropy.

The first approach to the asymptotics of quantization error is due to Luschgy and Pagès [2002]. They give us upper and lower bounds for e_n using the eigenvalues behaviour (regularly varying function) of the covariance operator and Shannon-Kolmogorov's ϵ-entropy respectively (see Theorem 4.6 and 4.12 in Luschgy and Pagès [2002]).

In Graf et al. [2003] and Dereich [2003] a different approach is proposed using regularly varying functions, small ball probabilities and their link with the quantization error, see for instance Theorem 1.2 in Graf et al. [2003] and Theorem 2.3 in Dereich [2003]. Sharp bounds are given for a wide class of Gaussian processes (see also Dereich et al. [2003]).

The most importants results for sharp asymptotics of quantization error are detailed in Luschgy and Pagès [2004a,b]. The general idea is mainly based on regularly varying functions and Shannon-Kolmogorov's ϵ-entropy. We recall here the Theorem 2.2 (a) in Luschgy and Pagès [2004a] which provides the sharp rate of convergence for the distortion.

***Theorem* 2.13**
If X is a Gaussian process with eigenvalues $\lambda_j \sim \varphi(j)$ as $j \to \infty$, where $\varphi : (s, \infty) \to (0, \infty)$ is a decreasing, regularly varying function at infinity of index $-b < -1$ for some $s \geq 0$. Set, for every $x > s$,

$$\psi(x) \stackrel{\triangle}{=} \frac{1}{x\varphi(x)},$$

then

$$e_n(X) \sim \left(\left(\frac{b}{2} \right)^{b-1} \frac{b}{b-1} \right)^{1/2} \psi(\ln n)^{-1/2}. \tag{2.5}$$

The most prevalent form for φ is

$$\varphi(x) = cx^{-b}(\ln x)^{-a}, \quad b > 1,\ a \in \mathbb{R},\ x > \max\{1, e^{-a/b}\},$$

and (2.5) turns to

$$e_n(X) \sim \left(\left(\frac{b}{2} \right)^{b-1} \frac{b}{b-1} \right)^{1/2} (\ln n)^{-(b-1)/2} (\ln \ln n)^{-a/2}.$$

For the particular case of diffusions the Shannon-Kolmogorov's ϵ-entropy plays a fundamental role, see for instance Luschgy and Pagès [2006], Dereich and Scheutzow [2006] and Dereich [2008]. Some asymptotics are also provided for the d-dimensional Brownian motion.

The statement of the previous theorem establishes the important role of the eigenvalues of the covariance operator. The theorems which states the asymptotic of eigenvalues comes from the works of Widom and Rosenblatt (see Theorem 1 in Widom [1964] and Theorem 3 in Rosenblatt [1963] for a deeper discussion).

In the next theorem we recall the sharp asymptotics for a wide class of Gaussian process. The asymptotic behaviour of e_n comes as a consequence of Theorem 2.13.

***Theorem* 2.14**
Let X be a Gaussian process and $e_n^2(X)$ be the quantization error associated to the process, then

1. *If X is a Wiener process, then*

$$e_n(X) \sim \frac{\sqrt{2}}{\pi}(\ln n)^{-1/2}. \tag{2.6}$$

2. *If X is a Brownian bridge in $[0,1]$ process, then*

$$e_n(X) \sim \frac{\sqrt{2}}{\pi}(\ln n)^{-1/2}.$$

3. *If X is an m-integrated Brownian motion, then*

$$e_n(X) \sim \pi^{-(m+1)}(m+1)^{m+1/2}\left(\frac{2m+2}{2m+1}\right)^{1/2}(\ln n)^{-(m+1/2)}.$$

4. *If X is a Gaussian diffusion, (i.e. is the unique solution of the equation $dX_t = A(t)X_t dt + dW_t$ with W_t a Wiener process, $X_0 = \xi \sim N(0,\sigma^2)$ and $A \in \mathcal{L}^2([0,1], dt)$.) then*

$$e_n(X) \sim \frac{\sqrt{2}}{\pi}(\ln n)^{-1/2}.$$

5. *If X is a fractional Brownian motion with Hurst exponent $H \in (0,1)$, then*

$$e_n(X) \sim \left(\frac{\Gamma(2H)\sin(\pi H)(1+2H)}{\pi}\right)^{1/2}\left(\frac{1+2H}{2\pi}\right)^{H}(\ln n)^{-H}.$$

6. *If X is a fractional Ornstein-Uhlenbeck process of third kind with Hurst exponent $H \in (0,1)$, then*

$$e_n(X) \sim \left(\frac{\theta\Gamma(2H)\sin(\pi H)(1+2H)}{\pi}\right)^{1/2}\left(\frac{1+2H}{2\pi}\right)^{H}(\ln n)^{-H}.$$

7. *If X is the standard stationary Ornstein-Uhlenbeck process, then*

$$e_n(X) \sim \frac{2\sqrt{\theta}}{\pi}(\ln n)^{-1/2}.$$

8. *If X is a Gaussian sheet (i.e. is a centered Gaussian fields $X = (X_t)_{t\in[0,1]^d}$), then*

$$e_n(X) \sim \frac{\sqrt{2}}{\pi^d(d-1)!}(\ln n)^{-1/2}(\ln\ln n)^{d-1}.$$

See for instance Luschgy and Pagès [2004a] for more details and a deeper discussion. Theorem 2.14 shows us that the asymptotic behaviour of the quantization error for a wide class of stochastic process has been well studied. However for the fOU(1) and fOU(2) there are not such asymptotic closed forms. In the next lemma we indicate the asymptotic behaviour of e_n for both processes as a consequence of Theorem 2.13.

***Lemma* 2.2**

1. *If X is a fractional Ornstein-Uhlenbeck process of first kind with Hurst exponent $H \in (0,1)$, then*

$$e_n(X) \sim \left(1+\frac{1}{H}\right)^{1/2} \left(\frac{1+2H}{2\pi}\right)^{H} (\ln n)^{-H}.$$

2. *If X is a fractional Ornstein-Uhlenbeck process of second kind with Hurst exponent $H \in (0,1)$, then*

$$e_n(X) \sim \theta \left(1+\frac{1}{H}\right)^{1/2} \left(\frac{1+2H}{2\pi}\right)^{H} (\ln n)^{-H}.$$

Proof

The asymptotic behaviour of the eigenvalues is a straightforward consequence of Theorem 3 in Rosenblatt [1963] for the spectral density associated to each fOU (see Subsection 1.1.5 for details). Using Theorem 2.13 the proof is complete.

□

Chapter 3

Stochastic algorithms

The basic stochastic approximation algorithms was introduced in the early 1950s by the works of Robbins and Monro and also by the works of Kiefer and Wolfowitz. This subject has been studied extensively from that moment due to the variability and versatility of their applications.

There are many areas where this approach has been used satisfactorily. One can mention the problems of estimating unknown parameters based on observation data containing information (pattern classification, adaptative control, etc), or the problem of searching a maximum of certain function and the two armed bandit. Almost all the problems mentioned before can be resumed or transformed into a root-seeking problem for an unknown function. Recursive methods for root finding are common in classical numerical analysis, and it is reasonable to expect that appropriate stochastic analogs would also perform well, see Duflo [1996] for specific information about this topic.

The classical formulation for stochastic algorithms is given by

$$X_{k+1} = X_k + \gamma_k F(X_k, \omega_{k+1}),$$

where $(\omega_k)_{k\geq 1}$ is a sequence of real-valued, mutually independent, and identically distributed, random variables with finite variance, $(X_k)_{k\geq 1}$ is a sequence in certain Euclidean space, $(\gamma_k)_{k\geq 1}$ is a sequence that tends to zero when k goes to infinity, (in more specific problems, some others properties are required for this sequence) and F is a function that takes different form for each particular problem.

The general setting of stochastic algorithms is the following. Let us define $f \subset \mathcal{O} \subset \mathbb{R}^d \rightsquigarrow \mathbb{R}^d$ as a continuous function in the open set $\mathcal{O}$, assume that there exists a unique point $X^* \in \overline{\mathcal{O}}$ such as $f(X^*) = a$, where a is certain known value. If X^* satisfies that $\langle X - X^*, f(X) - a\rangle < 0$ for all $X \neq X^*$ then take

$$X_{k+1} = X_k + \gamma_k(f(X_k) - a),$$

where $\sum \gamma_k = +\infty$ and $\gamma_k \searrow 0$. When X^* satisfies that $\langle X - X^*, f(X) - a\rangle > 0$ for all $X \neq X^*$, then take

$$X_{k+1} = X_k - \gamma_k(f(X_k) - a),$$

with the same conditions for $(\gamma_k)_{k\geq 1}$. To avoid explosion ones assume that $|f(X)| \leq K(1 + \|X\|)$, for all $X \in \mathcal{O}$ and some positive constant K. Therefore X_k converge to X^*.

The origins of recursive stochastic algorithms comes from the works of Robbins and Monro. They developed the recursive procedure for finding the root of a real-valued function. The idea is to take an observation at the current estimator of the root and use that observation to make a small correction in the estimate, then takes an observation at the new value of the estimator, and so forth. We want to point out that usually the mentioned function is unknown and we can see a part of it or a noisy version

One of the first applications of stochastic algorithms comes from a dosage problem in chemistry, related with the idea of find the dose x^* that produce, at a given level α, certain mean effect $f(x) = \mathbb{E}[F(x, \epsilon)]$ were ϵ is some random variable and F an unknown function. The problem is reduced to solve the equation $f(x^*) = \alpha$. The Kiefer Wolfowitz approach is analogous to the method proposed by Robbins and Monro, they use a finite difference form to locate an extrema of an unknown function, for example in the case of the dosage, the maximum mean effect given by the dose, see Duflo [1997] for more details about this formulation.

The stochastic gradient method (i.e. a method of minima searching using the gradient) is based on the integral representation of the criterion to be optimized and can be seen as a particular case of the Robbins-Monro algorithm when the function has an unique point of minima, and both, the function and its definition space are convex.

We will resume in the next subsection the classical results on stochastic algorithms.

3.1 Theoretical background

This section contains a brief summary of classical theorems and results for stochastic algorithms. All the results in this section can be found in Duflo [1996, 1997]. The following theorem is quite important in the deterministic setting.

***Theorem* 3.1**
Suppose that f is a continuous real function such that $f(x^*) = \alpha$ and such that, for all $x \in \mathbb{R}$,

$$\begin{aligned} \langle f(x) - \alpha, x - x^* \rangle &< 0 \\ |f(x)| &\leq K(1 + |x|), \end{aligned}$$

for some positive constant K. Suppose that $(\gamma_k)_{k\geq 1}$ is a positive decreasing sequence going to zero, such that $\sum \gamma_k = +\infty$ and that $(\epsilon_k)_{k\geq 1}$ is another sequence such that $\sum \gamma_k \epsilon_{k+1} < +\infty$. Then the sequence $(X_k)_{k\geq 1}$ defined by

$$X_{k+1} = X_k + \gamma_k(f(X_k) - \alpha + \epsilon_{k+1}),$$

converge to x^* for all initial values X_0.

In particular if we have a function $f : \mathbb{R}^d \to \mathbb{R}$ and we assume that $f \in \mathcal{C}^2(\mathbb{R}^d)$ then the gradient of f exist and is denoted by ∇f. Assuming that the matrix of second derivatives $D^2 f$ of f is positive definite on the open convex $\Lambda \subset \mathbb{R}^d$ implies that the function f is convex. If f attains a minimum at x^* then

$$\nabla f(x^*) = 0,$$

and if we search a minimum x^* of f we can use a *gradient algorithm*:

$$X_{k+1} = X_k - \gamma_k \nabla f(X_k),$$

where $(\gamma_k)_{k\geq 1}$ is a positive sequence going to zero. The convergence of this algorithm comes as a straightforward application of the preceding theorem.

The Newton's method is a gradient method and is described as follows

$$X_{k+1} = X_k - \gamma_k \big(D^2 f(X_k)\big)^{-1} \nabla f(X_k).$$

The most important theorem in stochastic approximation is due to Robbins and Siegmund.

Theorem 3.2 (Robbins-Siegmund)
Suppose that $(X_k)_{k\geq 1}$, $(\beta_k)_{k\geq 1}$, $(\xi_k)_{k\geq 1}$ and $(\eta_k)_{k\geq 1}$ are four positive sequences adapted to some filtration $\mathbb{F}$ and that

$$\mathbb{E}[X_{k+1}|\mathcal{F}_k] \leq (1+\beta_k)X_k + \xi_k - \eta_k, \qquad a.s.$$

Then on $\Gamma = \{\sum \beta_k < +\infty \text{ and } \sum \xi_k < +\infty\}$, almost surely, (X_k) converges to a finite random variable X^* and the serie $\sum \eta_k < +\infty$.

For the proof we refer the reader to Duflo [1997].

The well-known theorem of Robbins Monro reads as follows

Theorem 3.3 (Robbins-Monro)
Suppose that $(X_k)_{k\geq 1}$ and $(Y_k)_{k\geq 1}$ are two square-integrable sequences of random vectors adapted to some filtration $\mathbb{F}$ with values in $\mathbb{R}^d$ and that $(\gamma_k)_{k\geq 1}$ is a positive sequence decreasing towards zero of random variables adapted to $\mathbb{F}$ with finite initial value γ_0, such that the relation $X_{k+1} = X_k + \gamma_k Y_{k+1}$ holds. Suppose further that

$$\mathbb{E}[Y_{k+1}|\mathcal{F}_k] = f(X_k) \text{ and } \mathbb{E}[\|Y_{k+1} - f(X_k)\|^2|\mathcal{F}_k] = \sigma^2(X_k).$$

If the function f is continuous from $\mathbb{R}^d$ to $\mathbb{R}^d$ and $f(x^*) = 0$ and for $x \neq x^*$, $\langle f(x), x - x^*\rangle < 0$ and if one of the following assertions is satisfied,

A1. $d = 1$ and $|f(x)| \leq K(1+|x|)$ for some positive constant K and almost surely

$$\sum \gamma_k = +\infty, \text{ and } \sum \gamma_k^2 \sigma^2(X_k) < +\infty.$$

A2. $\sigma^2(x) + \|f(x)\|^2 = s^2(x) \leq K(1+\|x\|^2)$ for some positive constant K and almost surely

$$\sum \gamma_k = +\infty, \text{ and } \sum \gamma_k^2 < +\infty,$$

then $X_k \overset{a.s}{\to} x^*$.

In particular we have the following

Proposition* 3.1 *(Robbins-Monro procedure)
If $f(x) = \mathbb{E}[F(x,\epsilon)]$ where ϵ is some random variable and F an unknown function and $f(x^*) = \alpha$ then the procedure

$$X_{k+1} = X_k - \gamma_k(F(X_k, \epsilon_{k+1}) - \alpha),$$

where $(\epsilon_k)_{k\geq 1}$ is a sequence of random independent variables with the same law of ϵ and $(X_k)_{k\geq 1}$ is a square-integrable sequence of random vectors adapted to some filtration $\mathbb{F}$ with values in $\mathbb{R}^d$ converge a.s. to the dosage x^* ($X_k \overset{a.s}{\to} x^*$) when X^* satisfies that $\langle X - X^*, f(X) - a\rangle < 0$ for all $X \neq X^*$ and if the two of the following assertions are satisfied

1. *$(\gamma_k)_{k\geq 1}$ is a decreasing positive deterministic sequence going to zero, such that*

$$\sum \gamma_k = +\infty, \text{ and } \sum \gamma_k^2 < +\infty.$$

2. *$s^2(x) = \mathbb{E}[F^2(x,\epsilon)] \leq K(1 + \|x\|^2)$ for some constant K.*

We must also mention the algorithm proposed by Kiefer-Wolfowitz to find the maximum of an unknown function. The basics are the same as for Robbins-Monro but now the idea is to maximize the mean effect for $d = 1$, i.e.

$$x^* = \arg\max f(x),$$

where the function $f(x)$ is assumed concave. The previous equation leads to find x^* $\Phi(x^*) = 0$ where

$$\lim_{c\to 0} \frac{1}{2c}(f(x+c) - f(x-c)) = \Phi(x).$$

The algorithm is described as follows.

$$X_{k+1} = X_k + \frac{\gamma_k}{c_k}\Big(F(X_k + c_k, \epsilon_{k+1}^1) - F(X_k - c_k, \epsilon_{k+1}^2)\Big),$$

where $(\epsilon_k^i)_{k\geq 1}$, $i = 1, 2$ are sequences of mutually independent random variables with the same law of ϵ and $(X_k)_{k\geq 1}$ is a square-integrable sequence of random variables adapted to some filtration $\mathbb{F}$ with values in $\mathbb{R}$. The sequence $(\gamma_k)_{k\geq 1}$ satisfies the same conditions as in the Robbins-Monro algorithm and $(c_k)_{k\geq 1}$ is a suitable sequence which decreases toward zero.

Theorem* 3.4 *(Kiefer-Wolfowitz)
The Kiefer-Wolfowitz algorithm for finding the maximum x^* which guarantees the maximum effect is that $(X_k)_{k\geq 1}$ converge almost surely to x^* if the following assumptions are satisfied,

1. *The two sequences $(\gamma_k)_{k\geq 1}$ and $(c_k)_{k\geq 1}$ are deterministic, decrease to zero and satisfies*

$$\sum \gamma_k = +\infty, \ \sum \gamma_k c_k < +\infty \text{ and } \sum \gamma_k^2 c_k^{-2} < +\infty.$$

2. *$s^2(x) = \mathbb{E}(F^2(x,\epsilon)) \leq K(1 + |x|^2)$.*

3. The function f is in the class $\mathcal{C}^2$, strictly concave and

$$|f''(x)| \leq K(1+|x|).$$

Many algorithms have been developed in this area. Most of all are related to stochastic approximation. The main idea behind stochastic approximation is simple. It is a slight modification of the optimization method called gradient descent. In higher dimensions, the stochastic approach seems to be the best choice for quantization purposes.

The stochastic gradient method is defined as follows

Definition* 3.1 *Stochastic gradient method
Let g be a twice differentiable function from E to $\mathbb{R}$ such that dg has an integral representation on E with respect to $\mathbb{P}$, i.e.:

$$dg(x) = \int_{\mathbb{R}^d} dG(x,\xi)\mathbb{P}(d\xi),$$

with $dG : E \times \mathbb{R}^d \to \mathbb{R}$, $dG(x,\cdot) \in \mathcal{L}^1(\mathbb{P})$ for every $x \in E$. The stochastic gradient method is defined by a triplet of sequences $((X_k)_{k\geq 0}, (\xi_k)_{k\geq 1}, (\gamma_k)_{k\geq 1})$ with values respectively in E, $\mathbb{R}^d$ and $\mathbb{R}_+$ satisfying for every $k \geq 1$ that

$$X_{k+1} = X_k - \gamma_{k+1} dG(X_k, \xi_{k+1}),$$

where $(\xi_k)_{k\geq 1}$ are i.i.d random variables with distribution $\mathbb{P}$ and $(\gamma_k)_{k\geq 1}$ is a decreasing positive deterministic sequence going to zero such that

$$\sum \gamma_k = +\infty.$$

In applications, one generally prefers recursive algorithms, owing to their relative computational simplicity. After each new observation, one need not recompute the estimator from all the data collected to date. Each successive estimate is obtained as a simple function of the last estimate and the current observation. Recursive estimators are widely used in applications in communications and control theory. Indeed, recursive stochastic algorithms had been used in the control and communications area for tracking purposes even before the work of Robbins and Monro.

We will focus our attention into one of this algorithms: the *Competitive Learning Vector Quantization* which comes from the stochastic gradient method.

3.2 CLVQ method

The topic of quantization has been well studied because it is a powerful technique for data compression. The philosophy of vector quantization (VQ) comes from Shannon's rate distortion theory.

Roughly speaking vector quantization is a method that replaces, in a certain way, a sample with a huge amount of data into a few representative vectors of the data. In order to reduce the size of the information, or compress the data with a suitable method the

vector quantization is a powerful tool. In Gersho [1978] an extensive discussion of vector-quantization techniques and applications is given. Others applications of this theory can be found in fields such as cluster analysis, pattern recognition, etc.

Other useful application of this theory is for coding the bit rate of images and speeches. The VQ appears to be a good solution, because among others aspects, is a statistical method who can encode data for transmission and also can preserve the topology, or the initial characteristics of the data. This matter is clear if we note that vector quantization can be seen as a particular case of the Kohonen algorithm, see for example Fowler et al. [1993], Fowler et al. [1995] and references therein.

One fundamental problem in this particular area is the excessive amount of calculations that requieres the method to achieve good numerical results. The computational complexity of traditional VQ codebook design method has restricted their use in real time applications. The performance of a vector-quantization system depends on the composition of the codebook, the criteria to find an optimal codebook or quantizer and the calculations performed in order to find the best solution, this last one are very expensive. Aside this drawback, the vector quantization is an excellent tool in many fields.

The CLVQ is an on-line algorithm that can be seen as a particular case of the "stochastic gradient method with decreasing step" (see Duflo [1996] for more details and a rigorously definition of this method). It also appears as a version of the Kohonen algorithm with 0 neighborhood.

All the results will be given in $\mathbb{R}^d$ for convenience, but they straightforwardly extend to a Hilbert space.

The probability distribution to quantize is denoted by $\mathbb{P}$. We will assume that $\mathbb{P}$ is diffuse. Let $\mathcal{O} = \{x \in (\mathbb{R}^d)^n, \forall i \neq j,\ x_i \neq x_j\}$. This assumption ensures that $\mathbb{P}(x \in \mathcal{O}) = 1$. Throughout this section we will assume that our algorithms lives in $\mathcal{O}$.

The classical formulation of the method is very simple. There are two phases: *competitive* and *learning*.

The first one is the most time consuming, because it uses at each step a closest neighbour search to find the nearest point to the random vector generated at each iteration of the method. The second phase only actualizes the value of the quantizer. The algorithm can be described as follows:

Steps of *Competitive Learning Vector Quantization*

1. Generate the initial n-quantizers $x_0^1, \cdots, x_0^n$.
2. Generate (independently) a random vector ξ with distribution $\mathbb{P}$.
3. Find the winning index (*competitive phase*)
$$i_0(k+1) = \arg\min_i \|x_k^i - \xi\|_2.$$
4. Update the quantizers (*learning phase*)
$$\begin{cases} x_{k+1}^i = x_k^i - \gamma_{k+1}(x_k^i - \xi), & i = i_0(k+1) \\ x_{k+1}^i = x_k^i, & i \neq i_0(k+1) \end{cases}.$$

5. Repeat steps 2-4 until to satisfy some convergence criteria.

There are many variations of the method, for example some of them can be found in Yong et al. [1997]. We can found others modifications of the CLVQ in Krishnamurthy et al. [1990] and Bouton and Pagès [1997] and references therein.

3.2.1 Convergence results

As we mentioned before the CLVQ algorithm could be seen as a stochastic gradient descent method. The theorem which ensures the convergence of such algorithms reads as follows

Theorem 3.5 Convergence a.s.
Let $g : E \to \mathbb{R}_+$ be a continuously differentiable function whose differential dg admits an integral representation on E with respect to $\mathbb{P}$

$$dg(x) = \int_{\mathbb{R}^d} dG(x, \xi)\mathbb{P}(d\xi).$$

Assume that dg and dG satisfy

$$\lim_{|x|\to+\infty} g(x) = +\infty \text{ and } dg \text{ is Lipschitz continuous,}$$

$$dg(x) = O(g(x)), \text{ when } |x| \to +\infty.$$

Let $((X_n)_{n\geq 0}, (\xi_n)_{n\geq 1}, (\gamma_k)_{k\geq 1})$ be a stochastic gradient method with a positive gain parameter sequence satisfying

$$\sum_{k\geq 1} \gamma_k = +\infty, \qquad \sum_{k\geq 1} \gamma_k^2 < +\infty.$$

Then $g(X_k)$ a.s. converges to some nonnegative random variable g_∞ and X_k a.s. converges toward some random connected component χ^* of $\{dg = 0\} \cap \{g = g_\infty\}$. In particular, if $\{dg = 0\} = \{x^*\}$, then

$$X_k \to x^* \text{ a.s. when } k \to +\infty.$$

The previous theorem applies completely in dimension one if we use the quadratic distorsion and assume that the distribution $\mathbb{P}$ has a compact support and is absolutely continuous with a log-concave probability density function. In general we can not ensure that the algorithm converges to a global minimum. In higher dimension, uniqueness of stationary quantizers clearly often fails (see Graf and Luschgy [2000]), so that the CLVQ reaches, a priori, only locally optimal quantizers. Even if using a procedure of Lloyd the numerical precision can be improved, there is no reason to suppose that the obtained quantizers are, in some sense, the optimal ones.

3.3 Lloyd's method

The Lloyd's method is an algorithm that always converges to a local minimum because at each iteration the value of the distortion decreases. This algorithm is mostly used to improve the result of other methods and may be viewed as a fixed point iteration.

The implementation of this algorithm is very simple and is based on the search of stationary quantizers. Any stationary quantizer verify that it is the barycenter of its own Voronoï tessellation. The algorithm can be described as follows

Steps of Lloyd's method

1. Generate the initial n-quantizer $x_0^1, \cdots, x_0^n$.
2. Assign $x^i = x_0^i$ for all i.
3. Calculate the Voronoï tessellation for each quantizer $(x^1, \cdots, x^n)$.
4. Calculate de mass centroid (c_i) of each Voronoï region:

$$c_i = \frac{\int\limits_{C_i} \zeta \mathbb{P}(d\zeta)}{\int\limits_{C_i} \mathbb{P}(d\zeta)},$$

   where C_i is the Voronoï tessellation for x^i.
5. Update the values of x^i, $x^i = c_i \; \forall \; i = 1, \cdots, n$.
6. Calculate the value of the distortion $e_n^2(X)$ by the formulae:

$$e_n^2(X) = \sum_{i=1}^{n} \int\limits_{C_i} \|x^i - \zeta\|^2 \mathbb{P}(d\zeta).$$

7. Repeat steps 3-6 until reach some convergence criteria.

Of course it is generally impossible to compute the mass centroïd with the actual probability distribution $\mathbb{P}$. So the algorithm is applied to a huge sample from $\mathbb{P}$, for instance one to ten millions independent samples of $\mathbb{P}$.

Although the method has some good properties, there are some steps that are very time consuming and the local minimum obtained strongly depends on the initialization vector $(x_0^1, \cdots, x_0^n)$.

3.4 Kushner-Clark theorem

The Kushner-Clark theorem has been used to describe the behaviour of stochastic algorithms. It is also called *method of ordinary differential equation* and was introduced in Kushner and Clark [1978]. The most important results on *a.s.* convergence of stochastic algorithms is due to the mentioned theorem. The main idea comes from the link that can be made between the behaviour of the solution of the ODE ($\dot{x} = -h(x)$) and the asymptotic behaviour of each sample path of the stochastic algorithm.

Let us considerer the general recursive algorithm:

$$\begin{cases} X^0 \in \mathbb{R}^d, \\ \\ X_{k+1} = X_k - \gamma_{k+1} f(X_k, \xi_{k+1}), \end{cases} \tag{3.1}$$

where $k \in \mathbb{N}$, $(\xi_k)_{k\geq 1}$, $(X_k)_{k\geq 0}$ are two sequences on $\mathbb{R}^d$, γ_k is the gain of the method (i.e. $\gamma_k \searrow 0$) and H some general function, defined as $f(X_k, \xi_{k+1}) = H(X_k, \xi_{k+1}) - \eta_{k+1}$ where η_k is a noisy term such as $\eta_k \to 0$ when $k \to +\infty$.

Now let us define h as $h(X_k) = \mathbb{E}(H(X_k, \xi_{k+1}))$. Most of the results of convergence relies on the continuity of h and local properties of its zeros. In the Kushner-Clark approach, the convergence of $(X_k)_{k\geq 1}$ is "conditional $a.s.$" and related to the average ODE, which describes the mean behaviour of the stochastic algorithm. For instance the limiting points x^* of the algorithm in (3.1) will lies in $\{h = 0\}$.

The Kushner-Clark theorem need some previous definition and mathematical background. Let us write the general equation (3.1) in other form:

$$\begin{cases} X^0 \in \mathbb{R}^d, \\ \\ X_{k+1} = X_k - \gamma_{k+1} H(X_k, \xi_{k+1}) + \gamma_{k+1}\eta_{k+1}. \end{cases}$$

Defining $\Delta M_{k+1} = h(X_k) - H(X_k, \xi_{k+1})$ the previous formulation take the following form:

$$\begin{cases} X^0 \in \mathbb{R}^d, \\ \\ X_{k+1} = X_k - \gamma_{k+1} h(X_k) + \gamma_{k+1}(\Delta M_{k+1} + \eta_{k+1}). \end{cases} \tag{3.2}$$

Let us recall the general formulation needed for the asymptotic behaviour of a stochastic algorithm (see for instance Fort and Pagès [1996]). Let us define the stepwise functions $(X^{(k)})_{k\geq 0}$ by

$$\forall\, u \in \mathbb{R}_+,\ X_u^{(0)} \stackrel{\triangle}{=} X_k, \quad \text{if } u \in [\gamma_1 + \cdots + \gamma_k, \gamma_1 + \cdots + \gamma_{k+1}).$$

Then

$$\forall k \geq 1, \forall\, u \in \mathbb{R}_+, \qquad X_u^{(k)} \stackrel{\triangle}{=} X_{s_k+u}^{(0)},$$

where $s_k = \sum_{i=1}^k \gamma_i$.

We can rewrite the equation (3.2) using the stepwise functions for every $u \in [s_k, s_{k+1})$

$$\begin{aligned} X_u^{(0)} &= X_0 - \int_0^{s_k} h(X_t^{(0)})dt + \sum_{s=1}^{k} \gamma_s(\Delta M_s + \eta_s), \\ X_u^{(k)} &= X_0^{(k)} - \int_0^{u} h(X_t^{(k)})dt + R_u^{(k)}, \qquad \forall\, k \in \mathbb{N}, \end{aligned}$$

where

$$\begin{aligned} R_u^{(k)} \stackrel{\triangle}{=} & \int_{s_{k+l}}^{s_k+u} h(X_t^{(0)})dt + \sum_{s=k+1}^{k+l} \gamma_s(\Delta M_s + \eta_s) \\ & \forall\, k \in \mathbb{N} \text{ and } u \in [s_{k+l}, s_{k+l+1}), \end{aligned}$$

where $R_u^{(k)}$ can be seen as a reminder term, then if we state the following assumption

$$(\mathbf{R}) \equiv \begin{cases} (i.) \sum_{s=1}^{+\infty} \gamma_s \Delta M_s < +\infty \\ \\ (ii.) \lim_{s\to\infty} \eta_s = 0 \end{cases} .$$

Assuming that $(h(X_k))_{k\geq 0}$ is bounded, i.e.:

$$L \stackrel{\triangle}{=} \sup_{k\geq 0} \|h(X_k)\| < +\infty,$$

then it is straightforward that for every $K \in \mathbb{R}_+$

$$\sup_{u\in[0,K]} \|R_u^{(k)}\| \xrightarrow[k\to+\infty]{} 0.$$

Let us define

$$\phi_u^{(k)} \stackrel{\triangle}{=} \int_0^u h(X_t^{(k)})dt, \qquad u \in \mathbb{R}_+.$$

Therefore by the boundedness of $(h(X_k))_{k\geq 0}$ the $\phi^{(k)}$ are L-Lipschitz. Hence by the Arzelà-Ascoli theorem the sequence converge uniformly on compacts sets in $\mathbb{R}_+$ and that turns on to the boundedness of the sequence $(X_k)_{k\geq 0}$. Then $X^{(k)} = X_k - \phi^{(k)} + R^{(k)}$ also converge uniformly on compacts sets in $\mathbb{R}_+$, so by the previous assumptions and the formulation in equation (3.2) it follows that $X_{k+1} - X_k \to 0$ and the limiting points of $(X_k)_{k\geq 0}$ is a compact connected set. Under the above assumptions it holds

***Proposition* 3.2**

1. *If both sequences $(h(X_k))_{k\geq 0}$ and $(X_k)_{k\geq 0}$ are bounded and if assumption* **R** *holds, then:*
   - *The set of limiting points of $(X_k)_{k\geq 0}$ is a non-empty compact set.*
   - *$(X^{(k)})_{k\geq 0}$ converge uniformly on compacts set in $\mathbb{R}_+$.*
2. *Moreover if h is a continuous function, then every limit point of $(X^{(k)})_{k\geq 0}$ is a bounded continuous solution of the ordinary differential equation*

$$\dot{x} = -h(x) \quad x(0) = x_0^\infty,$$

where x_0^∞ is a limiting point of $(X_k)_{k\geq 0}$.

The proof can be found in Fort and Pagès [1996]. The previous proposition is the first result related with the classical Kushner-Clark statement. The concept of *stable attracting area* is needed to state the main theorem

Definition* 3.2 *Stable attracting area
Let $\dot{x} = -h(x)$ an ordinary differential equation. A neighborhood G_{x^*} of an equilibrium point x^* (i.e., $x^* \in \{h = 0\}$) of the ODE is a stable attracting area, if the followings conditions are satisfied:

1. *For every solution* $x^{(0)}$ *of the ODE starting at* x^0, $\forall\, x^0 \in G_{x^*}$, $\forall\, u \in \mathbb{R}_+$, $x^{(0)}(u) \in G_{x^*}$.
2. *For every solution* $x^{(0)}$ *of the ODE starting at* x^0, $\forall\, x^0 \in G_{x^*}$, $\lim_{u\to+\infty} x^{(0)}(u) = x^*$.
3. *For all compact set* K, $K \subset G_{x^*}$, $\forall\, \epsilon > 0$, $\exists \eta_{\epsilon,K} > 0$, *such that* $\forall x^0 \in K \bigcap B(x^*, \eta_{\epsilon,K})$ *and for every solution* $x^{(0)}$ *of the ODE starting at* x^0,

$$\sup_{u\in\mathbb{R}_+} \|x^{(0)}(u) - x^*\| \leq \epsilon.$$

A sufficient condition for the existence of G_{x^*} is that the function h is differentiable at x^* and that all the eigenvalues of its gradient $\nabla h(x^*)$ have a positive real part.

Now, let us recall the Kushner-Clark theorem which is a conditional convergence theorem. It identifies the possible limits as attractive equilibrium points of the associated ODE and states the conditional convergence as soon as a boundedness condition holds for the sequence, see Kushner and Clark [1978] for more information, then

***Theorem* 3.6**
If x^* *is an equilibrium point of the ODE* $\dot{x} = -h(x)$, h *a continuous function, the sequence* $(X_k)_{k\geq 1}$ *in (3.1) is bounded with* $\sum_{k\geq 1} \gamma_k = +\infty$ *and* $\sum_{k\geq 1} \gamma_k^2 < +\infty$ *and the assumption* **R** *holds, then*

$$X_k \to x^*, \qquad as \quad k \to +\infty,$$

on the event $A_K^{x^*} \stackrel{\triangle}{=} \{X_k \in K \subset G_{x^*}$ *infinitely often*$\}$, *where* K *is a compact set.*

In practice, the Kushner-Clark theorem needs the boundedness of the solutions $(X_k)_{k\geq 1}$ and the continuity of h. In some cases the choice of h leads to the boundedness of the solution (see for instance Fort and Pagès [1996]). Specifically the existence of Lyapunov functions will ensure the convergence.

Definition* 3.3 *Lyapunov function
A function V *is called* Lyapunov function *if satisfies the following:*
i. $V : \mathbb{R}^d \to \mathbb{R}_+$ *is continuously differentiable.*
ii. $\lim_{x\to\infty} V(x) = +\infty$, *so that* $\forall\, v \in \mathbb{R}^d$, $\{V \leq v\}$ *is a compact set.*
iii. $x \mapsto (\nabla V|h)(x)$ *is non negative.*

***Theorem* 3.7**
Assume that the sequence $(\gamma_k)_{k\geq 1}$, *the gain of the algorithm satisfies,*

$$\sum_{k\geq 1} \gamma_k = +\infty \qquad \sum_{k\geq 1} \gamma_k^2 < +\infty.$$

Assume that h *is continuous,* V *satisfies Definition 3.3 and that, furthermore, they fulfil*
1. ∇V *is* $[\nabla V]$*-Lipschitz and for every* $x \in \mathbb{R}^d$, $\|\nabla V(x)\|^2 \leq A(1 + V(x))$.
2. $\{(\nabla V|h) = 0\} = \{h = 0\}$ *and* $\forall\, v \in \mathbb{R}_+$, $\{h = 0\} \cap \{V \leq v\}$ *is finite.*
3. *Furthermore, assume that* $(\omega_k)_{k\geq 1}$, $(\eta_k)_{k\geq 1}$, H *and* h *satisfy for every* $k \in \mathbb{N}$*:*
   i. $\mathbb{E}(\|H(X_k, \omega_{k+1})\|^2|\mathcal{F}_k) \leq A(1 + V(X_k))$.
   ii. $\|h(X_k) - \mathbb{E}(H(X_k, \omega_{k+1})|\mathcal{F}_k)\|^2 + \mathbb{E}(\|\eta_{k+1}\|^2|\mathcal{F}_k) \leq A\gamma_{k+1}^2(1 + V(X_k))$.

then, almost surely, there exists $x^* \in \{h = 0\}$ *such that* $\lim_{k\to+\infty} X_k = x^*$.

Chapter 4

ACLVQ. Mixing CLVQ and Lloyd

We propose here a new method: the *Average Competitive Learning Vector Quantization*(ACLVQ) as a simple modification of the classical *Competitive Learning Vector Quantization*(CLVQ. We describe the scheme and discuss the convergence of the method. A simulation study was carried out to evaluate the performance of the method.

4.1 Scheme for the ACLVQ

The drawback of the CLVQ is the computation time due to the algorithm itself. This method only changes one quantizer (or codebook) at each iteration, and thus needs a large computation time to achieve good results.

We propose to generate a set of random vectors ξ instead of one as the CLVQ method do. In the *competitive phase* of the CLVQ, a single random vector is generated to find the "winning index". Our version changes that in generating a set of N random vectors. If we have the n-quantizer $x^1, \cdots, x^n$, where $x^i \in \mathbb{R}^d$, $d \geq 2$, $i = 1, \cdots, n$. After that we identify the nearest random vectors to each quantizer using some distance criterium. Then, in the *learning phase*, we take the mean of all groups formed in the previous step. Therefore, we obtain for each element of the n-quantizer (in the case that exists at least one random vector in its Voronoï tessel) a new random vector that is the mean of vectors fallen in its tessel.

In other words, the ACLVQ changes at least one quantizer each time, but if we generate a large number of random vectors at each iteration, then we can change all the elements of the quantizer at the same time. Obviously, the choice of the number of random vectors strongly depends on the number of quantizers.

Steps of the ACLVQ

1. Generate the initial n-quantizers $x_0^1, \cdots, x_0^n$.
2. Assign $x^i = x_0^i$ for all i.
3. Generate a set of *i.i.d.* random vectors $\xi = (\xi_1, \cdots, \xi_N)$ (with the same distribution $\mathbb{P}$).

4. Calculate the distance matrix between the n-quantizer and the random vectors (***competitive phase***).

   The procedure needed is described in the appendix, a new method for the calculation of Euclidean distance matrix between two sets of vectors in dimension greater than 2 is explained.

5. Find the groups for each quantizer in the step k.

   This is actually like to find the Voronoï region for each quantizer using for that the random vectors, i.e.:

$$I_k^j = \{i : \|\xi_i - x_k^j\| < \|\xi_i - x_k^m\|,\ \forall\, m \neq j\}, \quad j = 1, \cdots, n,\ i = 1, \cdots, N.$$

   That is the sequence of index associated to each quantizer. Obviously it is possible that it exists at least one $j : I_k^j = \emptyset$.

6. Calculate the mean in each group formed in the previous step. The principle is similar to find the mass centroid of a Voronoï region

$$\begin{cases} \widetilde{\xi}_j = \frac{\sum_{i \in I_k^j} \xi_i}{|I_k^j|} & I_k^j \neq \emptyset \\ \\ \widetilde{\xi}_j = 0 & I_k^j = \emptyset, \end{cases}$$

   where $|A|$ denotes the cardinal of the set A.

7. Update the quantizers (***learning phase***)

$$\begin{cases} x_{k+1}^i = x_k^i - \gamma_{k+1}(x_i^k - \widetilde{\xi}_i), & I_k^i \neq \emptyset \\ \\ x_{k+1}^i = x_k^i, & I_k^i = \emptyset \end{cases}$$

   where $\gamma_k > 0$ is the gain sequence as usual.

8. Repeat steps 2-7 until reach some convergence criteria.

To give the asymptotic behaviour of the *Average Competitive Learning Vector Quantization* we will write it in the Kushner-Clark settings.

4.2 Theoretical formulation of the ACLVQ

Let us formalize the previous scheme. If we define $\mathbf{X} = (X^i)_{1 \leq i \leq n}$, where $X^i \in \mathbb{R}^d$ for all $i = 1, \cdots, n$ and γ_k is the gain sequence as before, then the general model can be written as:

$$X_{k+1}^i = X_k^i - \gamma_{k+1} H_i(\mathbf{X}_k, \boldsymbol{\omega}_{k+1}), \tag{4.1}$$

where $H_i(\mathbf{X}_k, \boldsymbol{\omega}_{k+1}) = X_k^i - \Theta_i(\mathbf{X}_k, \boldsymbol{\omega}_{k+1})$, $(\boldsymbol{\omega}_k)_{k \geq 1}$, is a sequence of random vectors in $\mathbb{R}^{Nd}$: $\boldsymbol{\omega}_k = (\omega_k^1, \cdots, \omega_k^N)$, with $\omega_k^j \overset{i.i.d}{\sim} \mathbb{P}$ for all k and $j = 1, \cdots, N$, $N \in \mathbb{N}$. Therefore

$\boldsymbol{\omega}_k \sim \mathbb{P}_N = \bigotimes_{j=1}^N \mathbb{P}$ and obviously $\mathbf{X}_k$ is of the same kind as $\mathbf{X}$. The function Θ_i is defined by

$$\Theta_i(\mathbf{X}_k, \boldsymbol{\omega}_{k+1}) = \begin{cases} \frac{\sum_{j=1}^N \mathbf{1}_{C_i(\mathbf{X}_k)}(\omega_{k+1}^j)\omega_{k+1}^j}{\sum_{j=1}^N \mathbf{1}_{C_i(\mathbf{X}_k)}(\omega_{k+1}^j)} & \text{if } \sum_{j=1}^N \mathbf{1}_{C_i(\mathbf{X}_k)}(\omega_{k+1}^j) > 0 \\ 0 & \text{if } \sum_{j=1}^N \mathbf{1}_{C_i(\mathbf{X}_k)}(\omega_{k+1}^j) = 0, \end{cases}$$

where $C_i(\mathbf{X}_k)$ is the Voronoï region of X_k^i. The model can be rewritten as

$$X_{k+1}^i = \begin{cases} X_k^i - \gamma_{k+1}\left(X_k^i - \frac{\sum_{j=1}^N \mathbf{1}_{C_i(\mathbf{X}_k)}(\omega_{k+1}^j)\omega_{k+1}^j}{\sum_{j=1}^N \mathbf{1}_{C_i(\mathbf{X}_k)}(\omega_{k+1}^j)}\right) & \text{if } \sum_{j=1}^N \mathbf{1}_{C_i(\mathbf{X}_k)}(\omega_{k+1}^j) > 0 \\ X_k^i & \text{if } \sum_{j=1}^N \mathbf{1}_{C_i(\mathbf{X}_k)}(\omega_{k+1}^j) = 0 \end{cases}. \quad (4.2)$$

In the case of $N = 1$ we retrieve the classical CLVQ.

Our goal is the proof of the asymptotic convergence of the solution of (4.2) toward one "optimal" or stationary quantizer of the distribution. We use the Kushner-Clark theorem to prove the convergence of the algorithm. In order to use the theorem we must show some continuity property of the function $h_i(\mathbf{X}_k) = \mathbb{E}\left[H_i(\mathbf{X}, \boldsymbol{\omega}_{k+1})/\mathbf{X} = \mathbf{X}_k\right]$.

We have:

$$H_i(\mathbf{X}_k, \boldsymbol{\omega}_{k+1}) = \begin{cases} X_k^i - \frac{\sum_{j=1}^N \mathbf{1}_{C_i(\mathbf{X}_k)}(\omega_{k+1}^j)\omega_{k+1}^j}{N_k^i} & \text{if } N_{k+1}^i > 0 \\ X_k^i & \text{if } N_{k+1}^i = 0, \end{cases}$$

where $N_{k+1}^i = \sum_{j=1}^N \mathbf{1}_{C_i(\mathbf{X}_k)}(\omega_{k+1}^j)$, N_{k+1}^i is the number of times that a random vector lies in the Voronoï region of X_k^i. Furthermore

$$\begin{aligned} \mathbb{P}(N_{k+1}^i = 0) &= \mathbb{P}(\omega_{k+1}^j \notin C_i(\mathbf{X}_k)),\ \forall j = 1, 2, \cdots, N \\ &= (\mathbb{P}(\omega_{k+1}^1 \notin C_i(\mathbf{X}_k)))^N = (1 - \mathbb{P}(\omega_{k+1}^1 \in C_i(\mathbf{X}_k)))^N \\ &= (1 - \mathbb{P}(C_i(\mathbf{X}_k)))^N. \end{aligned}$$

Then $\mathbb{P}(N_{k+1}^i > 0) = 1 - (1 - \mathbb{P}(C_i(\mathbf{X}_k)))^N$.

Now, let us calculate $h_i(\mathbf{X}_k)$. From now on, to simplify the notations, we use $\boldsymbol{\omega}$ instead $\boldsymbol{\omega}_k$.

$$\begin{aligned} h_i(\mathbf{X}_k) &= X_k^i \mathbb{P}(N_{k+1}^i > 0) - \int\limits_{\{N_{k+1}^i > 0\}} \frac{1}{N_{k+1}^i} \sum_{j=1}^N \mathbf{1}_{C_i(\mathbf{X}_k)}(\omega_{k+1}^j)\omega_{k+1}^j \mathbb{P}_N(d\boldsymbol{\omega}) \\ &= X_k^i \mathbb{P}(N_{k+1}^i > 0) - \sum_{r=1}^N \frac{1}{r} \int\limits_{\{N_{k+1}^i = r\}} \sum_{j=1}^N \mathbf{1}_{C_i(\mathbf{X}_k)}(\omega_{k+1}^j)\omega_{k+1}^j \mathbb{P}_N(d\boldsymbol{\omega}). \end{aligned}$$

Now if we denote

$$\Omega^r = \{\omega_{k+1}^1 \in C_i(\mathbf{X}_k), \cdots, \omega_{k+1}^r \in C_i(\mathbf{X}_k), \omega_{k+1}^{r+1} \notin C_i(\mathbf{X}_k) \cdots, \omega_{k+1}^N \notin C_i(\mathbf{X}_k)\},$$

knowing that $\omega_k^j \overset{i.i.d}{\sim} \mathbb{P}$ it follows

$$\begin{aligned} h_i(\mathbf{X}_k) &= X_k^i \mathbb{P}(N_{k+1}^i > 0) - \sum_{r=1}^{N} \frac{1}{r} \binom{N}{r} \sum_{j=1}^{r} \int_{\Omega^r} \mathbf{1}_{C_i(\mathbf{X}_k)}(\omega_{k+1}^j) \omega_{k+1}^j \mathbb{P}_N(d\boldsymbol{\omega}) \\ &= X_k^i \mathbb{P}(N_{k+1}^i > 0) - \sum_{r=1}^{N} \binom{N}{r} \int_{\Omega^r} \mathbf{1}_{C_i(\mathbf{X}_k)}(\omega_{k+1}^1) \omega_{k+1}^1 \mathbb{P}_N(d\boldsymbol{\omega}). \end{aligned}$$

By the independence of ω^j it follows that

$$\int_{\Omega^r} \mathbf{1}_{C_i(\mathbf{X}_k)}(\omega^1)\omega^1 \mathbb{P}_N(d\boldsymbol{\omega}) = \mathbb{E}\mathbf{1}_{C_i(\mathbf{X}_k)}(\omega^1)\omega^1 \cdot \mathbb{P}^{r-1}(C_i(\mathbf{X}_k))(1 - \mathbb{P}(C_i(\mathbf{X}_k)))^{N-r}.$$

Therefore

$$\begin{aligned} h_i(\mathbf{X}_k) &= X_k^i \mathbb{P}(N_{k+1}^i > 0) - \frac{\mathbb{E}\mathbf{1}_{C_i(\mathbf{X}_k)}(\omega^1)\omega^1}{\mathbb{P}(C_i(\mathbf{X}_k))} \sum_{r=1}^{N} \binom{N}{r} \mathbb{P}^{r-1}(C_i(\mathbf{X}_k))(1 - \mathbb{P}(C_i(\mathbf{X}_k)))^{N-r} \\ &= \left[X_k^i - \frac{\mathbb{E}\mathbf{1}_{C_i(\mathbf{X}_k)}(\omega^1)\omega^1}{\mathbb{P}(C_i(\mathbf{X}_k))} \right] \mathbb{P}(N_{k+1}^i > 0). \end{aligned}$$

Defining

$$U_i(\mathbf{X}_k) = \frac{\mathbb{E}\mathbf{1}_{C_i(\mathbf{X}_k)}(\omega)\omega}{\mathbb{P}(C_i(\mathbf{X}_k))} = \frac{\int_{C_i(\mathbf{X}_k)} \omega \mathbb{P}(d\omega)}{\int_{C_i(\mathbf{X}_k)} \mathbb{P}(d\omega)} = \mathbb{E}\left[\mathbf{X}/\mathbf{X} \in C_i(\mathbf{X}_k)\right], \qquad \forall i = 1, 2, \cdots, n,$$

where $\omega \sim \mathbb{P}$. We have

$$h_i(\mathbf{X}_k) = \mathbb{E}\left[H_i(\mathbf{X}, \boldsymbol{\omega}_{k+1})/\mathbf{X} = \mathbf{X}_k\right] = (1 - (1 - \mathbb{P}(C_i(\mathbf{X}_k)))^N)(X_k^i - U_i(\mathbf{X}_k)). \tag{4.3}$$

From (4.3) it is quite obvious that $h_i(\mathbf{X}_k) = 0$ if and only if $X_k^i = U_i(\mathbf{X}_k)$. Which means that the zeros of $\mathbf{h} = (h_1, \cdots, h_n)$ are exactly the stationary quantizers.

In order to apply the Kushner-Clark theorem, we need to verify that h_i for $i = 1, \cdots, n$ is a continuous function. The following lemma ensures the continuity of $\mathbf{h}$.

***Lemma* 4.1**
The function $\mathbf{h} = (h_1, \cdots, h_n)$, *for* $\mathbf{x} = (x^1, \cdots, x^n)$, $x^i \in \mathbb{R}^d$ *for all* $i = 1, \cdots, n$ *is* $\mathbb{P}$*-a.s. continuous.*

Proof.

This is clear since $\mathbb{P}$ is supported by $\mathcal{O}$ ($\mathcal{O} \subset \mathbb{R}^d$ is an open set in $\mathbb{R}^d$): the functions $\int_{C_i(\mathbf{X}_k)} \mathbb{P}(d\omega)$, $\int_{C_i(\mathbf{X}_k)} \omega \mathbb{P}(d\omega)$ are continuous on $\mathcal{O}$ and the first one is positive on $\mathcal{O}$.

□

The convergence theorem reads as follows

Theorem* 4.1 *Convergence of the ACLVQ
If the sequence $\mathbf{X}_k = (X_k^1, \cdots, X_k^n)$ *defined by the stochastic algorithm given in equation (4.1) lives in a compact subset of* $\mathcal{O}$, *then it converges* $\mathbb{P}$*-a.s to an equilibrium point* $x^\star = (x^{\star,1}, \cdots, x^{\star,n})$ *of the function* $\mathbf{h}$, *which makes a stationary quantizer.*

Proof.

This is a direct consequence of the Kushner-Clark theorem.

□

***Remark* 4.1**
A way to improve the algorithm could be to choose at each step the number of random vectors N. So we may assume that $N = N(k)$, depending only on the step number (more sophisticated dependance could be studied). As soon as $N(k) > 1$ for all k the previous result holds.

***Remark* 4.2**
One can think that increasing N(k) will allow to stabilize the algorithm around a good optimal quantizer. However this case is not so interesting because we could reach a local minima.

In the examples on next section we show that we can achieve good results for small values of N_k^i, not always the same, because its value depends on the number of quantizers, as well as on the dimension of the problem.

4.3 Examples. A comparative study

In this section we carry out a simulation study of the ACLVQ method. We compare our results which the ones obtained by the classical CLVQ. We perform the comparison between both methods using our programs. We do not compare our results with these obtained by Pagès because due to the numerical implementation we do not achieve the same values for the quantization error.

Simulation study

In the sequel, d is the dimension of the problem, n is the number of quantizers, N is the number of random vectors generated in each iteration of the ACLVQ and K is the number of iterations performed for the ACLVQ method.

The simulation study has been organized in two parts

1. A brief simulation study to observe the relation of N and K.
2. A simulation study to compare the CLVQ method with the ACLVQ one.

1. First simulation study

For the first part we set

- $d = 2$, $n = 10, 14, 20$,
- $c = \{1/n, 1/2, 1, 2, 4, 5\}$ and $N_i^n = c_i \cdot n$, for $i = 1, \cdots, 6$ and
- the number of iterations for the CLVQ is $K_1^n = \{1 \cdot 10^6, 1.4 \cdot 10^6, 2 \cdot 10^6\}$ for $n = 10, 14, 20$ respectively.

We set the same initial vector x_0 for each n and all N_i^n. With the aim to obtain results of the same order the number of iterations of the ACLVQ has to change when i does for each n. In others words we need to set $K_i^n = K_1^n/N_i^n$ for all i. In that way we assure that we use the same number of random vector on the overall process for each combination of n and K_i^n. It is important to point out that CLVQ$\stackrel{\triangle}{=}$ACLVQ$(N = 1)$. In Table 4.1 we can see the results obtained in the simulation study. The quantization error could be seen in Figure 4.1. It is easy to check that the method proposed has robust performance.

$\mathbf{N_i^n}$	$\mathbf{n = 10}$	$\mathbf{n = 14}$	$\mathbf{n = 20}$
1	$\mathbf{0,329181}$	$\mathbf{0,244438}$	$\mathbf{0,176275}$
$\frac{\mathbf{1}}{\mathbf{2}} \cdot \mathbf{n}$	0,329624 ($\mathbf{0,0004}$)	0,244664 ($\mathbf{0,0002}$)	0,176750 ($\mathbf{0,0005}$)
$\mathbf{1 \cdot n}$	0,329627 ($\mathbf{0,0004}$)	0,246706 ($\mathbf{0,0023}$)	0,176634 ($\mathbf{0,0004}$)
$\mathbf{2 \cdot n}$	0,329358 ($\mathbf{0,0002}$)	0,246479 ($\mathbf{0,0020}$)	0,177140 ($\mathbf{0,0009}$)
$\mathbf{4 \cdot n}$	0,329322 ($\mathbf{0,0001}$)	0,247209 ($\mathbf{0,0028}$)	0,176684 ($\mathbf{0,0004}$)
$\mathbf{5 \cdot n}$	0,336227 ($\mathbf{0,0070}$)	0,246103 ($\mathbf{0,0017}$)	0,176821 ($\mathbf{0,0005}$)

Table 4.1: The quantization error and absolute value for each value of n

The absolute value is calculated between $\widehat{e^2}_{nN_i^n}$ and $\widehat{e^2}_{nN_1^n}$ assuming that this last one is the exact value. We calculate the values for $\widehat{e^2}_{nN_i^n}$ using a Monte Carlo method because with the ACLVQ we can not use the procedure to estimate $\widehat{e^2}_{nN_i^n}$ as the CLVQ do. The sample used was the same for all the cases (the size of the sample was 200 000 random vectors).

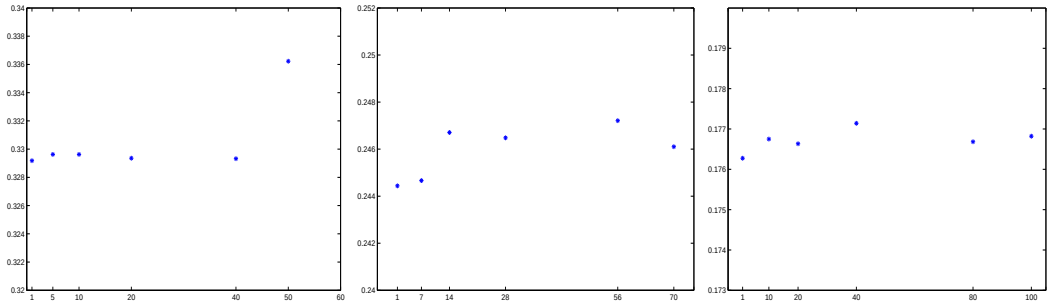

Figure 4.1: The quantization error for each n(10, 14, 20) and $N^n = (N_i^n)_{1 \le i \le 6}$.

In the Figures 4.2, 4.3 and 4.4 we can see that the quantizers obtained for each case are similar.

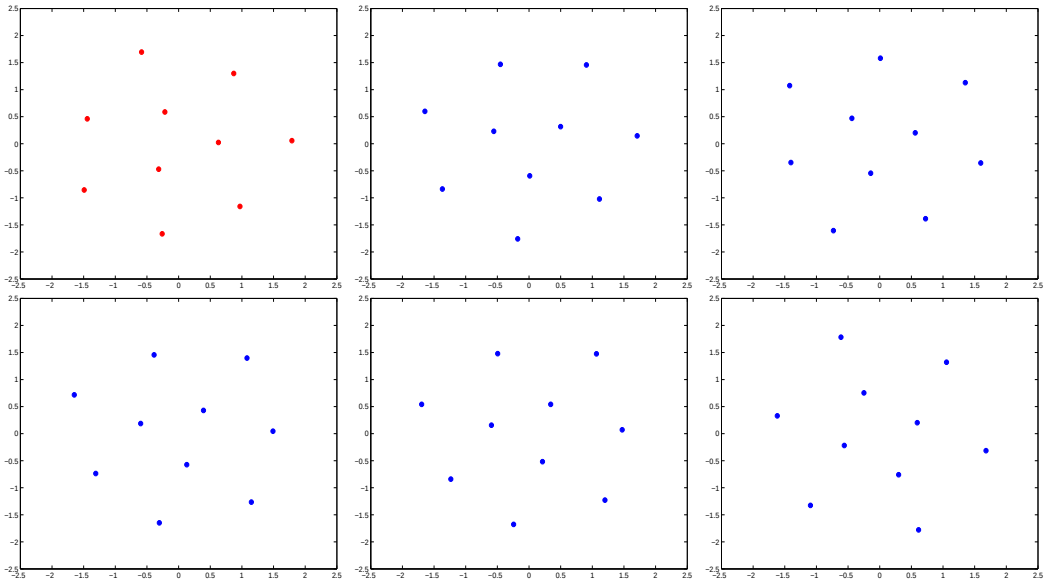

Figure 4.2: 10-quantizer for $d = 2$. In red the 10-quantizer obtained by the CLVQ method. In blue the 10-quantizer obtained by the ACLVQ method. Upper panel: from left to right the 10-quantizer with $N_1^{10}, N_2^{10}, N_3^{10}$ random vectors. Lower panel: from left to right the 10-quantizer with $N_4^{10}, N_5^{10}, N_6^{10}$ random vectors.

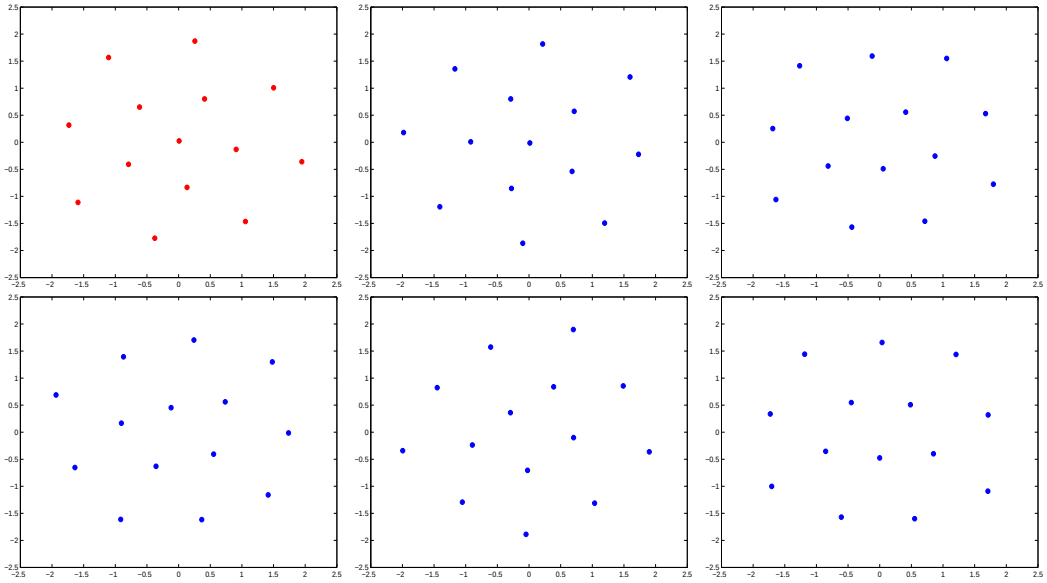

Figure 4.3: 14-quantizer for $d = 2$. In red the 14-quantizer obtained by the CLVQ method. In blue the 10-quantizer obtained by the ACLVQ method. **Upper panel**: from left to right the 14-quantizer with $N_1^{14}, N_2^{14}, N_3^{14}$ random vectors. **Lower panel**: from left to right the 20-quantizer with $N_4^{14}, N_5^{14}, N_6^{14}$ random vectors.

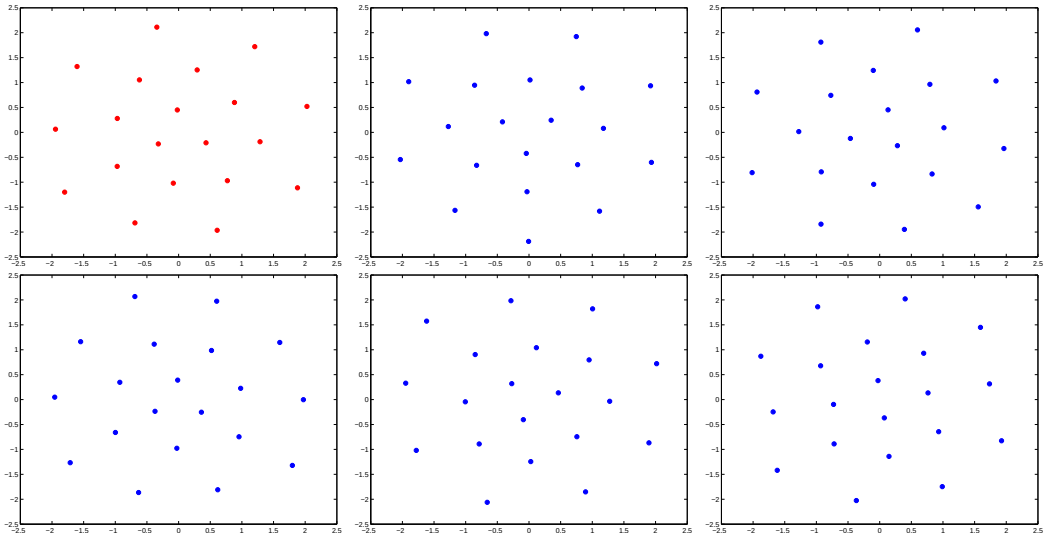

Figure 4.4: 20-quantizer for $d = 2$. In red the 20-quantizer obtained by the CLVQ method. In blue the 10-quantizer obtained by the ACLVQ method. Upper panel: from left to right the 20-quantizer with $N_1^{20}, N_2^{20}, N_3^{20}$ random vectors. Lower panel: from left to right the 20-quantizer with $N_4^{20}, N_5^{20}, N_6^{20}$ random vectors.

(d=10) In dimension 10 we only do a brief simulation study to observe the relation of N and K. We set

- $d = 10$, $n = 10, 20$,
- $c = \{1/n, 1/2, 1, 2, 4, 5\}$ and $N_i^n = c_i \cdot n$, for $i = 1, \cdots, 6$ and
- the number of iterations for the CLVQ is set $K_1^n = \{1 \cdot 10^6, 2 \cdot 10^6\}$ for $n = 10, 20$ respectively.

We set the same initial vector x_0 for each n and all N_i^n. The general idea remaind the same as for $d = 2$. In Table 4.2 we can see the results obtained in the simulation study. The quantization error is in Figure 4.5. The absolute value is calculated between $\widehat{e^2}_{nN_i^n}$ and $\widehat{e^2}_{nN_1^n}$ assuming that this last one is the exact value. We calculate $\widehat{e^2}_{nN_i^n}$ as before using a Monte Carlo method. The sample used was the same for all the cases (the size of the sample was 200 000 random vectors).

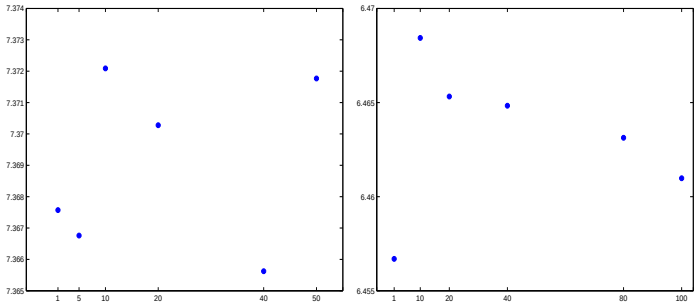

Figure 4.5: The quantization error for each n(10, 20) and $N^n = (N_i^n)_{1 \le i \le 6}$.

$\mathbf{N_i^n}$	$\mathbf{n = 10}$	$\mathbf{n = 20}$
1	$\mathbf{7,367572}$	$\mathbf{6,456694}$
$\frac{\mathbf{1}}{\mathbf{2}} \cdot \mathbf{n}$	7,366759 (0,0008)	6,468432 (0,0117)
$\mathbf{1 \cdot n}$	7,372088 (0,0045)	6,465321 (0,0086)
$\mathbf{2 \cdot n}$	7,370276 (0,0027)	6,464831 (0,0081)
$\mathbf{4 \cdot n}$	7,365623 (0,0019)	6,463129 (0,0064)
$\mathbf{5 \cdot n}$	7,371768 (0,0042)	6,460978 (0,0043)

Table 4.2: The quantization error and absolute value for each value of n

2. Second simulation study

The second step of the simulation study relies in the comparison between both methods. We set

- $d = 2$, $n = 10, 14$,
- $N^{10} = \{5, 10, 20\}$, $N^{14} = \{14, 28\}$ and
- $K^{10} = \{2 \cdot 10^5, 1 \cdot 10^5, 5 \cdot 10^4\}$,$K^{10}_{CLVQ} = 1 \cdot 10^6$; and $K^{14} = \{1 \cdot 10^5, 5 \cdot 10^4\}$, $K^{14}_{CLVQ} = 1.4 \cdot 10^6$.

We do for each combination 100 repetitions of the ACLVQ method with random initializations. The results for $n = 10$ can be seen in Figure 4.6 and Table 4.3. The quantization error for each N^{10} are in Figure 4.7.

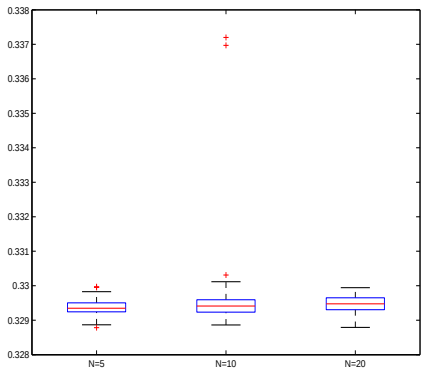

Figure 4.6: Box plot for the 100 samples used in the ACLVQ method for each value of N^{10}

$\mathbf{N^{10}}$	**Mean**	**STD**
5	0.3294	0.0002
10	0.3296	0.0011
20	0.3295	0.0003

Table 4.3: Mean and standard deviation for the 100 samples used in the ACLVQ method for each value of N^{10}.

The results for $n = 14$ are shown in Figure 4.8 and Table 4.4. The quantization error for each N^{14} are in Figure 4.9.

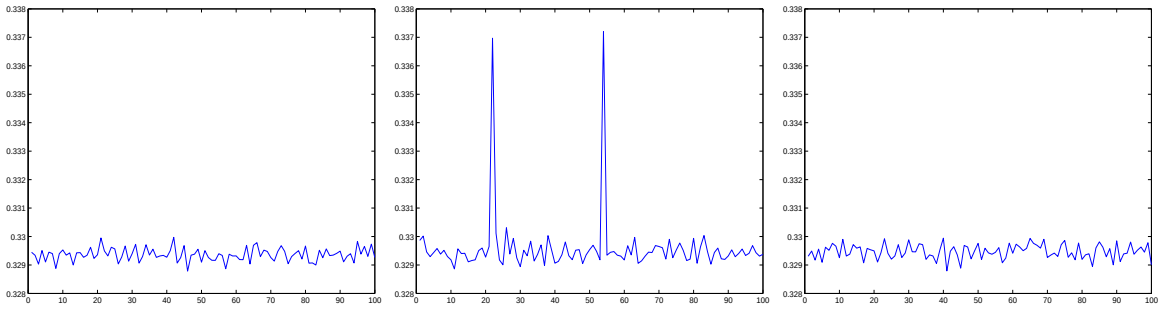

Figure 4.7: Quantization error for each N^{10} (5,10, 20).

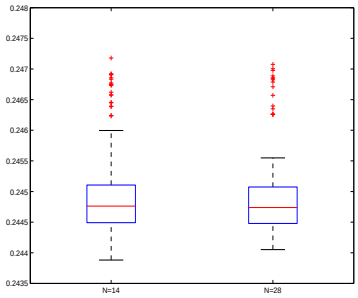

Figure 4.8: Box plot for the 100 samples used in the ACLVQ method for each value of N^{14}

$\mathbf{N^{14}}$	**Mean**	**STD**
14	0.24504	0.00086
28	0.24497	0.00078

Table 4.4: Mean and standard deviation for the 100 samples used in the ACLVQ method for each value of N^{14}.

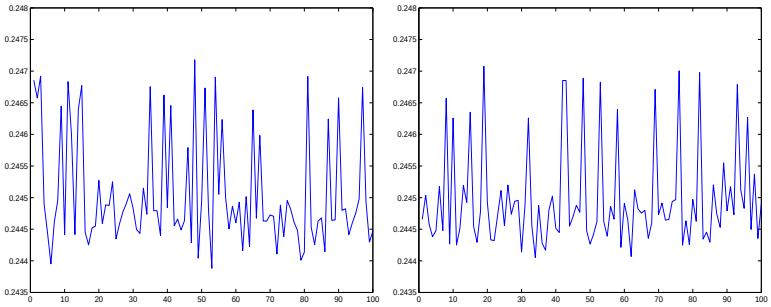

Figure 4.9: Quantization error for each N^{14} (14, 28).

Some considerations

This new method shows goods results in comparison with the classical CLVQ. The quantization error obtained by the ACLVQ is similar to the one obtained by the CLVQ, however we can not affirm that it is better. The advantage of our proposal is the computation time. We achieve similar results in less iterations and time that the CLVQ do. For that reason the ACLVQ could be used in real time problems considering its nature.

Chapter 5

Estimation of the Hurst parameter in fractional processes

In this chapter we introduce a method based on the Karhunen-Loève expansion of Gaussian process in order to estimate the Hurst parameter involved in fractional processes, specifically the fractional Brownian motion(fBm), the fractional Ornstein-Uhlenbeck(fOU) family and the fractional Brownian bridge(fBb). Our aim is to compare our results with the one obtained by the maximum likelihood method which show the validity of our proposal. The results presented here were accepted for publishing in the *Journal of Statistical Computation and Simulation*.

Our goal is to propose an estimator of the Hurst parameter of a fractional Gaussian process that performs almost as the maximum likelihood estimator but is a bit more robust to computational simulations.

Our framework is quite simple: we have observations of a random process at equally spaced or randomly distributed times, $X(t_1), \cdots, X(t_n)$. Assuming that the model which rules these observations is a gaussian fractional process (Brownian motion, Ornstein Uhlenbeck process, Brownian bridge) we estimate the Hurst parameter denoted by θ, $\theta \in (0, 1)$. We are strictly in the parametric estimation framework.

Obviously, the best estimator is obtain by maximizing the likelihood. It involves the numerical computation of the inverse of the covariance matrix and the logarithm of its determinant. The first is still tractable when n is of magnitude about 500 using LU decomposition, but the second is not.

Our alternative method is very simple: it relies on a simple result obtained via the Karhunen-Loève decomposition.

5.1 Estimation methods

In the sequel, for all the fractional process, there is one parameter of interest: θ which describes the roughness of the paths. For the fOU family we assume that $\mu = 0$ and θ and σ are known.

For the fBm, there are several methods for estimation of the Hurst parameter. In Dieker [2002] a general treatment is provided. In Jennane et al. [2001] the authors sum-

marizes some of these methods and offer numerical results. We also refer the reader to Coeurjolly [2001] and Cohen [2004].

The literature for the estimation of the Hurst parameter on fractional Ornstein-Uhlenbeck of first kind is small. The information provided for this process is more related with the estimation of the parameters involved into the Langevin equation. However, the maximum likelihood method (MLE) can be always used. For others members of the fOU family (second and third kind) the MLE works in the same way. There are others estimation methods from fractal theory for the fOU(3). (see Gneiting et al. [2011] for a deeper discussion).

5.2 General formulation

We will consider from now on that X belongs to one of the family (X^θ) of centered fractional Gaussian process with Hurst parameter θ in $(0,1)$. We assume that the observation time interval is $[0,1]$. Let us denote by θ_0 the real and unknown parameter of the process actually observe, X^{θ_0}. It is well-known that a Gaussian process has a unique representation in the basis of Karhunen-Loève. This basis is given by the diagonalization of the so-called covariance operator of the process X.

For each value of θ we denote by $(\lambda_k^\theta, \varphi_k^\theta)_{k\geq 1}$ the eigenvalues and associated normalized eigenfunctions of the covariance operator in decreasing order. Then exists an $i.i.d.$ sequence of standard normal random variables ξ_i^θ, such that in $\mathcal{L}^2[0,1]$:

$$X^\theta(t) = \sum_{k\geq 1} \sqrt{\lambda_k^\theta}\xi_k^\theta \varphi_k^\theta(t). \tag{5.1}$$

We observe the path of the process X^{θ_0} at $t_1, \cdots, t_n$ spaced or are $i.i.d.$ random variables independent of the process. Let $\mathbf{t} = (t_1, \cdots, t_n)'$, in view of we write:

$$X^\theta(\mathbf{t}) = \begin{pmatrix} X^\theta(t_1) \\ \vdots \\ X^\theta(t_n) \end{pmatrix}.$$

Hence we write

$$X^{\theta_0}(t) = \sum_{k\geq 1} \rho_k^{\theta,\theta_0} \varphi_k^\theta(t), \tag{5.2}$$

where

$$\rho_k^{\theta,\theta_0} = \int_0^1 X^{\theta_0}(t)\varphi_k^\theta(t)dt \quad \overset{\theta=\theta_0}{\Longrightarrow} \quad \rho_k^{\theta_0,\theta_0} = \sqrt{\lambda_k^{\theta_0}}\xi_k^{\theta_0}.$$

Now we define:

$$Y^{\theta,\theta_0}(\mathbf{t}) = (\Sigma^\theta)^{-\frac{1}{2}} X^{\theta_0}(\mathbf{t}),$$

where Σ^θ is the covariance matrix defined by

$$\Sigma^\theta = \left(\sum_{k\geq 1} \lambda_k^\theta \varphi_k^\theta(t_i)\varphi_k^\theta(t_j) \right)_{\substack{1\leq i\leq n \\ 1\leq j\leq n}}. \tag{5.3}$$

Clearly $Y^{\theta_0,\theta_0}(\mathbf{t}) \sim N(0, I_n)$ and $\|Y^{\theta_0,\theta_0}(\mathbf{t})\|^2/n$ *a.s.* converges toward 1.

The covariance function of X^{θ_0} using the basis $\{\varphi_k^\theta\}_{k\geq 1}$ is calculated as:

$$\begin{aligned} C^{\theta,\theta_0}(t,s) &= \mathbb{E}(X^{\theta_0}(t)X^{\theta_0}(s)) = cov\left(X^{\theta_0}(t), X^{\theta_0}(s)\right) \\ &= \mathbb{E}\sum_{\substack{l\geq 1\\ k\geq 1}} \langle X^{\theta_0}, \varphi_l^\theta\rangle\langle X^{\theta_0}, \varphi_k^\theta\rangle \varphi_l^\theta(t)\varphi_k^\theta(s) \\ &= \mathbb{E}\sum_{\substack{l\geq 1\\ k\geq 1}} \left(\int_0^1\int_0^1 X^{\theta_0}(u)X^{\theta_0}(v)\varphi_l^\theta(u)\varphi_k^\theta(v)dudv\right)\varphi_l^\theta(t)\varphi_k^\theta(s) \\ &= \sum_{\substack{l\geq 1\\ k\geq 1}} \left(\int_0^1\int_0^1 C^{\theta_0}(u,v)\varphi_l^\theta(u)\varphi_k^\theta(v)dudv\right)\varphi_l^\theta(t)\varphi_k^\theta(s), \end{aligned}$$

where $C^{\theta_0}(u,v)$ is the covariance function of X^{θ_0}. If we denote

$$A_{l,k}^{\theta,\theta_0} = \int_0^1\int_0^1 C^{\theta_0}(u,v)\varphi_l^\theta(u)\varphi_k^\theta(v)dudv, \tag{5.4}$$

the final expression for the covariance function of the process X^{θ_0} is

$$C^{\theta,\theta_0}(t,s) = \sum_{\substack{l\geq 1\\ k\geq 1}} A_{l,k}^{\theta,\theta_0}\varphi_l^\theta(t)\varphi_k^\theta(s). \tag{5.5}$$

Therefore the covariance matrix for $X^{\theta_0}(\mathbf{t})$ is

$$\Sigma^{\theta,\theta_0} = \left(\sum_{\substack{l\geq 1\\ k\geq 1}} A_{l,k}^{\theta,\theta_0}\varphi_l^\theta(t_i)\varphi_k^\theta(t_j)\right)_{\substack{1\leq i\leq n\\ 1\leq j\leq n}}. \tag{5.6}$$

For general θ the covariance matrix of Y^{θ,θ_0} is calculated as:

$$cov(Y^{\theta,\theta_0}(\mathbf{t})) = (\Sigma^\theta)^{-1/2}\Sigma^{\theta_0}(\Sigma^\theta)^{-1},$$

and we have:

$$\mathbb{E}\|Y^{\theta,\theta_0}(\mathbf{t})\|^2 = \mathbb{E}\left(X^{\theta_0}(\mathbf{t})'(\Sigma^\theta)^{-1}X^{\theta_0}(\mathbf{t})\right) = \mathrm{tr}(\Sigma^{\theta,\theta_0}(\Sigma^\theta)^{-1}).$$

Then we obviously define the contrast function h by:

$$h(\theta;\theta_0) = \lim_{n\to+\infty}\frac{\mathrm{tr}(\Sigma^{\theta,\theta_0}(\Sigma^\theta)^{-1})}{n} = \lim_{n\to+\infty}\frac{\mathbb{E}\|Y^{\theta,\theta_0}(\mathbf{t})\|^2}{n}. \tag{5.7}$$

It is straightforward that if $\theta = \theta_0$, then $\Sigma^{\theta_0,\theta_0} = \Sigma^{\theta_0}$ and

$$h(\theta_0;\theta_0) = \lim_{n\to+\infty}\frac{\mathrm{tr}(\Sigma^{\theta_0,\theta_0}(\Sigma^{\theta_0})^{-1})}{n} = \lim_{n\to+\infty}\frac{\mathrm{tr}(I_n)}{n} = \lim_{n\to+\infty}\frac{n}{n} = 1.$$

Obviously the function h take the value 1 when $\theta = \theta_0$ but we want to show a stronger result.

***Conjecture* 1**
Let X^θ be a fractional stochastic process with Hurst parameter $\theta \in (0,1)$ defined in $[0,1]$ and denote $X^{\theta_0}(\mathbf{t})$ for $\mathbf{t} = (t_1, \cdots, t_n)$ the partial observation of a path of the process X^{θ_0}. The times of observation $\mathbf{t}$ are assumed to be equally spaced or i.i.d random variables on $[0,1]$, independent of X^{θ_0}, with positive density on $[0,1]$. Then the function

$$h(\theta;\theta_0) = \lim_{n\to+\infty} \frac{\operatorname{tr}(\Sigma^{\theta,\theta_0}(\Sigma^\theta)^{-1})}{n},$$

satisfies the following property:

- ✓ *If $\theta = \theta_0$ then $h(\theta;\theta_0) = 1$.*
- ✓ *If $\theta > \theta_0$ then $h(\theta;\theta_0) = \infty$.*
- ✓ *If $\theta < \theta_0$ then $h(\theta;\theta_0) = 0$.*

The previous conjecture seems to be true but we could not have demonstrated it completely. For some particular cases, a full proof is provided. We carry out a simulation study in order to verify the validity of our proposal.

The previous result is obtained in a rather theoretical setting: the Karhunen-Loève decomposition is known and the limit $n \to \infty$ is taken. In practice we have to use an approximation of the Karhunen-Loève basis with finite n, we can only compute estimations of h, and finally it is not possible to estimate h at every possible value of θ. For the calculations needed to obtain the eigenvalues of each process we use the `eig` function provided by `MATLAB` which provides a good estimation of them. Using a time partition of $[0,1]$ of about 2000 points and the mentioned function we obtain the first 2000 eigenvalues and eigenfunctions of each process needed for the approximation of the Karhunen-Loève basis.

Now, to estimate h, we use $h_n(\theta;\theta_0) = \|Y^{\theta,\theta_0}(\mathbf{t})\|^2/n$, that we compute for "all" values θ in $(0,1)$.

Conditionally to the time values $\mathbf{t}$, $\|Y^{\theta_0,\theta_0}(\mathbf{t})\|^2$ is a χ^2 with n degrees of freedom, so that $\mathbb{E}\ h_n(\theta_0;\theta_0) = 1$ and $h_n(\theta_0;\theta_0)$ converges $a.s.$ toward 1.

Hence we propose to estimate the parameter θ_0 by:

$$\widehat{\theta_0} = \arg\min_\theta |h_n(\theta;\theta_0) - 1|.$$

To compute this estimator, it remains to numerically compute $(\Sigma^\theta)^{-1}$ which is possible even for about thousand points. We avoid the term $\log(\det\Sigma)$ that appears in the MLE estimation. In all the numerical experiments we observe a monotone behaviour of the empirical function h_n. We think that it is very unlikely that this empirical function attains more than one minimum, except may be for very small samples.

Then $\widehat{\theta_0}$ is computed by interpolation of the grid values. For a better understanding of the method proposed we describe the steps of the algorithm used for the simulation of a sample path of the process and then the calculation of the estimator.

Steps of the simulation and estimation procedures

We begin with the simulation algorithm of the process:

1. We set $n^\star = 2000$ and calculate the first $n^\star$ eigenvalues $(\widehat{\lambda_k^{\theta_0}})_{0\leq k\leq n^\star}$ and the eigenvectors $(\widehat{\varphi_k^{\theta_0}})_{0\leq k\leq n^\star}$ for a given θ_0 in the points of $\mathbf{t}^\star = (t_1^\star, \cdots, t_{n^\star}^\star)$ using the `eig` function provided by `MATLAB`.
2. We generate a set of $n^\star$ *i.i.d.* random variables $\xi = (\xi_1, \cdots, \xi_{n^\star})$ with a standard normal distribution.
3. We construct the fractional process X^{θ_0} by its Karhunen-Loève approximation.

$$X^{\theta_0}(t_i) \approx \sum_{k\geq 1}^{n^\star} \sqrt{\widehat{\lambda_k^{\theta_0}}}\xi_k \widehat{\varphi_k^{\theta_0}}(t_i), \qquad \forall i = 0, 1, \cdots, n^\star.$$

4. At this step we select a random set of points among the previous sample path: setting the value of n we take $X^{\theta_0}(\mathbf{t})$ for a uniformly chosen random set $\mathbf{t} = (t_1, \cdots, t_n) \subset \mathbf{t}^\star$.

The next three steps make the algorithm of estimation:

1. Calculate $(\Sigma^\theta)^{-\frac{1}{2}}$ for each $\theta \in \{0.05, 0.06, \cdots, 0.94, 0.95\}$ to obtain $Y^{\theta,\theta_0}(\mathbf{t}) = (\Sigma^\theta)^{-\frac{1}{2}} X^{\theta_0}(\mathbf{t})$.
2. Calculate the value of $h_n(\theta;\theta_0)$ for each θ.
3. Calculate the estimator $\widehat{\theta_0} = \text{argmin}_\theta |h_n(\theta;\theta_0) - 1|$ using a quadratic interpolation.

Some remarks about the estimation method

The simulation part in the previous description of the method is the most time consuming because it requires the creation of a covariance matrix and the extraction of its eigenvalues and eigenfunctions. This part of simulation is only needed for the creation of the process X^{θ_0} in the points $\mathbf{t}$ and do not appear when the observed process is provided.

With the aim to show the scope of our estimation method we provide the computation time needed to obtain $\widehat{\theta_0}$. For the fractional Brownian motion we generate 50 random samples for each combination of $n = 10, 50, 100, 150, 200, 300, 400, 500$ and $\theta_0 = 0.1$ (the choice of θ_0 do not significantly modify the computation time). Applying our estimation method we observe that

n	**Mean($\overline{\mathbf{T}}$)**	**STD**
10	0.1740	0.0097
50	1.1551	0.0133
100	5.1329	0.0275
150	13.4232	0.0465
200	26.4882	0.0794
300	73.7897	0.5203
400	160.1057	2.4224
500	293.8615	13.6876

Table 5.1: Mean and standard deviation for the estimation time in seconds of $\theta_0 = 0.1$ for each value of n by using 50 random observations of the fBm with the true parameter.

Regarding the mean of the computation time for each sample path length we see that for $n \leq 200$ to compute the estimator needs less than 30 seconds. A closer look of the observed data shows a quadratic behaviour of the mean time of the estimation algorithm versus the sample path length considered. By the least square method we obtain using the data of Table 5.1 that $\overline{T} \approx (4.\ 10^{-2}n\ -\ 3)^2$. All the computations were made with `MATLAB 7.5.0 R2007b` in a computer with the following characteristics: `CPU Intel Core 2 Duo P8400 2.26 GHz` and `2Gb RAM`.

5.3 Proof of the estimation theorem

We suppose that $\theta \neq \theta_0$. We begin by taking "approximations" of the matrices $\Sigma^{\theta}, \Sigma^{\theta,\theta_0}$. Let $N^\star$ an integer, we define the following matrices:

$$\Psi_{N^\star} = \left(\varphi_k^\theta(t_i)\right)_{\substack{1\leq i\leq n\\ 1\leq k\leq N^\star}} \quad A_{N^\star} = \left(A_{i,k}^{\theta,\theta_0}\right)_{\substack{1\leq i\leq N^\star\\ 1\leq k\leq N^\star}} \quad \Lambda_{N^\star} = \left(\lambda_{ik}^\theta\right)_{\substack{1\leq i\leq N^\star\\ N\leq k\leq N^\star}},$$

where $\lambda_{ik}^\theta = \lambda_k^\theta$ if $i = k$ and zero otherwise. Therefore we have the following approximations:

$$\Sigma_{N^\star}^\theta = \Psi_{N^\star}\Lambda_{N^\star}\Psi'_{N^\star} \quad \text{and} \quad \Sigma_{N^\star}^{\theta,\theta_0} = \Psi_{N^\star}A_{N^\star}\Psi'_{N^\star}.$$

Using well-known properties of the trace of a matrix and the fact that rank $\Sigma_{N^\star}^\theta = n$ it follows that $\operatorname{rank}(\Psi'_{N^\star}) = n$ ($a.s.$ if $\mathbf{t}$ is random and assuming that $N^\star \geq n$). Therefore it follows that

$$h(\theta;\theta_0) = \lim_{n\to+\infty} \frac{\operatorname{tr}(\Sigma_{N^\star}^{\theta_0}(\Sigma_{N^\star}^\theta)^{-1})}{n} = \lim_{n\to+\infty} \frac{\operatorname{tr}(A_{N^\star}\Lambda_{N^\star}^{-1})}{n} = \lim_{n\to+\infty} \frac{1}{n}\sum_{k\geq 1}^{N^\star} \frac{A_{k,k}^{\theta,\theta_0}}{\lambda_k^\theta}. \tag{5.8}$$

We conclude from (5.8) that we need to restrict the discussion on the quotient $A_{k,k}^{\theta,\theta_0}/\lambda_k^\theta$.

5.3.1 Fractional Brownian motion

In this case we will use the results of Bronski (Bronski [2003a]). This article provides expansion of the eigenvalues appearing in the Karhunen-Loève decomposition of the fBm, more precisely

$$\lambda_k^\theta = \frac{v_\theta}{k^{2\theta+1}} + o\left(k^{-\frac{(2\theta+2)(4\theta+3)}{4\theta+5}+\delta}\right), \qquad \forall\ \delta > 0, \quad k \gg 1,$$

where

$$v_\theta = \frac{\sin(\pi\theta)\Gamma(2\theta+1)}{\pi^{2\theta+1}},$$

and Γ is the usual Gamma function.

In the proof of the previous result the author define an integral operator $\mathbf{A}$ in the orthonormal basis $\{\sqrt{2}\sin\left(k+\frac{1}{2}\right)\pi x\}_{k=0}^\infty$, which is the Karhunen-Loève basis of the Wiener process. The $\mathbf{A}_{m,k}$ element for $\theta = \theta_0$ writes:

$$\mathbf{A}_{m,k} = \int_0^1\int_0^1 (u^{2\theta_0} + v^{2\theta_0} - |u-v|^{2\theta_0})\sin(m^*u)\sin(k^*v)dudv,$$

where $k^* = (k + 1/2)\pi$ and likewise m^*.

Bronski shows that $\mathbf{A}$ can be written as $\mathbf{A} = \mathbf{D} + \mathbf{O}$, where $\mathbf{D}$ is a diagonal piece and $\mathbf{O}$ an off-diagonal piece (the expressions for both are in Bronski [2003a]). In our case it is easy to see that

$$A_{m,k}^{1/2,\theta_0} = \mathbf{A}_{m,k}.$$

We have to study the quotient $A_{k,k}^{\theta,\theta_0}/\lambda_k^\theta$, hence we focus on the diagonal entries of $\mathbf{A}$. Therefore it follows

$$A_{k,k}^{1/2,\theta_0} = \frac{\sin(\pi\theta_0)\Gamma(2\theta_0+1)}{k^{*2\theta_0+1}} + O(k^{-2(\theta_0+1)}),$$

thus

$$\frac{A_{k,k}^{1/2,\theta_0}}{\lambda_k^{1/2}} = \sin(\pi\theta_0)\Gamma(2\theta_0+1)k^{2\theta_0-1} + O\Big(\frac{(k+1/2)^2}{k^{2(\theta_0+1)}}\Big).$$

Now it is straightforward that

$$\lim_{n\to+\infty}\frac{1}{n}\sum_{k\geq 1}^{N^\star}\frac{A_{k,k}^{1/2,\theta_0}}{\lambda_k^{1/2}} = \begin{cases} +\infty & \text{if } \theta_0 > 1/2 \\ 0 & \text{if } \theta_0 < 1/2 \end{cases}.$$

With this result we just get the result in the case $\theta = 1/2$, but the general setting follows the same way. Bronski [2003b] established an asymptotic expansion of the eigenvalues and eigenfunctions of the fBm

$$\begin{aligned} \lambda_k^\theta &\sim \frac{v_\theta}{k^{2\theta+1}} \qquad k \gg 1, \\ \varphi_k^\theta(t) &\sim \sqrt{2}\sin(k^*t) \qquad k \gg 1. \end{aligned} \tag{5.9}$$

Hence setting

$$\widetilde{\mathbf{A}}_{m,k} = \int_0^1\int_0^1 (u^{2\theta_0} + v^{2\theta_0} - |u-v|^{2\theta_0})\varphi_m^\theta(u)\varphi_k^\theta(v)dudv,$$

and using (5.9) we obtain

$$A_{k,k}^{\theta,\theta_0} = \widetilde{\mathbf{A}}_{k,k} \sim \mathbf{A}_{k,k} = \frac{\sin(\pi\theta_0)\Gamma(2\theta_0+1)}{k^{*2\theta_0+1}} + O(k^{-2(\theta_0+1)}). \tag{5.10}$$

That leads us to

$$\frac{A_{k,k}^{\theta,\theta_0}}{\lambda_k^\theta} \sim \sin(\pi\theta_0)\Gamma(2\theta_0+1)v_\theta^{-1}\frac{k^{2\theta+1}}{k^{*2\theta_0+1}} + O(\frac{k^{2\theta+1}}{k^{2(\theta_0+1)}}).$$

Therefore

$$\lim_{n\to+\infty}\frac{1}{n}\sum_{k\geq 1}^{N^\star}\frac{A_{k,k}^{\theta,\theta_0}}{\lambda_k^\theta} = \begin{cases} +\infty & \text{if } \theta_0 > \theta \\ 0 & \text{if } \theta_0 < \theta \end{cases}. \tag{5.11}$$

This ends the proof for the fBm.

5.3.2 Fractional Ornstein-Uhlenbeck processes

fOU(1)

Let $X^{\theta,1}$ be a fOU(1). Here we assume that α and σ are known.

There is no closed form for the covariance function or its eigenvalues, but one can relate them to the corresponding values for the fBm.

Let us denote $\Gamma^{\theta,B}$, $\{\varphi_k^{\theta,B}(t)\}_{k\geq 1}$ and $(\lambda_k^{\theta,B})_{k\geq 1}$ the covariance function, KL eigenfunctions and eigenvalues respectively for the fBm, and h_n^B the associated estimation function. The process $X^{\theta,1}$ can be expanded on $\{\varphi_k^{\theta,B}(t)\}_{k\geq 1}$, which is an orthonormal basis of $\mathcal{L}^2([0,1])$. Then

$$X^{\theta_0,1}(t) = \sum_{k\geq 1} \rho_k^{\theta,\theta_0} \varphi_k^{\theta,B}(t),$$

where

$$\rho_k^{\theta,\theta_0} = \int_0^1 X^{\theta_0,1}(t)\varphi_k^{\theta,B}(t)dt.$$

Therefore, the expression for $A_{l,k}^{\theta,\theta_0}$ in (5.4) still takes the form

$$\widehat{A}_{l,k}^{\theta,\theta_0} = \int_0^1 \int_0^1 \Gamma^{\theta_0,X}(u,v)\varphi_l^{\theta,B}(u)\varphi_k^{\theta,B}(v)dudv,$$

where $\Gamma^{\theta_0,X}$ is the covariance function of the fOU and

$$\widehat{\Sigma}^{\theta,\theta_0} = \left(\sum_{\substack{l\geq 1 \\ k\geq 1}} \widehat{A}_{l,k}^{\theta,\theta_0} \varphi_l^{\theta,B}(t_i)\varphi_k^{\theta,B}(t_j) \right)_{\substack{1\leq i\leq n \\ 1\leq j\leq n}}.$$

As in (5.8) we have

$$h^X(\theta;\theta_0) = \lim_{n\to+\infty} h_n^X(\theta;\theta_0) = \lim_{n\to+\infty} \frac{\mathrm{tr}(\widehat{\Sigma}^{\theta,\theta_0}(\Sigma^\theta)^{-1})}{n} = \lim_{n\to+\infty} \frac{1}{n} \sum_{k\geq 1}^{N^\star} \frac{\widehat{A}_{k,k}^{\theta,\theta_0}}{\lambda_k^\theta}.$$

Obviously we do not have the same properties for h_n^X like before for h_n^B when $\theta = \theta_0$. The inequality for the covariance function of the fOU(1) in (1.6) yields to

$$c'_\theta \; A_{k,k}^{\theta,\theta_0,B} \leq \widehat{A}_{k,k}^{\theta,\theta_0} \leq C'_\theta \; A_{k,k}^{\theta,\theta_0,B}$$

$$d'_\theta \lambda_k^{\theta,B} \leq \lambda_k^\theta \leq D'_\theta \lambda_k^{\theta,B}$$

where the constants $c'_\theta, C'_\theta, d'_\theta, D'_\theta$ are positive. The $A_{k,k}^{\theta,\theta_0,B}$ are as in (5.10) and the $(\lambda_k^\theta)_{k\geq 1}$ are the eigenvalues of the covariance operator of the fOU(1). Hence

$$c''_\theta h_n^B(\theta;\theta_0) \leq h_n^X(\theta;\theta_0) \leq C''_\theta h_n^B(\theta;\theta_0),$$

and finally

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{k\geq 1}^{N^\star} \frac{A_{k,k}^{\theta,\theta_0,X}}{\lambda_k^\theta} = \begin{cases} +\infty & \text{if } \theta_0 > \theta \\ 0 & \text{if } \theta_0 < \theta \end{cases}, \tag{5.12}$$

Again we note the spatial scale invariance of our estimator.

Therefore Conjecture 1 for the fOU(1) is also true.

fOU(2)

For the $X^{\theta,2}$ we propose a different approach. This process is by definition a function of the fBm, let us recall the expression

$$X_t^{\theta,2} = e^{-\alpha t} W^{\theta}_{a_{t,\theta}},$$

where $a_{t,\theta} = \theta e^{\alpha t/\theta}/\alpha$. The main idea is to use the previous equation to revert the fOU(2) into a fractional Brownian motion. This method can also be used for the fOU(1). For both cases we can construct a fBm $W^{\theta}(\mathbf{t})$ with Hurst parameter $\theta \in (0,1)$ in $\mathbf{t} = (t_1, \cdots, t_n)$ and $\mathbf{t} = (e^{\theta t_1/H}, \cdots, {}^{\theta t_n/H})$ respectively. Conjecture 1 applies with the correct modifications for this process. This assertion state that a correct estimation of the Hurst parameter on a fBm leads us to a correct estimation for the associated fOU.

fOU(3)

For the fOU of third kind $X^{\theta,3}$ we know the exact expression of the covariance function. Hence the same procedure can be used. However we do not know asymptotics results for the eigenfunctions of the associated KL basis, only it is known an approximative result for the eigenvalues due to Rosenblatt

$$\lambda_k^{\theta,3} \sim k^{-(1+2\theta)},$$

we refer the reader to Rosenblatt [1963] and Widom [1964] for details. Therefore we are not in position to give a proof of the Conjecture 1 for this case. However in the simulation study the results confirms the validity of the Conjecture.

□

The Conjecture can be reads as

***Theorem* 5.1**
We assume that X^{θ} is a fractional Brownian motion or a fractional Ornstein-Uhlenbeck (first kind) with Hurst parameter $\theta \in (0,1)$ defined on $[0,1]$ and $X^{\theta_0}(\mathbf{t})$ for $\mathbf{t} = (t_1, \cdots, t_n)$ is the partial observation of a path of the process X^{θ_0}. The times of observation $\mathbf{t}$ are assumed to be equally spaced or i.i.d random variables on $[0,1]$, independent of X^{θ_0}, with positive density on $[0,1]$. Then the function

$$h(\theta;\theta_0) = \lim_{n\to+\infty} \frac{\mathrm{tr}(\Sigma^{\theta_0}(\Sigma^{\theta})^{-1})}{n},$$

satisfies the following property:

- ✓ *If $\theta = \theta_0$ then $h(\theta;\theta_0) = 1$.*
- ✓ *If $\theta > \theta_0$ then $h(\theta;\theta_0) = \infty$.*
- ✓ *If $\theta < \theta_0$ then $h(\theta;\theta_0) = 0$.*

5.4 Simulation study and numerical results

We carry out a simulation study to compare the estimates from our method (KL method) with those of the MLE method. Both procedures were used to give estimates of the Hurst parameter in fractional processes. This study shows that the KL method performs well.

We work here with $h_n(\theta;\theta_0) = \|Y^{\theta,\theta_0}(\mathbf{t})\|^2/n$ the estimation of h, that we compute for "all" values θ in $(0,1)$. We leave the numerical implementation issue for the moment and we focus the discussion into the method itself. The simulation study has been organized in two parts

1. The first part concerns a simulation study with various values of n and only two representative values of the Hurst parameter (0.1, 0.7). This is done to compare the behaviour of the two estimators on a large range for n.
2. The second part concerns a study that compares the behaviour of the estimators with respect to the Hurst parameter. We chose a midrange value of n ($n = 200$), and values of the Hurst parameter varying from 0.06 to 0.9.

We do the first simulation only for the fBm. We assume the same behaviour for the others process due to the similitude between the asymptotics behaviour of their eigenvalues.

5.4.1 Fractional Brownian motion

For the fBm we set

- $n = 10; 50; 100; 150; 200; 300$, and
- $\theta_0 = 0.1$ and $\theta_0 = 0.7$.

The first value of θ_0 is close to 0 and is not easy to estimate. The second one illustrate an easier case. The value $n = 300$ is in the range of easy computation of the MLE.

For each combination of previous values we generate 50 random samples of the fractional process. Then we calculate the value of $h_n(\theta;\theta_0)$ for $\theta \in \{0.05, 0.06, \cdots, 0.15\}$ when $\theta_0 = 0.1$, and for $\theta \in \{0.65, 0.66, \cdots, 0.75\}$ when $\theta_0 = 0.7$. Finally we interpolate the results obtained for estimating θ_0 for which $h_n(\theta;\theta_0) = 1$. We use a parabolic interpolation because of the behaviour of h_n.

n	Mean	STD
10	0.1039	0.0455
50	0.0975	0.0285
100	0.0969	0.0169
150	0.0970	0.0122
200	0.0972	0.0090
300	0.0971	0.0070

n	Mean	STD
10	0.7178	0.0372
50	0.7054	0.0190
100	0.7027	0.0133
150	0.7026	0.0111
200	0.7010	0.0092
300	0.7003	0.0072

Table 5.2: Mean and standard deviation for the estimation of θ_0 for each value of n by using 50 random observations of the fBm with the true parameter. (Left: $\theta_0 = 0.1$. Right: $\theta_0 = 0.7$).

The numerical results in Table 5.2 allows to think that our estimator converges, and that the speed of convergence is of the correct magnitude. The choice of n should be for both values of θ_0 greater or equal to 100. According to the previous conclusion we perform a last simulation study to compare the performances of the KL method and the MLE.

The study provides quantitative results for different values of the parameter. The simulation environment allowed us to investigate the behaviour of our method and compare it with the MLE.

For the simulation scheme we set

- $\theta_0 \in \{0.06, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$,
- $n = 200$ for all values of θ_0.

The size of the random samples of the fBm is fixed at 100. For the KL method we calculate the value of $h_n(\theta; \theta_0)$ for $\theta \in \{\theta_0 - 0.05, \theta_0 - 0.04, \cdots, \theta_0 + 0.04, \theta_0 + 0.05\}$.

For the MLE we maximize the *log*-likelihood function

$$h_n^{\mathrm{MLE}}(\theta; \theta_0) = -\ln(\det(\Sigma^\theta(\mathbf{t}))) - (X_{\mathbf{t}}^{\theta_0})' \Sigma^\theta(\mathbf{t}) X_{\mathbf{t}}^{\theta_0},$$

where $X_{\mathbf{t_n}}^{\theta_0}$ is the random sample in the fixed points t_i and $\Gamma^\theta(\mathbf{t_n}) = (\Gamma^\theta(t_i, t_j))_{1 \le i,j \le n}$. The matrix $\Gamma^\theta(\mathbf{t_n})$ is positive definite for any $\mathbf{t}$ hence $\det(\Sigma^\theta(\mathbf{t})) > 0$, but it could be extremely small. So due to the small values of the eigenvalues associated to $\Gamma^\theta(\mathbf{t_n})$ we faced numerical difficulties: we observe that $\det(\Gamma^\theta(\mathbf{t_n})) = 0$ for a certain choice of θ and n. We avoid this numerical problem to compute the MLE replacing the previous function by

$$-\ln(\det(D_n^{\theta_0} \Sigma^\theta(\mathbf{t}))) - (X_{\mathbf{t}}^{\theta_0})' \Sigma^\theta(\mathbf{t}) X_{\mathbf{t}}^{\theta_0}.$$

Here $D_n^{\theta_0}$ is a numerical function which assures that $\det(D_n^{\theta_0} \Sigma^\theta(\mathbf{t})) > 0$ in the numerical implementation. Indeed using that $\det(aA) = a^k \det(A)$, where A is a square matrix of order k and $a \in \mathbb{R}$ it follows that

$$\ln(\det(\Gamma^\theta(\mathbf{t_n}))) = \ln(\det(D_n^{\theta_0} \Gamma^\theta(\mathbf{t_n}))) + n \ln(D_n^{\theta_0}),$$

where $D_n^{\theta_0}$ is invariant for all chosen θ. The choice for $D_{200}^{\theta_0}$ was determined numerically.

$$D_{200}^{\theta_0} = \begin{cases} 1 & \theta_0 \in (0,\, 0.2] \\ 10 & \theta_0 \in (0.2,\, 0.5) \\ 10^3 & \theta_0 \in [0.5,\, 0.7] \\ 10^5 & \theta_0 \in (0.7,\, 1) \end{cases}.$$

It is worth pointing out that $D_n^{\theta_0}$ actually depends on θ. In our case we use $|\theta - \theta_0| \le 0.05$ and under that hypothesis the numerical invariance of $D_{200}^{\theta_0}$ is assured.

The interpolation works in the same manner for both methods with the corresponding correction in each case. The results can be seen in Table 5.3.

The results of table 5.3 clearly show that the two methods are are numerically equivalent. Of curse the MLE estimator attains asymptotically the Cramer-Rao bound, but the KL method provides an almost precise estimator as the *M.L.E* do. This partial conclusion for the fractional Brownian motion has been confirmed by the study of fractional fOU(3) and the fBb.

θ_0	**Mean**(KL)	**STD**(KL)	**Mean**(MLE)	**STD**(MLE)
0.06	0.0601	0.0099	0.0596	0.0082
0.1	0.1010	0.0113	0.1010	0.0099
0.2	0.2021	0.0091	0.2015	0.0088
0.3	0.3016	0.0095	0.3015	0.0092
0.4	0.4012	0.0092	0.4004	0.0089
0.5	0.5002	0.0086	0.4997	0.0085
0.6	0.6013	0.0078	0.6008	0.0078
0.7	0.7005	0.0068	0.7000	0.0069
0.8	0.8009	0.0071	0.8002	0.0068
0.9	0.9010	0.0053	0.9002	0.0052

Table 5.3: Mean and standard deviation for several $\widehat{\theta_0}$ in the fBm using both methods.

5.4.2 Fractional Ornstein-Uhlenbeck family

fOU(1)

Here a simulation study is pointless. We can construct a fBm using a fOU(1) by the reversal formula of

$$X_{t_i}^{\theta_0,1} \approx e^{-\alpha t_i}\left(X_{t_0}^{\theta_0,1} + \sigma\sum_{j=1}^{n} e^{-\alpha t_j}\Delta W_{t_j}^{\theta_0}\right), \quad i = 1, \cdots, n.$$

Hence if $X_{\mathbf{t}}^{\theta_0,1}$ is observed we get

$$W_{t_i}^{\theta_0} \approx W_{t_{i-1}}^{\theta_0} + X_{t_i}^{\theta_0,1} - e^{-\alpha(t_i - t_{i-1})}X_{t_{i-1}}^{\theta_0,1}.$$

Obviously for the simulation study we construct first the fBm to obtain the fOU(1). Therefore the reversal formula is used. It is an obvious consequence that we will obtain the same fBm that was generated.

The subjacent idea is the parameter associated to W^{θ_0} and $X^{\theta_0,1}$ is the same. Hence there is no loss of generality to estimate θ_0 with $W^{\theta_0}(\mathbf{t})$ instead $X^{\theta_0,1}(\mathbf{t})$. The more important advantage of this method is that we use a simple procedure to construct a fBm and after we fall into the fBm setting and the results applies in the same manner.

fOU(2)

Let us recall the expression for the $X^{\theta,2}$

$$X_t^{H\theta,2} = e^{-\alpha t}W_{a_{t,\theta}}^{\theta},$$

where $a_{t,\theta} = \theta e^{\alpha t/\theta}/\alpha$. The idea is to use the previous equation to revert the fOU(2) into a fBm. We can construct a fBm $W^{\theta}(\mathbf{t})$ with Hurst parameter $\theta \in (0,1)$ in $\mathbf{t} = (t_1, \cdots, t_n)$. The construction of $W_{a_{\mathbf{t},\theta}}^{\theta}$ it is easily obtained using $W^{\theta}(\mathbf{t})$ by the self similarity of the fBm. Therefore a correct estimation of the Hurst parameter of a fBm leads us to a correct estimation for the associated fOU(2).

fOU(3)

In this case we perform a similar simulation study as before for the fBm. The first part related with the with the choice of n is avoided. The results in Table 5.4 shows the validity of our proposal.

We set $\theta_0 \in \{0.06, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$, $n = 200$ for all values of θ_0 and the MLE function $h_n^{\text{MLE}}(\theta; \theta_0)$ is exactly the same. For both methods we improved the numerical precision by a quadratic interpolation.

θ_0	**Mean**(KL)	**STD**(KL)	**Mean**(MLE)	**STD**(MLE)
0.06	0.0563	0.0194	0.0594	0.0127
0.1	0.0974	0.0131	0.0980	0.0112
0.2	0.2001	0.0111	0.1996	0.0103
0.3	0.3007	0.0112	0.3003	0.0105
0.4	0.4006	0.0078	0.3999	0.0081
0.5	0.5009	0.0082	0.5007	0.0078
0.6	0.6021	0.0083	0.6016	0.0078
0.7	0.7007	0.0075	0.7001	0.0075
0.8	0.8007	0.0063	0.8001	0.0062
0.9	0.9004	0.0055	0.8997	0.0054

Table 5.4: Mean and standard deviation for several $\widehat{\theta_0}$ in the fOU(3) using both methods.

5.4.3 Fractional Brownian bridge

The fractional Brownian bridge (fBb) B^θ with Hurst parameter $\theta \in (0, 1)$ is defined in [0, 1] by

$$B_t^\theta = W_t^\theta - \frac{|t|^{2\theta} + 1 - |t-1|^{2\theta}}{2} W_1^\theta,$$

where $\{W_t^\theta,\, t \geq 0\}$ is a fBm defined in [0, 1]. The exact expression of the covariance function is known. Thus the simulation scheme works as before for the fOU(3) and the fBm.

We set $\theta_0 \in \{0.06, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$, $n = 200$ for all values of θ_0 and the MLE function $h_n^{\text{MLE}}(\theta; \theta_0)$ is exactly the same. For both methods we improved the numerical precision by a quadratic interpolation. The results are in Table 5.5.

θ_0	**Mean**(KL)	**STD**(KL)	**Mean**(MLE)	**STD**(MLE)
0.06	0.0595	0.0113	0.0594	0.0099
0.1	0.0983	0.0097	0.0977	0.0084
0.2	0.2009	0.0100	0.2008	0.0096
0.3	0.3026	0.0082	0.3017	0.0082
0.4	0.4004	0.0097	0.3995	0.0093
0.5	0.5010	0.0083	0.5006	0.0079
0.6	0.6018	0.0087	0.6012	0.0084
0.7	0.7000	0.0078	0.6995	0.0076
0.8	0.8000	0.0068	0.7994	0.0066
0.9	0.9000	0.0054	0.8992	0.0054

Table 5.5: Mean and standard deviation for several $\widehat{\theta}_0$ in the fBb using both methods.

Chapter 6

Some asymptotics in functional quantization

This chapter is divided into two parts. The first one studies the asymptotic rate of the quantization error for a Gaussian processes with covariance function

$$\Gamma(t,s) = e^{-\theta|t-s|^2}.$$

In the second part, we discuss the main results of the chapter. We deal with the asymptotic behaviour of the maximal radius for the optimal quantizer of the Wiener process.

6.1 The squared exponential covariance function

The asymptotic behaviour for the quantization error of a Gaussian process X with squared exponential covariance function

$$\Gamma(t,s) = e^{-\theta|t-s|^2},$$

has not been studied, yet it appears frequently in applications using kriging.

6.1.1 Basic framework

Here we present some preliminaries to obtain the asymptotic rate. We denote X as a zero mean stochastic process and $(\lambda_k)_{k\geq 1}$ is the sequence of orderer eigenvalues associated to the covariance operator of the process. As usual $e_n^2(X)$ is the quantization error of X. We recall some results related with the upper bound of $e_n^2(X)$ stated in Luschgy and Pagès [2002] (see Section 4.1 therein).

$$e_n^2(X) \;\leq\; \sum_{j\geq m+1} \lambda_j + \inf\left\{ \sum_{j=1}^{m} \lambda_j n_j^{-2} :\; n_1, \cdots, n_m \in \mathbb{N} : \prod_{j=1}^{m} n_j \leq n \right\},$$

and

$$\inf\left\{ \sum_{j=1}^{m} \lambda_j n_j^{-2} :\; n_1, \cdots, n_m \in \mathbb{N} : \prod_{j=1}^{m} n_j \leq n \right\} \;=\; \sum_{j=1}^{m} \lambda_j z_j^{-2} = n^{-\frac{2}{m}} m \left(\prod_{j=1}^{m} \lambda_j \right)^{1/m},$$

where

$$z_j = n^{\frac{1}{m}} \lambda_j^{1/2} \left(\prod_{j=1}^{m} \lambda_j \right)^{-1/2m}.$$

Let us define

$$m_n^\star = \max \left\{ m \geq 1 : n^{\frac{2}{m}} \lambda_m \left(\prod_{j=1}^{n} \lambda_j \right)^{-1/m} \geq 1 \right\},$$

and

$$a_k = \frac{1}{2} \ln \left(\frac{\prod_{j=1}^{k} \lambda_j}{\lambda_k^k} \right) = \frac{k}{2} \ln \left(\frac{(\prod_{j=1}^{k} \lambda_j)^{1/k}}{\lambda_k} \right) = \frac{1}{2} \sum_{j=1}^{k} \ln \left(\frac{\lambda_j}{\lambda_k} \right).$$

Obviously $a_k \to +\infty$ when $k \to +\infty$ and it follows that

$$m_n^\star = \max \{ m \geq 1 : a_m \leq \ln n \}.$$

Furthermore

$$a_{m_n^\star} \leq \ln n < a_{m_n^\star + 1}.$$

6.1.2 Asymptotic rate of the quantization error

Under the conditions stated on Corollary 2 in Widom [1964] the asymptotic behaviour of the eigenvalues $(\lambda_k)_{k \geq 1}$ is $\ln \lambda_k \sim -k \ln k$. However we only need a result $\approx$ for the quantization error. Therefore, using the same argument on the proof of Corollary 4.13 (c) from Luschgy and Pagès [2002] that is based on Lemma 4.1.2 from Luschgy and Pagès [2002] allow us to assume without loss of generality that $\ln \lambda_k = -k \ln k$, hence

$$\begin{aligned} a_m &= \frac{1}{2} \sum_{k=1}^{m} \ln \left(\frac{\lambda_k}{\lambda_m} \right) = \frac{1}{2} \sum_{k=1}^{m} \ln \left(\frac{m^m}{k^k} \right) = \frac{1}{2} m^2 \ln m - \frac{1}{2} \sum_{k=1}^{m} k \ln k \\ &= \frac{1}{2} m^2 \ln m - \frac{m}{2} \sum_{k=1}^{m} \frac{k}{m} \ln \left(\frac{k}{m} \right) - \frac{\ln m}{2} \sum_{k=1}^{m} k \end{aligned}$$

It it obvious that

$$\frac{1}{m} \sum_{k=1}^{m} \frac{k}{m} \ln \left(\frac{k}{m} \right) \to \int_0^1 x \ln x dx, \quad \text{when } m \to +\infty.$$

Therefore we can write

$$\frac{1}{m} \sum_{k=1}^{m} \frac{k}{m} \ln \left(\frac{k}{m} \right) = \int_0^1 x \ln x dx + \xi_m,$$

where $\xi_m \to 0$ when $m \to +\infty$. Therefore

$$\begin{aligned} a_m &= \frac{1}{2}m^2 \ln m - \frac{m^2}{2}\int_0^1 x \ln x dx - \frac{m}{2}\xi_m - \frac{m(m+1)\ln m}{4} \\ &= \frac{1}{2}m^2 \ln m - \frac{m^2}{2}\left(\frac{x^2}{2}\ln x - \frac{x^2}{4}\right)\Big|_0^1 - \frac{m}{2}\xi_m - \frac{m(m+1)\ln m}{4} \\ &= \frac{m^2 \ln m}{4} + \frac{m^2}{8} - \frac{m}{2}\xi_m - \frac{m \ln m}{4}. \end{aligned}$$

Therefore

$$a_m \sim \frac{m^2 \ln m}{4}. \tag{6.1}$$

Then $\forall\, \epsilon_0 > 0$, $\exists\, m_0$, such as $\forall\, m \geq m_0$,

$$a_m \geq \frac{m^2 \ln m}{4(1+\varepsilon_0)}. \tag{6.2}$$

The previous inequality is needed for the main theorem proof. As a consequence of (6.1) and the definition of $m_n^\star$ it follows that

$$\frac{(m_n^\star)^2 \ln m_n^\star}{4} \sim \ln n. \tag{6.3}$$

Hence it exists ϵ_n such as $m_n^\star = 2\epsilon_n\sqrt{\ln n}$ for which (6.3) holds, therefore

$$\begin{aligned} \frac{(m_n^\star)^2 \ln m_n^\star}{4} &= \epsilon_n^2 \ln n \ \ln\left(2\epsilon_n\sqrt{\ln n}\right) \\ &= \ln n \left(\epsilon_n^2 \ln 2 + \epsilon_n^2 \ln \epsilon_n + \frac{\epsilon_n^2}{2}\ln\ln n\right). \end{aligned}$$

Clearly ϵ_n must be a decreasing sequence. Consequently

$$\frac{\epsilon_n^2}{2}\ln\ln n \sim 1.$$

Therefore

$$\epsilon_n \sim \sqrt{2}(\ln\ln n)^{-\frac{1}{2}},$$

and finally

$$m_n^\star \sim 2\sqrt{2}(\ln n)^{\frac{1}{2}}(\ln\ln n)^{-\frac{1}{2}}. \tag{6.4}$$

Let us define

$$R_n = \sum_{k \geq n+1} \lambda_k + n\lambda_n = \sum_{k \geq n+1} k^{-k} + n^{1-n}.$$

We can easily show that

$$R_n = O(n^{1-n}), \tag{6.5}$$

and

$$R_n = \Omega(n^{1-n}).$$

The proof is very simple. See Lemma 1 in Appendix C. The main theorem reads as follows

***Theorem* 6.1**
If X is a stochastic process with covariance function $\Gamma(t,s) = e^{-\theta|t-s|^2}$ then the eigenvalues associated to the process $(\lambda_k)_{k\geq 1}$ satisfies that $\ln \lambda_k \sim -k \ln k$ and

$$e_n(X) = O\left(\psi\left((2\sqrt{2})^{1-\delta}(\ln n)^{\frac{1}{2}}(\ln\ln n)^{-\frac{1}{2}}\right)^{-1}\right), \tag{6.6}$$

and

$$e_n(X) = \Omega\left(\psi\left((2\sqrt{2})^{1+\delta}(\ln n)^{\frac{1}{2}}(\ln\ln n)^{-\frac{1}{2}}\right)^{-1}\right). \tag{6.7}$$

where $\psi(n) = n^{n-1}$ for some $\delta > 0$.

Proof of Theorem 6.1

Let us begin proving (6.6). Using Lemma 4.3 in Luschgy and Pagès [2002] it follows

$$e_n^2(X) \leq c\left(\sum_{k\geq m_n^\star+1} \lambda_k + m_n^\star \lambda_{m_n^\star}\right),$$

then by (6.5) it is straightforward that

$$e_n^2(X) \;\leq\; cR_{m_n^\star} \leq c_1 \left(m_n^\star\right)^{1-m_n^\star}.$$

As a consequence of (6.4) $\forall\, \epsilon_1 > 0$, $\exists\, n_1$, such as $\forall\, n \geq n_1$ we have

$$m_n^\star \geq \frac{2\sqrt{2}}{1+\varepsilon_1}(\ln n)^{\frac{1}{2}}(\ln\ln n)^{-\frac{1}{2}}.$$

Hence by the decreasing behaviour of ψ^{-1} it follows that

$$\begin{aligned} e_n^2(X) &\leq c_1\psi\left(\frac{2\sqrt{2}}{1+\varepsilon_1}(\ln n)^{\frac{1}{2}}(\ln\ln n)^{-\frac{1}{2}}\right)^{-1} \\ &= c_1\psi\left((2\sqrt{2})^{1-\overline{\delta}}(\ln n)^{\frac{1}{2}}(\ln\ln n)^{-\frac{1}{2}}\right)^{-1}, \end{aligned}$$

where $\overline{\delta} = \frac{2\ln(1+\varepsilon_1)}{\ln 8}$.

That completes the proof for (6.6).

The proof of (6.7) is similar. We use Proposition 4.9 in Luschgy and Pagès [2002] instead Lemma 4.3. It is a simple matter that

$$\sum_{k\geq n+1} \lambda_k + n\lambda_{n+1} \geq \frac{n}{(n+1)^{n+1}} \geq \frac{1}{2}\frac{1}{(n+1)^n},$$

then $(n+1)^n = \Omega\left(\left(\sum_{k\geq n+1}\lambda_k + n\lambda_{n+1}\right)^{-1}\right)$. We need to find a function ϕ^* for which

$$\phi^*(a_n) = \Omega\left(\left(\sum_{k\geq n+1}\lambda_k + n\lambda_{n+1}\right)^{-1}\right),$$

or $\phi^*(a_n) = \Omega\left((n+1)^n\right)$ in order to use Proposition 4.9. Let us define

$$\phi^*(n) = (g(n))^{g(n)-1},$$

where

$$g(n) = C^* \left(\frac{n}{\ln n}\right)^{\frac{1}{2}}.$$

By (6.2) and the increasing behaviour of g we get

$$\begin{aligned} g(a_n) &\geq g\left(\frac{1}{4(1+\varepsilon_0)} n^2 \ln n\right) = C^* \left(\frac{(4(1+\varepsilon_0))^{-1} n^2 \ln n}{2\ln n + \ln\ln n - \ln 4(1+\varepsilon_0)}\right)^{\frac{1}{2}} \\ &= \frac{C^*}{2\sqrt{1+\varepsilon_0}} n \left(2 + \frac{\ln\ln n}{\ln n} - \frac{\ln 4(1+\varepsilon_0)}{\ln n}\right)^{-\frac{1}{2}} \\ &\geq \frac{C^*}{2\sqrt{2}\sqrt{1+\varepsilon_0}(1+\varepsilon_2)} n = \frac{C^*}{2\sqrt{2}(1+\varepsilon_3)} n, \qquad \forall\, n \geq n_2. \end{aligned}$$

Taking $C^* > 2\sqrt{2}(1+\varepsilon_3)$ it is an immediate consequence that

$$g(a_n) \geq n+1.$$

Therefore

$$\phi^*(a_n) = \Omega\left((n+1)^n\right).$$

Consequently it follows

$$\phi^*(a_n) \quad = \Omega\left(\textstyle\sum_{k\geq n+1} \lambda_k + n\lambda_{n+1}\right)^{-1}.$$

The function ϕ^* is rapidly varying of index ∞. Therefore, applying Proposition 4.9 of Luschgy and Pagès [2002] it follows for $\varepsilon_4 > \varepsilon_3 > 0$ that

$$\begin{aligned} e_n(X) &= \Omega\left(\phi^*\left(\ln n\right)^{-1/2}\right) = \Omega\left(\psi\left(2\sqrt{2}(1+\varepsilon_4)(\ln n)^{\frac{1}{2}}(\ln\ln n)^{-\frac{1}{2}}\right)^{-1/2}\right) \\ &= \Omega\left(\psi\left((2\sqrt{2})^{1+\underline{\delta}}(\ln n)^{\frac{1}{2}}(\ln\ln n)^{-\frac{1}{2}}\right)^{-1/2}\right), \end{aligned}$$

where $\underline{\delta} = \frac{2\ln(1+\varepsilon_4)}{\ln 8}$.

Taking $n^* = \max\{n_0, n_1, n_2\}$ we can choose $\varepsilon_0, \varepsilon_1$ and ε_2 conveniently such as

$$\delta = \overline{\delta} = \underline{\delta} \qquad \forall\, n \geq n^*,$$

and the proof is complete.

□

***Remark* 6.1**

It is important to emphasize that we do not have exactly the same order on both limits for the quantization error. However we can take δ small enough. Yet, it is not clear that the result holds when $\delta \to 0$.

6.1.3 Kriging. Two dimensions

Kriging is a common method used in the analysis of spatial data, it is a procedure for constructing a minimum error variance estimate at a location where the true value is unknown. It is an optimal geostatistical interpolation technique that considers both the distance and the degree of variation between known data points when estimating values in unknown areas. It can also be viewed as a nonparametric regression method against observed values of surrounding data points, weighted according to spatial covariance values for estimating curves and surfaces. It is typically derived as a best linear unbiased estimator.

Kriging is a geostatistical method based on statistical models that include the statistical relationship among the measured points. Applied properly, kriging allows to derive weights that result in optimal and unbiased estimates. It attempts to minimize the error variance and set the mean of the prediction errors to zero. It have the capability of producing a prediction surface and provide some measure of the accuracy of the predictions.

Formally, the aim of kriging is to estimate the value of some surface or curve Z on a point x_0 using the information of the known data and it is described as follows

$$\widehat{Z}(x_0) = \mu(x_0) + \sum_{x_i \in V(x_0)} w_i(x_0) Z(x_i),$$

where

- $V(x_0)$ is a neighborhood of x_0. (The size of $V(x_0)$ depends on the nature of the problem).
- The x_i for $i = 1, 2, \cdots, |V(x_0)|$, are the known data "around" x_0.
- $\mu(x_0)$ is the trend of the model.
- The w_i are the weights associated to $Z(x_i)$ and are derived from the covariance function of the data. In other words the weights are chosen or estimated such that the kriging variance or estimation variance

$$\begin{aligned} \sigma_e^2 &= V(Z(x_0)) + \sum_{\substack{x_i \in V(x_0) \\ x_j \in V(x_0)}} w_i(x_0) w_j(x_0) \mathrm{Cov}(Z(x_i), Z(x_j)) \\ &\quad -2 \sum_{x_i \in V(x_0)} w_i(x_0) \mathrm{Cov}(Z(x_0), Z(x_i)), \end{aligned}$$

  is minimized.

The classical types or kriging are

Simple kriging: Assume that the trend component $\mu(x_0) = m$ is a known constant.

Ordinary kriging: Assume that the trend component $\mu(x_0) = \mu$ is unknown. It is the most commonly used type of kriging. It is highly reliable and is recommended for most data sets constant.

Universal kriging: Assume that the trend component satisfies

$$\mu(x_0) = \sum_{x_i \in V(x_0)} \beta_i m_i(x_0).$$

For a deeper discussion and a fuller treatment on kriging theory we refer the reader to the work of Cressie [1993] and references therein.

Curve estimation. Squared covariance function

We wish to integrate the quantization theory into the field of kriging. It is not our purpose to give a different estimation for $\widehat{Z}(x_0)$. We intend to introduce a method based on optimal quantizers of Gaussian process as an additional tool for the curve or surface estimation.

In that direction we work with the squared covariance function $\Gamma(t) = e^{-\theta t^2}$ with $\theta = 20$ and $\theta = 100$.

First we find for each θ the 5000-optimal quantizer Q_{5000} for each value of θ. The calculations was made using the approximated eigenvalues of the process and the classical method based on product quantizers. We generate a path of the process $X(\mathbf{t})$ in some points $\mathbf{t} = (t_1, \cdots, t_n)$ (we take $n = 100$). Using a random sample of the path $X(\mathbf{t_N})$, where $\mathbf{t_N} = \{t_{i_1}, \cdots, t_{i_N},\ i_k \in \{1, 2, \cdots, n\}, \forall\, k\}$ we construct the kriging estimate of the process $\widehat{X}_N(\mathbf{t})$. Finally we find the best two components of Q_{5000} near of $X(\mathbf{t_N})$ and $\widehat{X}_N(\mathbf{t})$ (in some cases are the same).

In Figures 6.1 and 6.2 we see the behaviour of the nearest quantizer of the process to the points and to the kriging estimate. In each case the results are not good. The quantizers in some of the cases capture well the curve, but in others not so well.

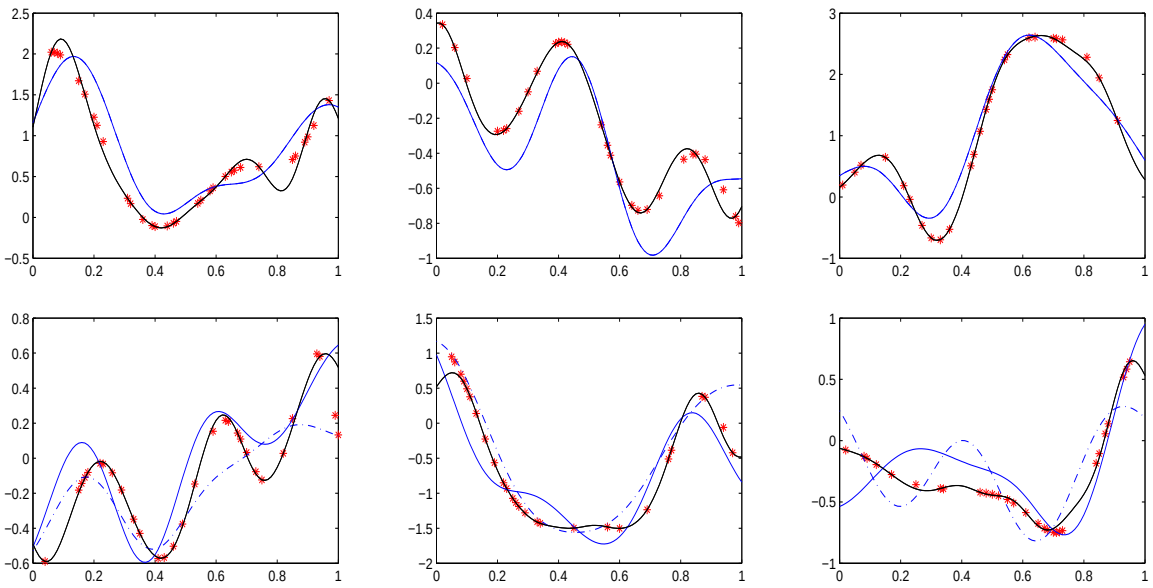

Figure 6.1: $\theta = 20$ with $*$ for $X(\mathbf{t_{30}})$, black $-$ for $\widehat{X}_{30}(\mathbf{t})$, blue $-$ for $Q_{5000}(X(\mathbf{t_{30}}))$, blue $\cdot-$ for $Q_{5000}(\widehat{X}_{30}(\mathbf{t}))$.

The quantization theory do no work properly on kriging setting. Perhaps another approach can be used to give some useful information. A larger number of quantizers were used to obtain better results, however in the light of the previous figures we do not think that a better improvement can be achieved. Moreover if we think at the calculation time needed to construct the optimal quantizers. It seems that it strongly depends on the choice of the points $\mathbf{t_N}$, see Figure 6.3.

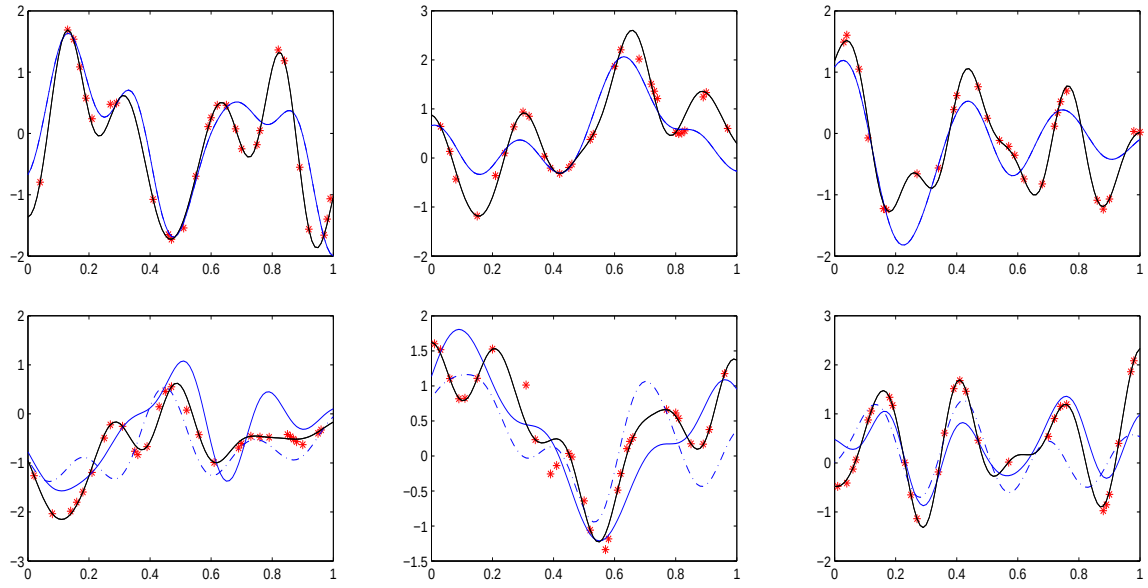

Figure 6.2: $\theta = 100$ with $*$ for $X(\mathbf{t_{30}})$, black $-$ for $\widehat{X}_{30}(\mathbf{t})$, blue $-$ for $Q_{5000}(X(\mathbf{t_{30}}))$, blue $\cdot-$ for $Q_{5000}(\widehat{X}_{30}(\mathbf{t}))$.

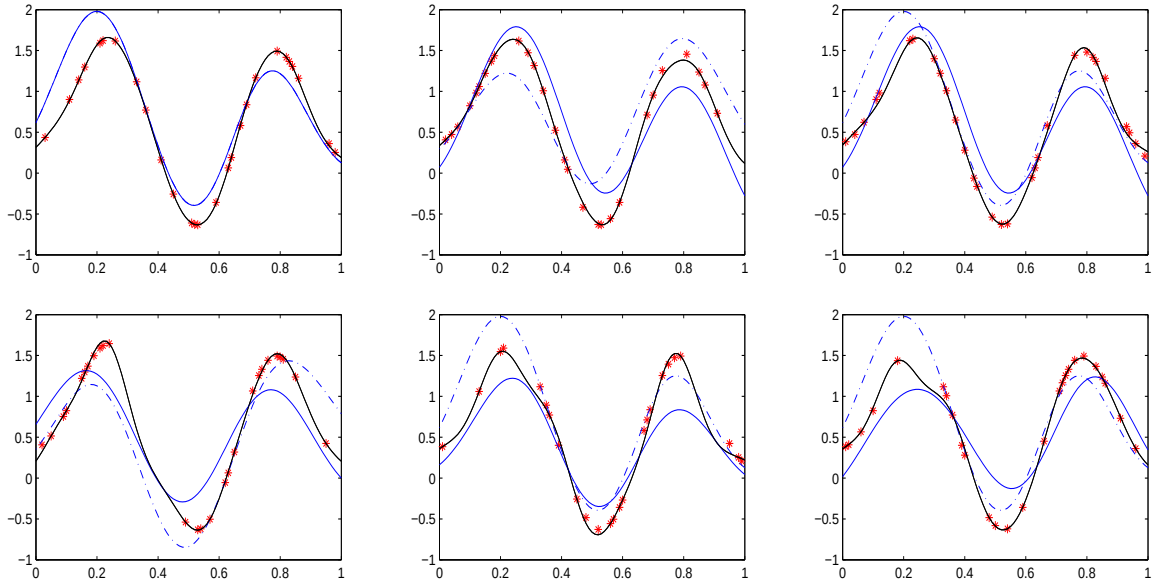

Figure 6.3: The same process with $\theta = 20$ and different choices of $\mathbf{t_{30}}$ with $*$ for $X(\mathbf{t_{30}})$, black $-$ for $\widehat{X}_{30}(\mathbf{t})$, blue $-$ for $Q_{5000}(X(\mathbf{t_{30}}))$, blue $\cdot-$ for $Q_{5000}(\widehat{X}_{30}(\mathbf{t}))$.

6.2 Some asymptotics for the maximal radius of the Wiener process

We present here some asymptotic results for the maximal radius of the Wiener process. We follow the work presented in the PhD thesis of Sagna [2008]. The author provides some upper and lower estimate for the maximal radius in the finite dimensional framework. The extension to infinite dimension carried out some extra difficulties because we do not have some of the good properties of the optimal quantizers in $\mathbb{R}^d$. Our aim is to obtain similar results for the maximal radius of the optimal sequence $(\alpha_n)_{n\geq 1}$ of n-quantizers for the Wiener process. The maximal radius sequence is defined for $n \geq 1$ by

$$\rho_n = \rho(\alpha_n) = \max\{\|a\|_{\mathcal{L}^2}, a \in \alpha_n\}.$$

This section is intended as an attempt to approach us to the correct asymptotics of the maximal radius of the Wiener process. In order to achieve our goal we present four different ways that give us a general idea of the general asymptotics. We begin with a initial approach based on the arguments used by Sagna in $\mathbb{R}^d$. The second approach is related with the search of some specific kind or set of quantizers which allow us to conclude some asymptotic for the maximal radius of the Wiener process. The third one is concerned on some stationary quantizers, more specifically a variation on the quantization design proposed by Luschgy et al. [2010]. The fourth approach is based on the study of the maximal radius of the optimal quantizers for the linear approximation of the Brownian motion and finally we conclude with some ideas for the Brownian motion.

6.2.1 First approach for the Brownian motion

Our first approach is based on the method used by Sagna in his PhD thesis. However the extension of his methodology do not work properly in the infinite dimensional case. Due to the behaviour of the Wiener measure we can not state a lower bound for $e_n^2(W) - e_{n+1}^2(W)$ which is the first step to find the upper limit for ρ_n. For the lower limit, using his argument we can only achieve the rate $\log\log n$.

The accuracy of $\log\log n$ for the limit inferior of the maximal radius does not seem to be the optimal one. We numerically construct the sequence $(\rho_n)_{n\geq 1}$ until $n = 10160$ using the optimal n-quantizers for the Wiener process calculated by Pagès. Using the database given in `http://www.quantize.maths-fi.com/Gaussian_process_database` we computed the sequence $(\rho_n)_{n\geq 1}$. We think that the correct asymptotic for the maximal radius of the Wiener process is $\log n$. In figure 6.4 we can see the behaviour of ρ_n^2 with $\log n$ and $\sqrt{2}\pi^{-2}\log\log n$ (this last one comes from the first approach of the lower limit, see Theorem 6.3).

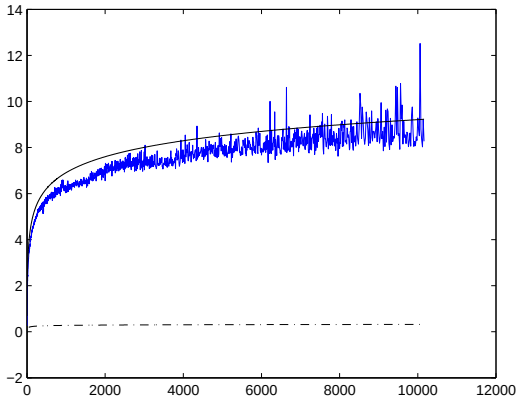

Figure 6.4: Graphic of ρ_n^2 with $\ln n$(black $-$) and $\frac{2}{\pi^2}\ln\ln n$(black $-.$).

It is easy to see that the behaviour of ρ_n^2 is closer to $\log n$ instead $\frac{2}{\pi^2}\log\log n$. In Figure 6.5 we plot the maximal radius versus $\ln n$ and $\ln\ln n$ respectively and we confirm that the behaviour of ρ_n^2 is similar to $\ln n$, then it seems that the correct order of the maximal radius is $\ln n$.

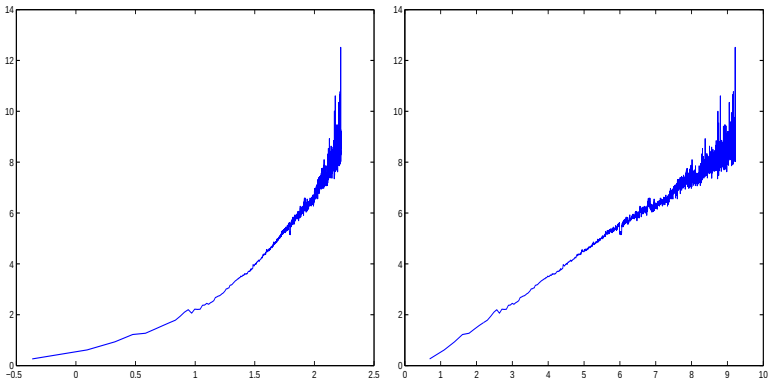

Figure 6.5: Left: graphic of ρ_n^2 versus $\ln\ln n$. Right: graphic of ρ_n^2 versus $\ln n$.

Lower bound for the maximal radius of the Brownian motion

We can state the following result for the lower bound of $(\rho_n)_{n\geq 1}$.

***Theorem* 6.2**
Let W be a Wiener process with probability measure $\mathbb{P}$ in $\mathcal{L}^2$. Let us assume that $(\alpha_n)_{n\geq 1}$ is an optimal sequence of n-quantizers. Let $(W_k)_{k\geq 1}$ be an i.i.d sequence of H-valued random variables with probability distribution $\mathbb{P}$. Then

$$\liminf_{n\to+\infty}\left(\rho_n - \mathbb{E}\left(\max_{k\leq[(\ln n)^{1/2}]}\|W_k\|\right)\right) \geq -\frac{\sqrt{2}}{\pi}.$$

Proof of Theorem 6.2

For all $k \leq N$ it follows that

$$\|W_k\| \leq \|W_k - \widehat{W}_k^{\alpha_n}\| + \rho_n,$$

where $\widehat{W}_k^{\alpha_n} = \sum_{a \in \alpha_n} a \mathbf{1}_{W_k \in C_a(\alpha_n)}$.

Taking the maximum in the previous equation we have

$$\max_{1 \leq k \leq N} \|W_k\| \leq \max_{1 \leq k \leq N} \|W_k - \widehat{W}_k^{\alpha_n}\| + \rho_n \leq \sum_{k=1}^{N} \|W_k - \widehat{W}_k^{\alpha_n}\| + \rho_n.$$

Taking expected value in the previous equation, it follows that

$$\begin{aligned} \mathbb{E} \max_{1 \leq k \leq N} \|W_k\| &\leq N\, \mathbb{E}\|W_1 - \widehat{W}_1^{\alpha_n}\| + \rho_n \leq N \left(\mathbb{E}\|W_1 - \widehat{W}_1^{\alpha_n}\|^2\right)^{1/2} + \rho_n \\ &= N\, e_n(W) + \rho_n. \end{aligned}$$

Taking $N = [(\ln n)^{1/2}]$ it follows that

$$\liminf_{n \to +\infty} \left(\rho_n - \mathbb{E}\left(\max_{k \leq [(\ln n)^{1/2}]} \|W_k\|\right)\right) \geq -\lim_{n \to +\infty} (\ln n)^{\frac{1}{2}} e_n(W).$$

Obviously the right side of the last inequality is finite and the proof is complete. The constant $\sqrt{2}/\pi$ comes from the asymptotic rate of Brownian motion (see for instance Luschgy and Pagès [2004a]).

□

The lower bound for $(\rho_n)_{n \geq 1}$ is provided by the following theorem.

***Theorem* 6.3**

Let W be a Wiener process with probability measure $\mathbb{P}$ in $\mathcal{L}^2([0,1], dt)$. Let us assume that $(\alpha_n)_{n \geq 1}$ is an optimal sequence of n-quantizers and $(\rho_n)_{n \geq 1}$ the optimal sequence of maximal radius. Then

$$\liminf_{n \to \infty} \frac{\rho_n}{\sqrt{\ln \ln n}} \geq \sqrt{\frac{\lambda_1}{2}} = \frac{\sqrt{2}}{\pi}$$

where $\lambda_1 = 4/pi^2$ is the greatest eigenvalue of the covariance operator of the process.

Proof

We have that

$$\begin{aligned} e_n^2(W) &= \mathbb{E} \min_{a \in \alpha_n} \|W - a\|^2 \mathbf{1}_{\{\|W\| < 2\rho_n\}} + \mathbb{E} \min_{a \in \alpha_n} \|W - a\|^2 \mathbf{1}_{\{\|W\| \geq 2\rho_n\}} \\ &\geq \mathbb{E} \min_{a \in \alpha_n} \|W - a\|^2 \mathbf{1}_{\{\|W\| \geq 2\rho_n\}} \\ &\geq \rho_n^2 P(\|W\| \geq 2\rho_n). \end{aligned}$$

It is a simple matter that

$$\begin{aligned} P(\|W\|^2 \geq 4\rho_n^2) &= P\left(\sum_{i\geq 1} \lambda_i \xi_i^2 \geq 4\rho_n^2\right) \\ &\geq P(\lambda_1 \xi_1^2 \geq 4\rho_n^2) = P\left(|\xi_1| \geq \frac{2\rho_n}{\sqrt{\lambda_1}}\right). \end{aligned}$$

It is know that $\xi_1 \sim N(0,1)$ therefore the distribution of $|\xi_1|$ it is know. Hence using

$$\int_x^{+\infty} e^{-u^2/2} du \geq \frac{x}{x^2+1} e^{-x^2/2} \qquad \forall\, x \geq 0,$$

it follows that

$$\begin{aligned} e_n^2(W) &\geq \rho_n^2 P(\|W\|^2 \geq 4\rho_n^2) \geq \rho_n^2 P\left(|\xi_1| \geq \frac{2\rho_n}{\sqrt{\lambda_1}}\right) \\ &\geq \sqrt{\frac{8\lambda_1}{\pi}} \frac{\rho_n^3}{4\rho_n^2 + \lambda_1} \exp\{-4\rho_n^2/2\lambda_1\}. \end{aligned}$$

Then

$$\begin{aligned} \ln e_n^2(W) &\geq 3\ln \rho_n + \ln\left(\frac{8\lambda_1}{\pi}\right) - \ln(4\rho_n^2 + \lambda_1) - 2\rho_n^2/\lambda_1 \\ 2\rho_n^2/\lambda_1 &\geq \ln\left(\frac{1}{e_n^2(W)}\right) + 3\ln\rho_n + \ln\left(\frac{8\lambda_1}{\pi}\right) - \ln(4\rho_n^2 + \lambda_1) \\ &\geq \ln\left(\frac{1}{e_n^2(W)}\right). \end{aligned}$$

By the asymptotic behaviour of $e_n(W) \sim \frac{\sqrt{2}}{\pi}(\ln n)^{-1/2}$ it follows that

$$\liminf_{n\to\infty} \frac{\rho_n}{\sqrt{\ln\ln n}} \geq \sqrt{\frac{\lambda_1}{2}} = \frac{\sqrt{2}}{\pi}.$$

□

6.2.2 Quasi optimal quantizers

Our purpose is to create a class of quantizers that allow us to obtain some results for the maximal radius of the Wiener process. We can construct a set A_n^ε formed for a class of n-quantizers included the optimal ones. In that case the general properties obtained for A_n^ε could be also applied to the optimal n-quantizer. In that direction let us introduce the notion of n-optimal ε-quantizer.

***Definition* 6.1**
The set α_n^ε a sequence of quantizers it is an asymptotic n-optimal ε-quantizer for the stochastic process X if it exists $\varepsilon > 0$ such that for all $n \geq n_0 (n_0 \in \mathbb{N})$,

$$e_n^2(X, \alpha_n^\varepsilon) \leq (1+\varepsilon) e_n^2(X),$$

where $e_n^2(X)$ is the optimal quantization error for X associated to the n-optimal quantizer α for X.

***Remark* 6.2**
It is important to point out that a sequence of optimal n-quantizer is also an asymptotic n-optimal ε-quantizer because

$$e_n^2(X) < (1+\varepsilon)e_n^2(X), \qquad \varepsilon > 0.$$

The next step was to create one specific n-optimal ε-quantizer which allows us to conclude some asymptotics for the maximal radius of the Brownian motion. We present the following lemma that applies for any self-similar process with stationary increments. According to that lemma we construct two new quantizers.

***Lemma* 6.1**
If $X_{[0,1/k]}$ is a self similar process in $[0,1/k]$ then for the optimal n-quantizer $\beta^{(k)}$ it follows that

$$e_n^2(X_{[0,1/k]}) = \frac{1}{k^{2H+1}} e_n^2(X_{[0,1]}) \quad \textit{and} \quad \beta^{(k)} = k^{-H}\,\beta,$$

where $H \in (0,1)$ is the self similar index and β is the optimal n-quantizer for $X_{[0,1]}$.

Proof

Let us write

$$\begin{aligned}
e_n^2(X_{[0,1/k]}) &= \sum_{i=1}^{n} \mathbb{E}\left(\mathbf{1}_{C_i}\|X_{[0,1/k]} - \beta^{(k)}\|^2\right) \\
&= \sum_{i=1}^{n} \mathbb{E}\left(\mathbf{1}_{C_i} \int\limits_{[0,1/k]} (X_{[0,1/k]}(t) - \beta_i^{(k)}(t))^2 dt\right) \\
&= \frac{1}{k}\sum_{i=1}^{n} \mathbb{E}\left(\mathbf{1}_{C_i} \int\limits_{[0,1]} (X_{[0,1/k]}(s/k) - \beta_i^{(k)}(s/k))^2 ds\right) \\
&= \frac{1}{k^{2H+1}}\sum_{i=1}^{n} \mathbb{E}\left(\mathbf{1}_{C_i} \int\limits_{[0,1]} (X_{[0,1]}(s) - k^H\beta_i^{(k)}(s/k))^2 ds\right) \\
&= \frac{1}{k^{2H+1}} e_n^2(X_{[0,1]}),
\end{aligned}$$

and that complete the proof.

□

From now on we restrict our attention on the Brownian motion. The main idea is to construct a suitable quantizer by dividing the time interval $[0,1]$. Set $k = \frac{\ln n}{\ln m} \in \mathbb{R}_+$ with $m, n \in \mathbb{N}$ ($m^k = n$) and $m < n$. We work with two concatenated m^k-quantizers which of course may be not optimal:

1. We consider a continuous n-quantizer for the Brownian motion in $[0,1]$ made of the concatenation of the optimal m-quantizer of the process on each time interval $\left[\frac{i-1}{k}, \frac{i}{k}\right]$. We write this quantizer as $\alpha^{(1)}_{\otimes k} = (a^{(1)}_1 \cdots, a^{(1)}_n)$.
2. We consider a non continuous n-quantizer for the Brownian motion in $[0,1]$ using the optimal m-quantizer for the Brownian motion in $[0,1/k]$ and the optimal $n^{\frac{i-1}{k}}$-quantizer for the $N(0, \frac{i-1}{k})$ for $i = 2, \cdots, k$ on each time interval $\left[\frac{i-1}{k}, \frac{i}{k}\right]$. We will write this quantizer as $\alpha^{(2)}_{\otimes k} = (a^{(2)}_1 \cdots, a^{(2)}_n)$.

Obviously in the same way we can construct a wide class of these concatenated quantizers. In the sequel we leave the index on the optimal quantizer when its size is different to n. Let us denote by

- $\alpha_n = (a_1 \cdots, a_n)$ the optimal n-quantizer for a Brownian motion W in $[0,1]$.
- $\alpha^{(k)}_n = (a^{(k)}_1 \cdots, a^{(k)}_n)$ the optimal n-quantizer for a Brownian motion in $[0,1/k]$ named $W_{[0,1/k]}$.
- $\beta^{(i)}_n = (b^i_1 \cdots, b^i_{m^{(i)}})$ the optimal $m^{(i)}$-quantizer of a normal distribution $N(0, \frac{i-1}{k})$ for $i = 2, \cdots, k$ where $m^{(i)} = n^{\frac{i-1}{k}}$ and C^i_r is the Voronoï tessel for b^i_r for $r = 1, \cdots, m^{(i)}$.
- C_r and $C^{(k)}_r$ are the Voronoï tessel for a_r and $a^{(k)}_r$ respectively in $\mathcal{L}^2([0,1])$ and $\mathcal{L}^2([0,1/k])$ for $r = 1, \cdots, n$.
- $C^{(1)}_j$ and $C^{(2)}_j$ are the Voronoï tessel for $a^{(1)}_j$ and $a^{(2)}_j$ respectively in $\mathcal{L}^2([0,1])$ for $j = 1, \cdots, n$.

Therefore for $a^{(1)}_j \in \alpha^{(1)}_{\otimes k}$ $(j = 1, \cdots, n)$ we write

$$a^{(1)}_j = a^{(k)}_{l^1_j}(t)\mathbf{1}_{[0,\frac{1}{k}]}(t) + \sum_{i=2}^{k} \left(a^{(k)}_{l^i_j}\left(t - \frac{i-1}{k}\right) + (i-1)a^{(k)}_{l^{i-1}_j}\left(\frac{1}{k}\right)\right) \mathbf{1}_{]\frac{i-1}{k}, \frac{i}{k}]}(t), \quad (6.8)$$

where $l^i_j \in \{1, \cdots, m\}$ for all $j = 1, \cdots, n$ and $i = 1, \cdots, k$. In the same way for $a^{(2)}_j \in \alpha^{(2)}_{\otimes k}$ $(j = 1, \cdots, n)$ we write

$$a^{(2)}_j = a^{(k)}_{l^1_j}(t)\mathbf{1}_{[0,\frac{1}{k}]}(t) + \sum_{i=2}^{k} \left(a^{(k)}_{l^i_j}\left(t - \frac{i-1}{k}\right) + b^i_{j_i}\right) \mathbf{1}_{]\frac{i-1}{k}, \frac{i}{k}]}(t), \quad (6.9)$$

where $l^i_j \in \{1, \cdots, m\}$ and $j_i \in \{1, \cdots, m^{(i)}\}$ for all $j = 1, \cdots, n$ and $i = 1, \cdots, k$.

Technical results

We will show that $\alpha^{(2)}_{\otimes k}$ is a n-optimal ε-quantizer. For $\alpha^{(1)}_{\otimes k}$ we do not have a demonstration in that sense but we think that it has a similar behaviour.

For the second concatenated quantizer $\alpha^{(2)}_{\otimes k}$ it holds

$$e^2_n(W, \alpha^{(2)}_{\otimes k}) = \mathbb{E} \min_{a^{(2)}_j \in \alpha^{(2)}_{\otimes k}} \|W - a^{(2)}_j\|^2_{\mathcal{L}^2} = \mathbb{E}\left(\sum_{j=1}^{n} \mathbf{1}_{C^{(2)}_j}(W) ||W - a^{(2)}_j||^2_{\mathcal{L}^2} \right),$$

and if we set $W_{(i)} = W(\frac{i-1}{k}) \sim N(0, \frac{i-1}{k})$ and $\mathcal{L}^2(\frac{i}{k}) = \mathcal{L}^2([\frac{i-1}{k}, \frac{i}{k}])$ it follows that

$$\left\|W - a_j^{(2)}\right\|^2_{\mathcal{L}^2} = \left\|W - a_{l_j^1}^{(k)}\right\|^2_{\mathcal{L}^2(\frac{1}{k})} + \sum_{i=2}^{k} \left\|W - a_{l_j^i}^{(k)}(t-(i-1)/k) - b_{j_i}^i\right\|^2_{\mathcal{L}^2(\frac{i}{k})},$$

and

$$\begin{aligned}
\left\|W - a_{l_j^i}^{(k)}(t-(i-1)/k) - b_{j_i}^i\right\|^2_{\mathcal{L}^2(\frac{i}{k})} &= \int_{\frac{i-1}{k}}^{\frac{i}{k}} \left(W(t) - W_{(i)} + W_{(i)} - a_{l_j^i}^{(k)}(t-(i-1)/k) - b_{j_i}^i\right)^2 dt \\
&= \left\|W - W_{(i)} - a_{l_j^i}^{(k)}(t-(i-1)/k)\right\|^2_{\mathcal{L}^2(\frac{i}{k})} + \frac{\left(W_{(i)} - b_{j_i}^i\right)^2}{k} \\
&\quad + 2\int_{\frac{i-1}{k}}^{\frac{i}{k}} \left(W(t) - W_{(i)} - a_{l_j^i}^{(k)}(t-(i-1)/k)\right)\left(W_{(i)} - b_{j_i}^i\right) dt.
\end{aligned}$$

Hence

$$\begin{aligned}
e_n^2(W, \alpha_{\otimes k}^{(2)}) &= \mathbb{E} \min_{a_j^{(2)} \in \alpha_{\otimes k}^{(2)}} \|W - a_j^{(2)}\|^2_{\mathcal{L}^2} = \mathbb{E}\left(\sum_{j=1}^{n} \mathbf{1}_{C_j^{(2)}}(W) \|W - a_j^{(2)}\|^2_{\mathcal{L}^2}\right) \\
&= \mathbb{E}\left(\sum_{j=1}^{n} \mathbf{1}_{C_j^{(2)}}(W) \|W - a_{l_j^1}^{(k)}\|^2_{\mathcal{L}^2(\frac{1}{k})}\right) + \frac{1}{k}\sum_{i=2}^{k} \mathbb{E}\left(\sum_{j=1}^{n} \mathbf{1}_{C_j^{(2)}}(W)(W_{(i)} - b_{j_i}^i)^2\right) \\
&\quad + \sum_{i=2}^{k} \mathbb{E}\left(\sum_{j=1}^{n} \mathbf{1}_{C_j^{(2)}}(W) \|W - W_{(i)} - a_{l_j^i}^{(k)}(t-(i-1)/k)\|^2_{\mathcal{L}^2(\frac{i}{k})}\right) \\
&\quad + 2\sum_{i=2}^{k} \mathbb{E}\left(\sum_{j=1}^{n} \mathbf{1}_{C_j^{(2)}}(W) \int_{\frac{i-1}{k}}^{\frac{i}{k}} \left(W(t) - W_{(i)} - a_{l_j^i}^{(k)}(t-(i-1)/k)\right)\left(W_{(i)} - b_{j_i}^i\right) dt\right)
\end{aligned}$$

The process $W(t) - W_{(i)}$ in $[\frac{i-1}{k}, \frac{i}{k}]$ has the same law that a Brownian motion in $[0, \frac{1}{k}]$, thus

$$\begin{aligned}
k\mathbb{E}\left(\sum_{j=1}^{n} \mathbf{1}_{C_j^{(2)}}(W) \|W - a_{l_j^1}^{(k)}\|^2_{\mathcal{L}^2(\frac{1}{k})}\right) &= \mathbb{E}\left(\sum_{j=1}^{n} \mathbf{1}_{C_j^{(2)}}(W) \|W - a_{l_j^1}^{(k)}\|^2_{\mathcal{L}^2(\frac{1}{k})}\right) \\
&\quad + \sum_{i=2}^{k} \mathbb{E}\left(\sum_{j=1}^{n} \mathbf{1}_{C_j^{(2)}}(W) \|W - W_{(i)} - a_{l_j^i}^{(k)}(t-(i-1)/k)\|^2_{\mathcal{L}^2(\frac{i}{k})}\right)
\end{aligned}$$

For each t in $[\frac{i-1}{k}, \frac{i}{k}]$ we also know that $\left(W(t) - W_{(i)} - a_{l_j^i}^{(k)}(t-(i-1)/k)\right)$ is independent of $\left(W_{(i)} - b_{j_i}^i\right)$. Therefore in the Voronoï tessel $C_{l_j^i}^{(k)}$ of $a_{l_j^i}^{(k)}$ we have that

$$\mathbb{E}\left(\mathbf{1}_{C_{l_j^i}^{(k)}}\left(W(t) - W_{(i)} - a_{l_j^i}^{(k)}(t-(i-1)/k)\right)\right) = 0.$$

Therefore

$$\begin{aligned}
e_n^2(W,\alpha_{\otimes k}^{(2)}) &= \mathbb{E}\min_{a_j^{(2)}\in\alpha_{\otimes k}^{(2)}}\|W-a_j^{(2)}\|_{\mathcal{L}^2}^2 = \mathbb{E}\left(\sum_{j=1}^n \mathbf{1}_{C_j^{(2)}}(W)\|W-a_j^{(2)}\|_{\mathcal{L}^2}^2\right)\\
&= k\mathbb{E}\left(\sum_{j=1}^n \mathbf{1}_{C_j^{(2)}}(W)\|W-a_{l_j^i}^{(k)}\|_{\mathcal{L}^2(\frac{1}{k})}^2\right)+\frac{1}{k}\sum_{i=2}^k\mathbb{E}\left(\sum_{j=1}^n \mathbf{1}_{C_j^{(2)}}(W)(W_{(i)}-b_{j_i}^i)^2\right)\\
&\quad +2\sum_{i=2}^k\mathbb{E}\left(\sum_{j=1}^n \mathbf{1}_{C_j^{(2)}}(W)\int_{\frac{i-1}{k}}^{\frac{i}{k}}\left(W(t)-W_{(i)}-a_{l_j^i}^{(k)}\left(t-(i-1)/k\right)\right)\left(W_{(i)}-b_{j_i}^i\right)dt\right)\\
&\leq k\mathbb{E}\left(\sum_{j=1}^n \mathbf{1}_{C_{l_j^1}^{(k)}}(W)\|W-a_{l_j^1}^{(k)}\|_{\mathcal{L}^2(\frac{1}{k})}^2\right)+\frac{1}{k}\sum_{i=2}^k\mathbb{E}\left(\sum_{j=1}^n \mathbf{1}_{C_{j_i}^i}(W)(W_{(i)}-b_{j_i}^i)^2\right)\\
&\quad +2\sum_{i=2}^k\mathbb{E}\left(\sum_{j=1}^n \mathbf{1}_{C_{l_j^i}^{(k)}}(W)\int_{\frac{i-1}{k}}^{\frac{i}{k}}\left(W(t)-W_{(i)}-a_{l_j^i}^{(k)}\left(t-(i-1)/k\right)\right)\left(W_{(i)}-b_{j_i}^i\right)dt\right)\\
&= ke_m^2(W,\alpha_m^{(k)})+\frac{1}{k}\sum_{i=2}^k\frac{i-1}{k}e_{m^{(i)}}^2(Z_i)
\end{aligned}$$

where $Z_i\sim N(0,1)$. By the asymptotic behaviour of $e_{m^{(i)}}^2(Z_i)$ it follows that

$$\begin{aligned}
e_n^2(W,\alpha_{\otimes k}^{(2)}) &= \mathbb{E}\min_{a_j^{(2)}\in\alpha_{\otimes k}^{(2)}}\|W-a_j^{(2)}\|_{\mathcal{L}^2}^2\\
&\leq ke_m^2(W,\alpha_m^{(k)})+C_2\frac{1}{k}\sum_{i=2}^k\frac{i-1}{k}\left(\frac{1}{n^{\frac{i-1}{k}}}\right)^2.
\end{aligned}$$

The final term of the previous equation can be rewritten as

$$\frac{1}{k}\sum_{i=1}^{k-1}\frac{i}{k}\left(\frac{1}{n}\right)^{\frac{2i}{k}},$$

and

$$\frac{1}{k}\sum_{i=1}^{k-1}\frac{i}{k}\left(\frac{1}{n}\right)^{\frac{2i}{k}}\leq\frac{1}{k}\sum_{i=1}^{k}\frac{i}{k}\left(\frac{1}{n}\right)^{\frac{2i}{k}}\overset{k\to+\infty}{\longrightarrow}\int_0^1 u\left(\frac{1}{n}\right)^{2u}du=\int_0^1 ue^{-2u\ln n}du,$$

where

$$\begin{aligned}
\int_0^1 ue^{-2u\ln n}du &= -\frac{ue^{-2u\ln n}}{2\ln n}\Big|_0^1+\int_0^1\frac{e^{-2u\ln n}}{2\ln n}du\\
&= -\frac{1}{2n^2\ln n}-\frac{e^{-2u\ln n}}{4(\ln n)^2}\Big|_0^1=-\frac{1}{2n^2\ln n}+\frac{1}{4(\ln n)^2}-\frac{1}{4(n\ln n)^2}.
\end{aligned}$$

We also know by Lemma 6.1 that

$$e_m^2(W,\alpha_m^{(k)})=\frac{1}{k^2}e_m^2(W,\alpha_m)=\frac{1}{k^2}e_m^2(W).$$

Therefore it follows

$$\begin{aligned} e_n^2(W,\alpha_{\otimes k}^{(2)}) &\leq ke_m^2(W,\alpha_m^{(k)}) + C_2\frac{1}{k}\sum_{i=2}^{k}\frac{i-1}{k}\left(\frac{1}{n^{\frac{i-1}{k}}}\right)^2 \\ &\leq \frac{1}{k}e_m^2(W) + C_2'\frac{n^2-2\ln n-1}{4(n\ln n)^2}. \end{aligned}$$

For $n \geq n_0$ it exists $\varepsilon_{(1)} > 0$ such as

$$C_2'\frac{n^2-2\ln n-1}{4(n\ln n)^2}e_m^{-2}(W) \leq \varepsilon_{(1)}.$$

Using that $e_n^2(W) \sim \frac{2}{\pi^2}(\ln n)^{-1}$ it is obvious that

$$e_{m^k}^2(W) \sim \frac{2}{k\pi^2}(\ln m)^{-1} \sim \frac{1}{k}e_m^2(W).$$

Then it exist $\varepsilon_{(2)} > 0$ for all $m \geq m_0$ such as

$$\frac{1}{k}e_m^2(W) \leq (1+\varepsilon_{(2)})e_{m^k}^2(W),$$

and

$$e_n^2(W,\alpha_{\otimes k}^{(2)}) \leq (1+\varepsilon_{(1)}+\varepsilon_{(2)})e_{m^k}^2(W_{[0,1]}).$$

Therefore a straightforward conclusion is that $\alpha_{\otimes k}^{(2)}$ is a m^k-optimal $(\varepsilon_{(1)}+\varepsilon_{(2)})$-quantizer. This procedure can not be used in the same way for the first concatenated quantizer $\alpha_{\otimes k}^{(2)}$. However we think that its behaviour should be similar. In that direction we made some numerical calculation and simulations to observe the behaviour of the quantization error for both concatenated quantizers.

Some numerical results

For numerical purposes we set

$$(n,m,k) = \{(16,2,4);(16,4,2);(25,5,2);(32,2,5);(81,3,4);(243,3,5);(256,4,4)\}.$$

The results could be seem in Table 6.1. The quantization error for the optimal n-quantizer of the Wiener process is also included. It is easy to see that for the dispositions with small n, $n = 16, 25$ the results for both concatenated quantizers are similar. However for the others values of (n,m,k) we observe that $e_n^2(W,\alpha_{\otimes k}^{(1)}) < e_n^2(W,\alpha_{\otimes k}^{(2)})$. We know that this information is not enough to affirm that $\alpha_{\otimes k}^{(1)}$ is a n-optimal ε-quantizer but it seems to be true. It is important to point out that these results strongly depend on the choice made for (n,m,k).

Some of the quantizers that produces the results in Table 6.1 are in Figures 6.6, 6.7, 6.8 and 6.9. The others configurations are in Appendix E

Number of Quantizers(n)	$e_n^2(W)$	$e_n^2(W, \alpha_{\otimes k}^{(1)})$	$e_n^2(W, \alpha_{\otimes k}^{(2)})$
$n = 16 = 2^4$	0.0789	0.1082	0.1052
$n = 16 = 4^2$	0.0789	0.0852	0.0842
$n = 25 = 5^2$	0.0680	0.0744	0.0734
$n = 32 = 2^5$	0.0634	0.0904	0.0921
$n = 81 = 3^4$	0.0503	0.0593	0.0774
$n = 243 = 3^5$	0.0403	0.0481	0.0728
$n = 256 = 4^4$	0.0400	0.0452	0.0672

Table 6.1: Quantization error for both concatenated quantizers and the optimal ones for the Wiener.

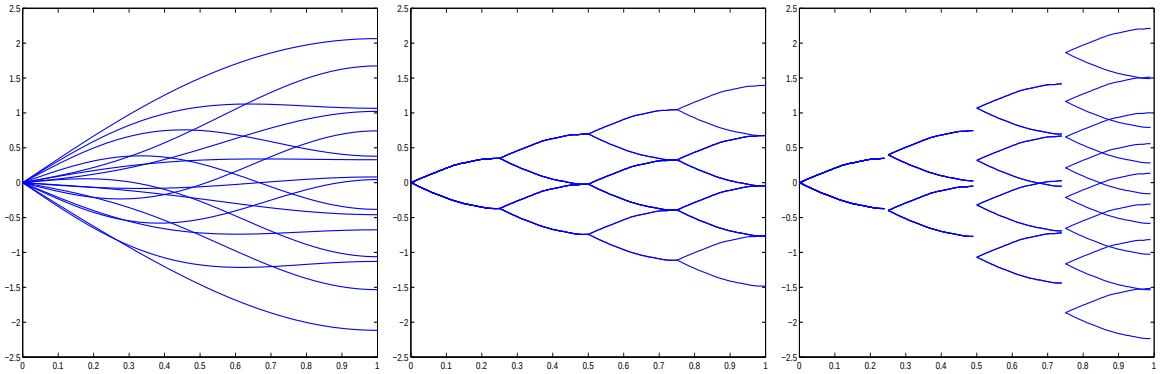

Figure 6.6: *From left to right*: 16-optimal quantizer for the Brownian motion W, $16(2^4)$-quantizer $(\alpha_{\otimes k}^{(1)})$ and $16(2^4)$-quantizer $(\alpha_{\otimes k}^{(2)})$.

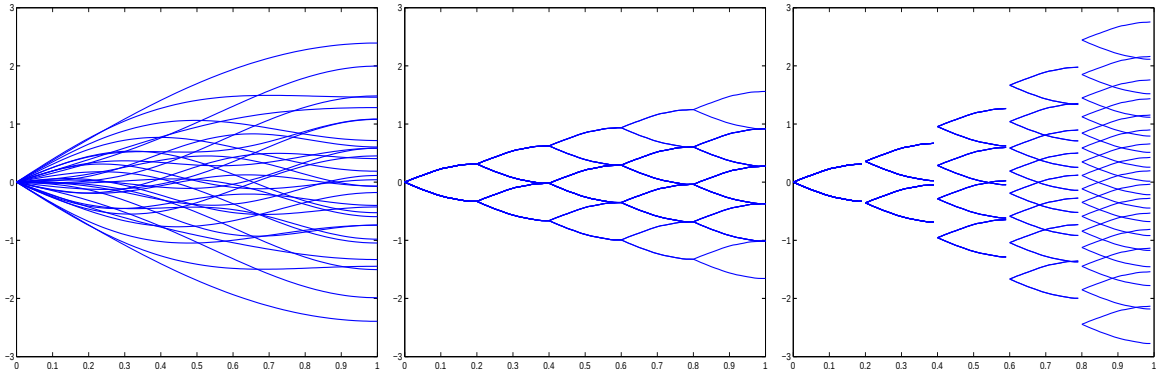

Figure 6.7: *From left to right*: 32-optimal quantizer for the Brownian motion W, $32(2^5)$-quantizer $(\alpha_{\otimes k}^{(1)})$ and $32(2^5)$-quantizer $(\alpha_{\otimes k}^{(2)})$.

As before for the quantization error the choice of (n, m, k) is important for the shape of the concatenated quantizers. The calculations results of the maximal radius for these quantizers are in Table 6.2. It easy to see that the maximal radius exhibits a large

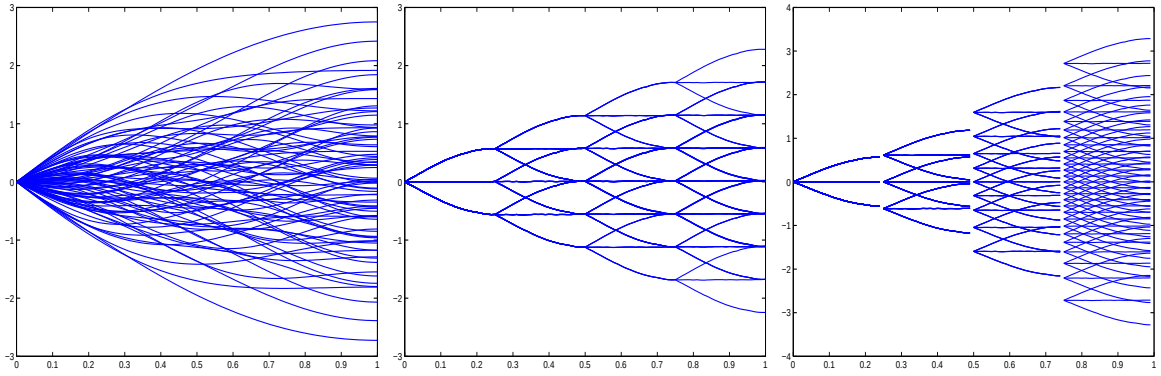

Figure 6.8: *From left to right*: 81-optimal quantizer for the Brownian motion W, $81(3^4)$-quantizer $(\alpha^{(1)}_{\otimes k})$ and $81(3^4)$-quantizer $(\alpha^{(2)}_{\otimes k})$.

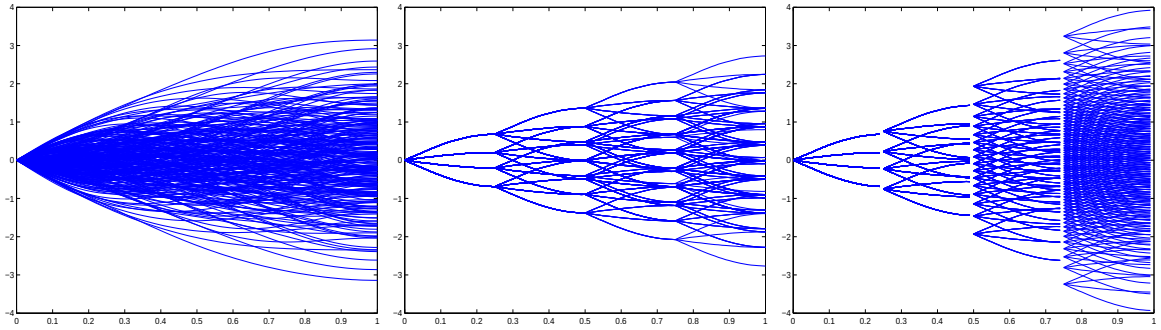

Figure 6.9: *From left to right*: 256-optimal quantizer for the Brownian motion W, $256(4^4)$-quantizer $(\alpha^{(1)}_{\otimes k})$ and $256(4^4)$-quantizer $(\alpha^{(2)}_{\otimes k})$.

variation in the results for the concatenated quantizer. Observing the results in Table 6.2 with the associated figures of the concatenated quantizers it seems that

$$\rho_n^2(\alpha^{(1)}_{\otimes k}) < \rho_n^2(W) < \rho_n^2(\alpha^{(2)}_{\otimes k}).$$

This assertion only holds as a conjecture. A deeper analysis on (n, m, k) is needed besides a full proof. However this fact "confirm" our initial idea to find some general properties on some class of quantizers (n-optimal ε-quantizer) that allow us to obtain some results for the maximal radius of the Wiener process.

Some analysis of the maximal radius for the n-optimal ε-quantizer

We work here with the second concatenated quantizer $\alpha^{(2)}_{\otimes k}$. As usual $\rho_n = \max\{\|a\|, a \in \alpha\}$ is the maximal radius for the optimal n-quantizer of the Wiener process. Let us define by $\rho_n(\varepsilon) = \max\{\|a\|, a \in \alpha^\varepsilon\}$ the maximal radius for a n-optimal ε-quantizer α^ε and

$$\begin{aligned} \rho_n^\varepsilon &= \inf\{\max_{a\in\alpha^\varepsilon} \|a\|, \text{ for all the } n\text{-optimal } \varepsilon\text{-quantizers}\} \\ &= \inf\{\rho_n(\varepsilon), \text{ for all the } n\text{-optimal } \varepsilon\text{-quantizers}\}. \end{aligned}$$

Number of Quantizers(n)	$\rho_n^2(\alpha_{\otimes k}^{(1)})$	$\rho_n^2(W)$	$\rho_n^2(\alpha_{\otimes k}^{(2)})$
$n = 16 = 2^4$	0,8228	2,2088	1,6907
$n = 16 = 4^2$	1,5647	2,2088	1,7137
$n = 25 = 5^2$	1,9544	2,6962	2,2105
$n = 32 = 2^5$	1,0074	2,9414	2,5283
$n = 81 = 3^4$	1,9470	3,8226	3,7033
$n = 243 = 3^5$	2,3840	5,0819	5,3478
$n = 256 = 4^4$	2,8476	5,1325	5,3273
$n = 1024 = 2^{10}$	1,9337	6,8964	7,8884
$n = 1024 = 4^5$	3,4935	6,8964	7,5666

Table 6.2: Maximal radius for both concatenated quantizers and the optimal one for the Wiener.

We proof that $\alpha_{\otimes k}^{(2)}$ is a n-optimal ε-quantizer then

$$\rho_{m^k}^{\varepsilon} \leq \rho_{m^k}^{\otimes},$$

where $\rho_{m^k}^{\otimes}$ is the maximal radius for $\alpha_{\otimes k}^{(2)}$. By Lemma 6.1 it is easy to check that

$$\alpha_m^{(k)} = \frac{1}{\sqrt{k}}\,\alpha_m.$$

Hence the maximal radius $\rho_m^{(k)}$ for $\alpha_m^{(k)}$ satisfies that

$$\rho_m^{(k)} = \frac{1}{\sqrt{k}}\rho_m.$$

In the sequel we work with three different norms: $\|\cdot\|_{\mathcal{L}^1}$, $\|\cdot\|_{\mathcal{L}^2}$ and $\|\cdot\|_\infty$. For simplicity we begin our analysis with this last one.

Norm ∞: The maximal radius $\rho_n^{\otimes}$ for $\alpha_{\otimes k}^{(2)}$ satisfies that

$$\begin{aligned}
\rho_n^{\otimes} = \rho_{m^k}^{\otimes} = \|a_j^k\|_\infty &= \left\| a_{l_j^1}(t)\mathbf{1}_{[0,\frac{1}{k}]}(t) + \sum_{i=2}^{k}\left(a_{l_j^i}\left(t-(i-1)/k\right) + b_{j_i}^i\right)\mathbf{1}_{]\frac{i-1}{k},\frac{i}{k}]}(t)\right\|_\infty \\
&= \left\|\left(a_{l_j^k}\left(t-(k-1)/k\right) + b_{j_k}^k\right)\mathbf{1}_{]\frac{k-1}{k},1]}(t)\right\|_\infty \\
&= \rho_m^{(k)} + \widehat{\rho}_{m^{(k)}} = \frac{1}{\sqrt{k}}\rho_m + \widehat{\rho}_{m^{(k)}},
\end{aligned}$$

where $\widehat{\rho}_{m^{(k)}} = \max\{\|b\|_2, b \in \beta\}$ (β is an optimal $n^{\frac{k-1}{k}}$-quantizer for the $N(0, \frac{k-1}{k})$). Now

$$\widehat{\rho}_{m^{(k)}} \leq \sqrt{k-1}\,\widehat{\rho}_{m^{(1)}},$$

where $\widehat{\rho}_{m^{(1)}} = \max\{\|b\|_2, b \in \beta\}$ (β is an optimal $n^{\frac{1}{k}}$-quantizer for the $N(0, \frac{k-1}{k})$). Hence

$$\widehat{\rho}_{m^{(k)}} \leq \sqrt{k-1}\,\widehat{\rho}_{m^{(1)}} = \sqrt{k-1}\sqrt{\frac{k-1}{k}}\widehat{\rho}_{m^{(0)}} = \frac{k-1}{\sqrt{k}}\widehat{\rho}_{m^{(0)}}, \tag{6.10}$$

where $\widehat{\rho}_{m^{(0)}} = \max\{\|b\|_2, b \in \beta\}$ (β is an optimal $n^{\frac{1}{k}}$-quantizer for the $N(0,1)$). Therefore

$$\begin{aligned}\rho_n^{\otimes} = \rho_{m^k}^{\otimes} = \|a_j^k\|_{\infty} &= \rho_m^{(k)} + \widehat{\rho}_{m^{(i)}} = \frac{1}{\sqrt{k}}\rho_m + \widehat{\rho}_{m^{(k)}} \\ &\leq \frac{1}{\sqrt{k}}\rho_m + \frac{k-1}{\sqrt{k}}\widehat{\rho}_{m^{(0)}}\end{aligned}$$

The relation between ρ_m and $\widehat{\rho}_{m^{(0)}}$ is unknown. Then we obtain that

$$\rho_{m^k}^{\varepsilon} \leq \rho_{m^k}^{\otimes} \leq \frac{1}{\sqrt{k}}(\rho_m - \widehat{\rho}_{m^{(0)}}) + \sqrt{k}\widehat{\rho}_{m^{(0)}}.$$

Thus

$$\frac{\rho_n^{\varepsilon}}{\sqrt{\ln n}} \leq \frac{\sqrt{\ln m}}{\ln n}(\rho_m - \widehat{\rho}_{m^{(0)}}) + \frac{\widehat{\rho}_{m^{(0)}}}{\sqrt{\ln m}}.$$

This last equation implies that

$$\limsup \frac{\rho_n^{\varepsilon}}{\sqrt{\ln n}} \leq \frac{\widehat{\rho}_{m^{(0)}}}{\sqrt{\ln m}} < +\infty,$$

However this conclusion is not sufficient to obtain information about

$$\liminf \frac{\rho_n}{\sqrt{\ln n}} \quad \text{or} \quad \limsup \frac{\rho_n}{\sqrt{\ln n}}.$$

Norm $\mathcal{L}^2$: Here we have

$$\begin{aligned}\rho_n^{\otimes} = \|a_j^k\|_{\mathcal{L}^2} &= \left\| a_{l_j^1}(t)\mathbf{1}_{[0,\frac{1}{k}]}(t) + \sum_{i=2}^{k}\left(a_{l_j^i}\left(t-(i-1)/k\right) + b_{j_i}^i\right)\mathbf{1}_{]\frac{i-1}{k},\frac{i}{k}]}(t)\right\|_{\mathcal{L}^2} \\ &\leq \left\| a_{l_j^1}(t)\mathbf{1}_{[0,\frac{1}{k}]}(t)\right\|_{\mathcal{L}^2} + \sum_{i=2}^{k}\left\|\left(a_{l_j^i}\left(t-(i-1)/k\right) + b_{j_i}^i\right)\mathbf{1}_{]\frac{i-1}{k},\frac{i}{k}]}(t)\right\|_{\mathcal{L}^2} \\ &\leq \frac{1}{\sqrt{k}}\rho_m + \sum_{i=2}^{k}\left(\left\|a_{l_j^i}\left(t-(i-1)/k\right)\mathbf{1}_{]\frac{i-1}{k},\frac{i}{k}]}(t)\right\|_{\mathcal{L}^2} + \left\|b_{j_i}^i\mathbf{1}_{]\frac{i-1}{k},\frac{i}{k}]}(t)\right\|_{\mathcal{L}^2}\right) \\ &= \sqrt{k}\rho_m + \frac{1}{\sqrt{k}}\sum_{i=2}^{k}\left\|b_{j_i}^i\right\|_2 = \sqrt{k}\rho_m + \frac{1}{\sqrt{k}}\sum_{i=2}^{k}\widehat{\rho}_{m^{(i)}},\end{aligned}$$

where $\widehat{\rho}_{m^{(i)}} = \max\{\|b\|_2, b \in \beta_n^{(i)}\}$ for $i = 2, \cdots, k$ (we recall that $\beta_n^{(i)} = (b_1^i \cdots, b_{m^{(i)}}^i)$ is the optimal $n^{\frac{i-1}{k}}$-quantizer for the $N(0, \frac{i-1}{k})$ for $i = 2, \cdots, k$). By the same idea used in (6.10) we state that

$$\widehat{\rho}_{m^{(i)}} \leq \frac{i-1}{\sqrt{k}}\rho_m.$$

Therefore

$$\begin{aligned}\rho_n^{\otimes} = \|a_j^k\|_{\mathcal{L}^2} &\leq \sqrt{k}\rho_m + \frac{1}{\sqrt{k}}\sum_{i=2}^{k}\widehat{\rho}_{m^{(i)}} \leq \sqrt{k}\rho_m + \frac{1}{\sqrt{k}}\widehat{\rho}_{m^{(0)}}\sum_{i=2}^{k}\frac{i-1}{\sqrt{k}} \\ &= \sqrt{k}\rho_m + \frac{k-1}{2}\widehat{\rho}_{m^{(0)}}.\end{aligned}$$

Using this procedure we can not obtain an useful result.

Norm $\mathcal{L}^1$: Here we made one last attempt to obtain a result useful for the maximal radius of the optimal quantizer of the Wiener process. As before we have that

$$\begin{aligned}
\rho_n^{\otimes} = \|a_j^k\|_{\mathcal{L}^1} &= \left\| a_{l_j^1}(t)\mathbf{1}_{[0,\frac{1}{k}]}(t) + \sum_{i=2}^{k} \left(a_{l_j^i}\left(t-(i-1)/k\right) + b_{j_i}^i \right) \mathbf{1}_{]\frac{i-1}{k},\frac{i}{k}]}(t) \right\|_{\mathcal{L}^1} \\
&\leq \left\| a_{l_j^1}(t)\mathbf{1}_{[0,\frac{1}{k}]}(t) \right\|_{\mathcal{L}^1} + \sum_{i=2}^{k} \left\| \left(a_{l_j^i}\left(t-(i-1)/k\right) + b_{j_i}^i \right) \mathbf{1}_{]\frac{i-1}{k},\frac{i}{k}]}(t) \right\|_{\mathcal{L}^1} \\
&\leq \frac{1}{\sqrt{k}}\rho_m + \sum_{i=2}^{k} \left(\left\| a_{l_j^i}\left(t-(i-1)/k\right) \mathbf{1}_{]\frac{i-1}{k},\frac{i}{k}]}(t) \right\|_{\mathcal{L}^1} + \left\| b_{j_i}^i \mathbf{1}_{]\frac{i-1}{k},\frac{i}{k}]}(t) \right\|_{\mathcal{L}^1} \right) \\
&= \sqrt{k}\rho_m + \frac{1}{k}\sum_{i=2}^{k} \left\| b_{j_i}^i \right\|_1 = \sqrt{k}\rho_m + \frac{1}{k}\sum_{i=2}^{k} \widehat{\rho}_{m^{(i)}}.
\end{aligned}$$

Using that

$$\widehat{\rho}_{m^{(i)}} \leq \frac{i-1}{\sqrt{k}}\rho_m,$$

it follows

$$\begin{aligned}
\rho_n^{\otimes} = \|a_j^k\|_{\mathcal{L}^1} &\leq \sqrt{k}\rho_m + \frac{1}{\sqrt{k}}\sum_{i=2}^{k} \widehat{\rho}_{m^{(i)}} \leq \sqrt{k}\rho_m + \frac{1}{k}\widehat{\rho}_{m^{(0)}} \sum_{i=2}^{k} \frac{i-1}{\sqrt{k}} \\
&= \sqrt{k}\rho_m + \frac{k-1}{2\sqrt{k}}\widehat{\rho}_{m^{(0)}} \leq \sqrt{k}(\rho_m + \frac{\widehat{\rho}_{m^{(0)}}}{2}).
\end{aligned}$$

Hence we obtain that

$$\rho_{m^k}^{\varepsilon} \leq \rho_{m^k}^{\otimes} \leq \sqrt{k}(\rho_m + \frac{\widehat{\rho}_{m^{(0)}}}{2}).$$

Thus

$$\frac{\rho_n^{\varepsilon}}{\sqrt{\ln n}} \leq \frac{\rho_m}{\sqrt{\ln m}} + \frac{\widehat{\rho}_{m^{(0)}}}{2\sqrt{\ln m}}.$$

This last equation implies that

$$\limsup \frac{\rho_n^{\varepsilon}}{\sqrt{\ln n}} \leq \frac{\rho_m}{\sqrt{\ln m}} + \frac{\widehat{\rho}_{m^{(0)}}}{2\sqrt{\ln m}} < +\infty.$$

We obtain a similar result like in the case of the norm uniform.

Some conclusions

The initial idea about the concatenated quantizers was promising in the beginning. However we do not achieve goods results in the direction that we need. We find some interesting results related with the the concatenated quantizers but neither of them were useful.

6.2.3 Asymptotics for the maximal radius of stationary quantizers

In the sequel we work with the product quantizers defined in Pagès and Printems [2005] (is the IV quantization design presented in Luschgy et al. [2010]). The main idea is to use the Karhunen-Loève expansion of a Gaussian process X to produce a functional quantization for X using at most n elementary quantizers with some *product quantizer* of the form

$$\widehat{X}_t = \sum_{j\geq 1} \sqrt{\lambda_j}\widehat{\xi}_j\varphi_j(t), \tag{6.11}$$

where $\{\varphi_j\}_{j\geq 1}$ is the KL basis for X and $(\lambda_j)_{j\geq 1}$ the orderer eigenvalues associated to $\{\varphi_j\}_{j\geq 1}$, $\widehat{\xi}_j$ is the optimal n_j-quantizer $N(0,1)$ and $n_1 \times \cdots \times n_k \leq n$, $n_1, n_2, \cdots, n_k \geq 1$ for all k. Note that for k large enough $n_k = 1$ and that implies $\widehat{\xi}_j = 0$. Therefore the serie above turns on a finite sum when n is fixed for k large enough. The $n_1 \times \cdots \times n_k$-quantizer χ that produces (6.11) is of the form

$$\chi_{\underline{i}}(t) = \sum_{j\geq 1} \sqrt{\lambda_j} x_{i_j}^{n_j} \varphi_j(t), \tag{6.12}$$

where $x^{n_j} = (x_1^{n_j}, \cdots, x_{n_j}^{n_j})$ is the unique n_j-quantizer of the standard normal distribution in one dimension, $\underline{i} = (i_1, i_2, \cdots, i_k)$ and $i_j \in \{1, \cdots, n_j\}$ for all $j = 1, 2, \cdots, k$.

Let us denote by $\mathcal{O}(X, n)$ the set

$$\mathcal{O}(X,n) = \{\chi : \chi \text{ is the KL product quantizer of size at most } n \text{ defined by } (6.12)\}.$$

For each n we can define D_n as

$$D_n = \max\{l : n_1 \times \cdots \times n_l \leq n,\ n_1, \cdots, n_l \geq 2\}.$$

It is easy to check for n fixed that $D_n \leq k$. Obviously D_n is unbounded when $n \to +\infty$, hence the finite sum in (6.12) turns on an infinite sum, but the best quantizer χ for each n could not be the one with D_n summands in (6.12). However the infinity of the sum could be a problem. In order to avoid this we can work with a suitable subset of $\mathcal{O}(X, n)$. Fixing a constant $k_0 \in \mathbb{N}$ for all n we can write the following.

$$\chi_{\underline{i}}^{k_0}(t) = \sum_{j=1}^{k_0} \sqrt{\lambda_j} x_{i_j}^{n_j} \varphi_j(t), \tag{6.13}$$

where $x^{n_j} = (x_1^{n_j}, \cdots, x_{n_j}^{n_j})$ is the unique n_j-quantizer of the standard normal distribution in one dimension, $\underline{i} = (i_1, i_2, \cdots, i_{k_0})$ and $i_j \in \{1, \cdots, n_j\}$ for all $j = 1, 2, \cdots, k_0$. Therefore we set

$$\mathcal{O}_{k_0}(X,n) \ = \ \left\{\chi^{k_0} : \chi^{k_0} \text{ is the KL product quantizer of size at most } n \text{ defined by } (6.13)\right\}.$$

By the definition above it is easy to check that

$$\mathcal{O}_{k_0}(X,n) \subset \mathcal{O}(X,n),$$

for all n and k_0 fixed.

***Theorem* 6.4**
If we denote

$$\rho_{n,k_0} = \rho_n^2(\chi^{k_0}) = \max\{\|\chi_{\underline{i}}^{k_0}\|_{\mathcal{L}^2}, \chi_{\underline{i}}^{k_0} \in \mathcal{O}_{k_0}(X,n)\},$$

as the maximal radius for $\chi^{k_0} \in \mathcal{O}_{k_0}(X,n)$ *then*

$$c_{k_0} \leq \liminf_{n\to+\infty} \frac{\rho_{n,k_0}}{\sqrt{\ln n}} \leq \limsup_{n\to+\infty} \frac{\rho_{n,k_0}}{\sqrt{\ln n}} \leq C_{k_0}, \tag{6.14}$$

where c_{k_0}, C_{k_0} *are some positive constants.*

Proof.

If we define the maximal radius for the optimal n_j-quantizer x^{n_j} of the $N(0,1)$ as $\rho_{n_j} = \max\{|a|, a \in x^{n_j}\}$ then it follows that

$$\rho_{n,k_0}^2 = \rho_n^2(\chi^{k_0}) = \sum_{j=1}^{k_0} \lambda_j \rho_{n_j}^2.$$

We know that $\prod_{j=1}^{k_0} n_j \leq n$. We can always assume that there exists a $n_1 \times \cdots \times n_{k_0}$-quantizer χ^{k_0} for which $\prod_{j=1}^{k_0} n_j \sim n$ when $n \to +\infty$. From now on we work with this quantizer. This last assertion allows us to conclude that exist at least one subsequence $((n_r)_j)_{r\geq 1}$ of at least one n_j such as $(n_r)_j \to +\infty$ when $n \to +\infty$. Let us define the following set

$$J = \{j, 1 \leq j \leq k_0 : \exists (n_r)_j \to +\infty \text{ when } n \to +\infty\},$$

then it follows that

$$\begin{aligned} \sum_{j\in J} \lambda_j \rho_{n_j}^2 \leq \rho_n^2(\chi^{k_0}) &= \sum_{j\notin J} \lambda_j \rho_{n_j}^2 + \sum_{j\in J} \lambda_j \rho_{n_j}^2 \\ &\leq C_J + \sum_{j\in J} \lambda_j \rho_{n_j}^2, \end{aligned}$$

where C_J is certain finite and positive constant for all n.

If $|J| = 1$ for all n then we can write $J = \{j_0\}$. Hence

$$\lambda_{j_0} \rho_{n_{j_0}}^2 \leq \rho_n^2(\chi^{k_0}) \leq C_J + \lambda_{j_0} \rho_{n_{j_0}}^2,$$

and obviously $n_{j_0} \approx n$ (moreover we can write $c_{j_0} n \leq n_{j_0} \leq n$ with $c_{j_0} \leq 1$), then

$$\lambda_{j_0} \frac{\rho_{n_{j_0}}^2}{\ln n} \leq \frac{\rho_n^2(\chi^{k_0})}{\ln n} \leq \frac{C_J + \lambda_{j_0} \rho_{n_{j_0}}^2}{\ln n} = C_j(n) + \lambda_{j_0} \frac{\rho_{n_{j_0}}^2}{\ln n},$$

where $C_j(n) = C_J / \ln n$.

The previous equation obviously depends on j_0. In order to get a general result on $\mathcal{O}_{k_0}(X,n)$ for all χ^{k_0} with only one element in J we need to avoid this dependency. This can be easily avoided regarding that we fixed k_0. Therefore it follows that

$$\lambda_{k_0} \frac{\rho_{n_{j_0}}^2}{\ln n} \leq \frac{\rho_n^2(\chi^{k_0})}{\ln n} \leq C_j(n) + \lambda_1 \frac{\rho_{n_{j_0}}^2}{\ln n}.$$

Taking limit when $n \to +\infty$ it follows that

$$(6\lambda_{k_0})^{1/2} \leq \liminf_{n\to+\infty} \frac{\rho_n(\chi^{k_0})}{\sqrt{\ln n}} \leq \limsup_{n\to+\infty} \frac{\rho_n(\chi^{k_0})}{\sqrt{\ln n}} \leq (6\lambda_1)^{1/2}.$$

This last assertion comes from Sagna [2008]. Obviously (6.14) follows after that. The same idea can be used when $|J| \geq 2$ for all n.

Let us assume that $|J| = j^0 \geq 2$. To simplify notation we assume that the first j^0 indexes are in J. Therefore we write

$$\sum_{j=1}^{j^0} \lambda_j \rho_{n_j}^2 \leq \rho_n^2(\chi^{k_0}) \leq C_0 + \sum_{j=1}^{j^0} \lambda_j \rho_{n_j}^2,$$

where C_0 is certain finite and positive constant for all n.

Moreover

$$\lambda_{j^0} \sum_{j=1}^{j^0} \rho_{n_j}^2 \leq \rho_n^2(\chi^{k_0}) \leq C_0 + \lambda_1 \sum_{j=1}^{j^0} \rho_{n_j}^2.$$

For every $j \in J$ we know that exist a subsequence $(n_r)_j \to +\infty$ when n go toward infinity, then

$$\limsup_{n\to+\infty} \sum_{j=1}^{j^0} \frac{\rho_{n_j}^2}{\ln n} \quad \leq \quad \sum_{j=1}^{j^0} \limsup_{n\to+\infty} \frac{\rho_{n_j}^2}{\ln n_j} \leq 6j^0.$$

Therefore we have

$$\limsup_{n\to+\infty} \frac{\rho_n(\chi^{k_0})}{\sqrt{\ln n}} \leq (6j^0\lambda_1)^{1/2}.$$

For the lim inf we proceed as follows.

$$\sum_{j=1}^{j^0} \frac{\rho_{n_j}^2}{\ln n} \quad = \quad \sum_{j=1}^{j^0} \frac{\rho_{n_j}^2}{\ln n_j} \frac{\ln n_j}{\ln n},$$

and for all $j \in J$ we have that

$$\liminf_{n\to+\infty} \frac{\rho_{n_j}^2}{\ln n_j} \geq 6.$$

Therefore we can write that

$$\frac{\rho_{n_j}^2}{\ln n_j} = 6 + \varepsilon_j^n,$$

where $\varepsilon_j^n \to 0$ for all $j \in J$ when $n \to +\infty$. Hence we have that

$$\sum_{j=1}^{j^0} \frac{\rho_{n_j}^2}{\ln n} \quad \geq \quad 6\frac{\ln \prod_{j=1}^{k_0} n_j}{\ln n} + \sum_{j=1}^{j^0} \varepsilon_j^n \frac{\ln n_j}{\ln n}.$$

Taking lim inf when $n \to +\infty$ it follows that

$$\liminf_{n\to+\infty} \sum_{j=1}^{j^0} \frac{\rho_{n_j}^2}{\ln n} \quad \geq \quad 6.$$

This assertion follows from the fact that

$$\frac{\ln \prod_{j=1}^{k_0} n_j}{\ln n} \sim 1,$$

because $\prod_{j=1}^{k_0} n_j \sim n$ and $n \to +\infty$. Obviously the second term

$$\sum_{j=1}^{j^0} \varepsilon_j^n \frac{\ln n_j}{\ln n} \to 0.$$

Therefore we obtain

$$\liminf_{n\to+\infty} \frac{\rho_n^2(\chi^{k_0})}{\ln n} \quad = \quad \lambda_{j^0} \liminf_{n\to+\infty} \sum_{j=1}^{j^0} \frac{\rho_{n_j}^2}{\ln n} \geq 6\lambda_{j^0}.$$

Finally we have

$$(6\lambda_{j^0})^{1/2} \leq \liminf_{n\to+\infty} \frac{\rho_n(\chi^{k_0})}{\sqrt{\ln n}} \leq \limsup_{n\to+\infty} \frac{\rho_n(\chi^{k_0})}{\sqrt{\ln n}} \leq (6j^0\lambda_1)^{1/2}.$$

As before for the case of $|J| = 1$ we need to avoid the dependency on j^0 for the lower limit. Using the same idea we obtain that

$$(6\lambda_{k_0})^{1/2} \leq \liminf_{n\to+\infty} \frac{\rho_n(\chi^{k_0})}{\sqrt{\ln n}} \leq \limsup_{n\to+\infty} \frac{\rho_n(\chi^{k_0})}{\sqrt{\ln n}} \leq (6j^0\lambda_1)^{1/2},$$

and that conclude the proof

□

***Remark* 6.3**
It is easy to check that the constants c_{k_0} and C_{k_0} in (6.14) are

$$c_{k_0} = (6\lambda_{k_0})^{1/2}, \qquad C_{k_0} = (6|J|\lambda_1)^{1/2}.$$

It is straightforward that the behaviour of both constants depends on k_0:

$$\lim_{k_0\to\infty} c_{k_0} = 0, \quad \textit{and} \quad \lim_{k_0\to\infty} C_{k_0} = +\infty.$$

***Remark* 6.4**
The procedure used to prove Theorem 6.4 remains valid for others quantizer design like the proposed in Luschgy et al. [2010]. Specifically design I, II and III. The first one is associated to the optimal quantizer (see for instance Subsection 2.2.2). More specifically a version of Theorem 6.4 could be written if we take in Design I a fixed dimension d_n for all value of n and for Design II and III a fixed value for m.

6.2.4 Asymptotics for the maximal radius of linear quantizers

Here we work directly with an approximation of the stochastic process X in $[0,1]$. Let us denote by π a fixed time partition as follows

$$\pi = \{0 = t_0 < t_1 < \cdots < t_{k-1} < t_k = 1\}.$$

The linear interpolation of X using π is

$$X_L^\pi(t) = \sum_{i=0}^{k-1} \left[\frac{t_{i+1} - t}{\Delta_i} X(t_i) + \frac{t - t_i}{\Delta_i} X(t_{i+1}) \right] \mathbf{1}_{]t_i, t_{i+1}]}(t),$$

where $\Delta_i = t_{i+1} - t_i$, $\forall\, i = 1, 2, \cdots, k-1$. The exponent π in $X_L^\pi(t)$ will often be dropped when no confusion can arise. From now on we denote the length of the longest subinterval of π by

$$\delta_\pi = \sup_{0 \le i \le k-1} |\Delta_i|.$$

Let us define by H^π the linear set of polygonal functions X_L in $[0,1]$ with interpolation knots t_j, $j = 0, 1, \cdots, k$. Obviously $H^\pi \subset H = \mathcal{L}^2([0,1], dt)$. Let us define the bounded linear operator T^π by

$$\begin{array}{rcl} T^\pi : (\mathcal{L}^2, \|\cdot\|_{\mathcal{L}^2}) & \longrightarrow & (H^\pi, \|\cdot\|_{\mathcal{L}^2}) \\ X & \overset{\pi}{\rightsquigarrow} & T^\pi(X) = X_L^\pi. \end{array}$$

It is clear that T^π is surjective. We define $A^\pi = (T^\pi)^{-1}(f^\pi)$ where $f^\pi \in H^\pi$, obviously $A^\pi \neq \emptyset$ because f^π belongs to it. The norm defined in H^π is the same as in H. For $X_L \in H^\pi$ it holds

$$\|X_L\|_{\mathcal{L}^2}^2 = \frac{1}{3} \sum_{i=0}^{k-1} \Delta_i (X^2(t_{i+1}) + X^2(t_i) + X(t_i)X(t_{i+1})).$$

The previous formula allows us to compute the norm of X_L exactly knowing the vector $X^\pi = (X(t_0), \cdots, X(t_n)) \in \mathbb{R}^{k+1}$. The following norm $\|\cdot\|_{\mathbb{R}^{k+1}}^2$ is equivalent to the Euclidean one in $\mathbb{R}^{k+1}$

$$\|X^\pi\|_{\mathbb{R}^{k+1}}^2 \overset{\triangle}{=} \|X_L\|_{\mathcal{L}^2}^2. \tag{6.15}$$

More precisely we have

***Lemma* 6.2**
The norm $\|\cdot\|_{\mathbb{R}^{k+1}}^2$ associated to a fix partition π satisfies that

$$\frac{\delta_{(1)}}{6} \|X\|_2^2 \le \|X\|_{\mathbb{R}^{k+1}}^2 \le \delta_\pi \|X\|_2^2,$$

where $\|\cdot\|_2$ is the Euclidean norm in $\mathbb{R}^{k+1}$ and $\delta_{(1)} = \min_{0 \le i \le k-1} |t_{i+1} - t_i|$.

Proof

Let us begin with the right inequality

$$\begin{aligned}
\|X\|^2_{\mathbb{R}^{k+1}} &= \frac{1}{3}\sum_{i=0}^{k-1}\Delta_i(X^2(t_{i+1})+X^2(t_i)+X(t_i)X(t_{i+1})) \\
&\leq \frac{\delta_\pi}{3}\sum_{i=0}^{k-1}(X^2(t_{i+1})+X^2(t_i)+X(t_i)X(t_{i+1})) \\
&= \frac{\delta_\pi}{3}(2\|X\|_2^2-X^2(t_0)-X^2(t_k))+\frac{\delta_\pi}{6}\sum_{i=0}^{k-1}(2X(t_i)X(t_{i+1})) \\
&\leq \frac{\delta_\pi}{3}(2\|X\|_2^2-X^2(t_0)-X^2(t_k))+\frac{\delta_\pi}{6}(2\|X\|_2^2-X^2(t_0)-X^2(t_k)) \\
&\leq \frac{\delta_\pi}{3}2\|X\|_2^2+\frac{\delta_\pi}{6}2\|X\|_2^2=\delta_\pi\|X\|_2^2.
\end{aligned}$$

For the left inequality we follow in the same way

$$\begin{aligned}
\|X\|^2_{\mathbb{R}^{k+1}} &\geq \frac{\delta_{(1)}}{3}\sum_{i=0}^{k-1}(X^2(t_{i+1})+X^2(t_i)+X(t_i)X(t_{i+1})) \\
&= \frac{\delta_{(1)}}{3}(2\|X\|_2^2-X^2(t_0)-X^2(t_k))+\frac{\delta_{(1)}}{6}\sum_{i=0}^{k-1}(2X(t_i)X(t_{i+1})) \\
&\geq \frac{\delta_{(1)}}{3}(2\|X\|_2^2-X^2(t_0)-X^2(t_k))-\frac{\delta_{(1)}}{6}(2\|X\|_2^2-X^2(t_0)-X^2(t_k)) \\
&\geq \frac{\delta_{(1)}}{3}\|X\|_2^2-\frac{\delta_{(1)}}{6}\|X\|_2^2=\frac{\delta_{(1)}}{6}\|X\|_2^2.
\end{aligned}$$

□

Linear approximation of a Brownian motion

The piecewise linear interpolator W_L of the Brownian motion W with interpolation knots t_i, $\forall\, i=0,1,\cdots,k$ is

$$T^\pi(W(t))=W_L^\pi(t)=\sum_{i=0}^{k-1}\left[\frac{t_{i+1}-t}{\Delta_i}W(t_i)+\frac{t-t_i}{\Delta_i}W(t_{i+1})\right]1_{]t_i,t_{i+1}]}(t), \tag{6.16}$$

and satisfies the following

***Proposition* 6.1**
For any partition $\pi=\{0=t_0<t_1<\cdots<t_{k-1}<t_k=1\}$ let W_L^π be the linear interpolation of the Wiener process W. It holds

$$\lim_{\delta_\pi\to 0}\mathbb{E}\|W-W_L^\pi\|_{\mathcal{L}^2}=0. \tag{6.17}$$

Proof

Using the expression of W_L^π it follows that

$$\begin{aligned} W(t) - W_L^\pi(t) &= \sum_{i=0}^{k-1} \left[\frac{t_{i+1}-t}{\Delta_i}(W(t)-W(t_i)) + \frac{t-t_i}{\Delta_i}(W(t)-W(t_{i+1})) \right] 1_{]t_i,t_{i+1}]}(t) \\ &\stackrel{\mathcal{D}}{=} \sum_{i=0}^{k-1} \left[W(t-t_i) - \frac{t-t_i}{\Delta_i} W(t_{i+1}-t_i) \right] 1_{]t_i,t_{i+1}]}(t) \\ &\stackrel{\mathcal{D}}{=} \sum_{i=0}^{k-1} \Delta_i^{1/2} \left[W\Big(\frac{t-t_i}{\Delta_i}\Big) - \frac{t-t_i}{\Delta_i} W(1) \right] 1_{]t_i,t_{i+1}]}(t), \end{aligned}$$

where $\stackrel{\mathcal{D}}{=}$ means equal in distribution. It is simple to verify that for all $t \in]t_i, t_{i+1}]$

$$W\Big(\frac{t-t_i}{\Delta_i}\Big) - \frac{t-t_i}{\Delta_i} W(1) \stackrel{\mathcal{D}}{=} B(s) \quad \forall\ s \in [0,1],$$

where B is a Brownian bridge. Therefore

$$W(t) - W_L(t) \stackrel{\mathcal{D}}{=} \sum_{i=0}^{k-1} \Delta_i^{1/2} B_i((t-t_i)/\Delta_i) 1_{]t_i,t_{i+1}]}(t),$$

where B_i is a Brownian bridge for all $i = 1, 2, \cdots, k-1$. Obviously

$$\mathbb{E}(W - W_L) = 0.$$

Furthermore it follows that

$$\begin{aligned} \mathbb{E}\|W - W_L\|_{\mathcal{L}^2}^2 &= \mathbb{E} \int_0^1 (W(t) - W_L(t))^2 dt \\ &= \mathbb{E} \int_0^1 \left(\sum_{i=0}^{k-1} \Delta_i^{1/2} B_i((t-t_i)/\Delta_i) 1_{]t_i,t_{i+1}]}(t) \right)^2 dt \\ &= \mathbb{E} \int_0^1 \sum_{i=0}^{k-1} \Delta_i B_i^2((t-t_i)/\Delta_i) 1_{]t_i,t_{i+1}]}(t) dt \\ &= \mathbb{E} \sum_{i=0}^{k-1} \Delta_i \int_{t_i}^{t_{i+1}} B_i^2((t-t_i)/\Delta_i) dt = \mathbb{E} \sum_{i=0}^{k-1} \Delta_i^2 \int_0^1 B_i^2(s) ds \\ &= \sum_{i=0}^{k-1} \Delta_i^2 \int_0^1 \mathbb{E} B_i^2(s) ds = \frac{1}{6} \sum_{i=0}^{k-1} \Delta_i^2. \end{aligned}$$

Obviously (6.17) follows.

□

For Gaussian processes we have also convergence in maximal quadratic mean

$$\lim_{\delta_\pi \to 0} \max_{t\in[0,1]} \mathbb{E}(W(t) - W_L^\pi(t))^2 = 0.$$

See for instance Hüsler et al. [2003] and references therein. The rate of approximation for the linear interpolation in mean square is also optimal in some sense for certain process (we refer the reader to Seleznjev [1996]). The optimal properties of linear interpolation for sample paths of random processes is extensively discussed in Su and Cambanis [1993].

***Theorem* 6.5**

1. *The n-quantization error $e_n^2(W_L^\pi)$ satisfies*

$$\lim_{\delta_\pi \to 0} e_n^2(W_L^\pi) = e_n^2(W), \tag{6.18}$$

   and

2. *if $\widetilde{e}_n^2(W^\pi)$ is the n-quantization error, using the norm defined in (6.15), for the Gaussian vector $W^\pi = (W(t_1), \cdots, W(t_k))$ in $\mathbb{R}^k$ then*

$$e_n^2(W_L^\pi) = \widetilde{e}_n^2(W^\pi). \tag{6.19}$$

Proof

Let us prove equation (6.18). The n-quantization error for W_L in H^π is

$$e_n^2(W_L^\pi) = \mathbb{E} \min_{1\le j\le n} ||W_L^\pi - \widehat{a}_{j,L}^\pi||_{\mathcal{L}^2}^2,$$

where $\widehat{a}_{j,L}^\pi \in \widehat{\alpha}_L^\pi$ for all $j = 1, \cdots, n$. The set $\widehat{\alpha}_L^\pi$ is an optimal n-quantizer for W_L^π in H^π. Therefore the functions $\widehat{a}_{j,L}^\pi$ are in H^π.

Then if $\alpha = (a_1 \cdots, a_n)$ is an optimal n-quantizer for W in H we have that

$$\begin{aligned}
e_n^2(W) &= \mathbb{E} \min_{1\le j\le n} ||W - a_j||_{\mathcal{L}^2}^2 = \mathbb{E}\left(\sum_{j=1}^n \mathbf{1}_{C_j}(W)||W - a_j||_{\mathcal{L}^2}^2\right) \\
&\le \mathbb{E}\left(\sum_{i=1}^n \mathbf{1}_{\widehat{C}_j^\pi}(W_L^\pi)||W - \widehat{a}_{j,L}^\pi||_{\mathcal{L}^2}^2\right) \\
&\le \mathbb{E}\sum_{j=1}^n \mathbf{1}_{\widehat{C}_j^\pi}(W_L^\pi)||W - W_L^\pi||_{\mathcal{L}^2}^2 + \mathbb{E}\sum_{j=1}^n \mathbf{1}_{\widehat{C}_j^\pi}(W_L^\pi)||W_L^\pi - \widehat{a}_{j,L}^\pi||_{\mathcal{L}^2}^2 \\
&\quad +2\left|\mathbb{E}\sum_{j=1}^n \mathbf{1}_{\widehat{C}_j^\pi}(W_L^\pi)\langle W - W_L^\pi, W - \widehat{a}_{j,L}^\pi\rangle\right| \\
&\le \mathbb{E}||W - W_L^\pi||_{\mathcal{L}^2}^2 + e_n^2(W_L^\pi) + 2\left(\mathbb{E}||W - W_L^\pi||_{\mathcal{L}^2}^2\right)^{1/2} e_n(W_L^\pi),
\end{aligned}$$

where $(C_j)_{1\leq j\leq n}$ and $(\widehat{C}_j^\pi)_{1\leq j\leq n}$ are the Voronoï partition associated to α and $\widehat{\alpha}_L^\pi$ respectively. Therefore by Proposition 6.1 it follows

$$e_n^2(W) \leq \liminf_{\delta_\pi\to 0} e_n^2(W_L^\pi). \tag{6.20}$$

For each $a_i \in \alpha$ we set

$$a_{j,L}^\pi(t) = \sum_{i=0}^{k-1}\left[\frac{t_{i+1}-t}{\Delta_i}a_j(t_i) + \frac{t-t_i}{\Delta_i}a_j(t_{i+1})\right] 1_{]t_i,t_{i+1}]}(t).$$

Note that $a_{j,L}^\pi$ is always constructed using a_j for any partition, hence it follows that

$$a_{j,L}^\pi \to a_j, \quad \text{when } \delta_\pi \to 0, \tag{6.21}$$

uniformly in t for all j. Therefore we have that

$$e_n^2(W_L^\pi) \;=\; \mathbb{E}\sum_{j=1}^n \mathbf{1}_{\widehat{C}_j^\pi}(W_L^\pi)||W_L^\pi - \widehat{a}_{j,L}^\pi||_{\mathcal{L}^2}^2 \leq \mathbb{E}\sum_{j=1}^n \mathbf{1}_{C_j}(W_L^\pi)||W_L^\pi - a_{j,L}^\pi||_{\mathcal{L}^2}^2.$$

Applying again Proposition 6.1 and (6.21) we have

$$\limsup_{\delta_\pi\to 0} e_n^2(W_L^\pi) \leq \limsup_{\delta_\pi\to 0} \mathbb{E}\sum_{j=1}^n \mathbf{1}_{C_j}(W_L^\pi)||W_L^\pi - a_{j,L}^\pi||_{\mathcal{L}^2}^2 = e_n^2(W).$$

Obviously using the equation above and (6.20) we obtain that

$$\limsup_{\delta_\pi\to 0} e_n^2(W_L^\pi) \leq e_n^2(W) \leq \liminf_{\delta_\pi\to 0} e_n^2(W_L^\pi),$$

and equation (6.18) follows.

Equation (6.19) is a direct application of Theorem 2.11. Taking in the setting of the theorem $T_1 \equiv H^\pi$, $T_2 \equiv \mathbb{R}^k$ and the isometry

$$\begin{aligned} I : H^\pi &\longrightarrow \mathbb{R}^k \\ W_L^\pi &\rightsquigarrow I(W_L^\pi) = W^\pi, \end{aligned}$$

we complete the proof.

□

***Remark* 6.5**

A straightforward conclusion of Theorem 6.5 is that if $\alpha = (a_1, \cdots, a_n)$ is an optimal n-quantizer for W it follows that

$$a_{j,L}^\pi(t) = \sum_{i=0}^{k-1}\left[\frac{t_{i+1}-t}{\Delta_i}\beta_j^i + \frac{t-t_i}{\Delta_i}\beta_j^{i+1}\right]\mathbf{1}_{]t_i,t_{i+1}]}(t), \qquad \forall\, j = 1, \cdots, n,$$

where $\beta^\pi = \{\beta_1, \cdots \beta_n\}$ is an optimal n-quantizer for W^π and

$$a_j(t_i) = \beta_j^i \qquad \forall\, a_j \in \alpha,\, i = 1, 2, \cdots, k.$$

The main theorem concerning the maximal radius of the linear interpolation of the Brownian motion is an immediate conclusion of the preceding theorem and the results of Sagna [2008].

Let us recall briefly the basic hypothesis used by Sagna to obtain the asymptotics for the maximal radius of an optimal quantizer of an $\mathbb{R}^d$-valued vector X.

- The random vector X has a distribution $\mathbb{P}$ with unbounded support and moment of order r.
- The distribution $\mathbb{P}$ for X must also satisfies that

$$\mathbb{P}(dx) \geq \varepsilon_0 \mathbf{1}_{x \in \overline{B}(x_0, r_0)} \lambda_d(dx), \qquad \varepsilon_0,\, r_0 > 0,\ x_0 \in \mathbb{R}^d,$$

  where λ_d is the Lebesgue measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and $\overline{B}(x_0, r_0)$ is a closed ball in $\mathbb{R}^d$. This assumption is fulfilled by the usual distributions, in particular all the Gaussian ones.
- The norm used in the framework of Sagna [2008] was the Euclidean one. This last hypothesis could be relaxed for some of the results provided by Sagna.

In our case the Gaussian vector $W^\pi = (W(t_1), \cdots, W(t_{k-1}), W(t_k))$ defined in $\mathbb{R}^k$ has a multidimensional normal distribution $\mathbb{P} \equiv N(0, \Sigma)$, where

$$\Sigma = (\sigma_{ij})_{1 \leq i,j \leq k},$$

with $\sigma_{ii} = t_i$ and $\sigma_{ij} = t_i \wedge t_j$ for all $1 \leq i, j \leq k$.

Obviously this distribution satisfies all the assumptions needed to obtain the asymptotics for the maximal radius of an optimal n-quantizer for W^π. However, we need to find the correct constants for the asymptotics of the maximal radius under the norm $\|\cdot\|_{\mathbb{R}^k}$ instead of the Euclidean one. Obviously the asymptotic rate for $\widetilde{e}_n^2(W^\pi)$ is a particular case of the Zador's theorem stated on Graf and Luschgy [2000]. Therefore, we can claim that the principal results on Sagna [2008] are valid in our framework.

Let us denote by $\beta^\pi = \{\beta_1, \cdots \beta_n\}$ an optimal n-quantizer for W^π and the maximal radius associated by

$$\rho_{n,\pi} = \max\{\|\beta\|_{\mathbb{R}^k}, \beta \in \beta^\pi\}.$$

Therefore we have

***Lemma* 6.3**
Let $(\beta_n^\pi)_{n \geq 1}$ the optimal sequence of n-quantizers for the distribution $\mathbb{P}$ of W^π, $(\rho_{n,\pi})_{n \geq 1}$ the optimal sequence of maximal radius. Let $e^{\|W^\pi\|_{\mathbb{R}^k}^2} \in \mathcal{L}^{0+}(\mathbb{P})$. Set

$$\theta^\star = \sup\{\theta > 0, \limsup_{x \to +\infty} e^{\theta x^2} \overline{F}_2(x) < +\infty\} = \sup\{\theta > 0, \mathbb{E} e^{\theta \|W^\pi\|_{\mathbb{R}^k}^2} < +\infty\},$$

and

$$\theta_\star = \inf\{\theta > 0, \liminf_{x \to +\infty} e^{\theta x^2} \mathbb{P}(\|W^\pi\|_{\mathbb{R}^k} > x) > 0\},$$

then $\theta^\star, \theta_\star \in (0, +\infty]$

$$\left(\frac{2+k}{k\theta_\star}\right)^{1/2} \leq \liminf_{n \to +\infty} \frac{\rho_{n,\pi}}{(\ln n)^{1/2}} \leq \limsup_{n \to +\infty} \frac{\rho_{n,\pi}}{(\ln n)^{1/2}} \leq 2 \left(\frac{2+k}{k\theta^\star}\right)^{1/2}.$$

The previous lemma is a particular case of Criterion 3.2.1 and 3.2.4 from Sagna [2008] with the Euclidean norm. In our case the proof follows in the same way.

The main theorem reads as follows

***Theorem* 6.6**
The maximal radius $\rho_{n,L}$ for the linear interpolation W_L using π defined by

$$\rho_{n,L} = \rho(\alpha_L) = \max\{\|a\|_{\mathcal{L}^2}, a \in \alpha_L\}, \qquad \forall\, n \in \mathbb{N},$$

satisfies that

$$\left(\frac{2+k}{k\theta_\star}\right)^{1/2} < \liminf_{n\to+\infty} \frac{\rho_{n,L}}{\sqrt{\ln n}} \leq \limsup_{n\to+\infty} \frac{\rho_{n,L}}{\sqrt{\ln n}} < 2\left(\frac{2+k}{k\theta^\star}\right)^{1/2}.$$

where $\theta^\star, \theta_\star \in (0, +\infty]$.

Proof

Let be $\alpha_L = (a_{1,L}, \cdots, a_{n,L})$ an optimal n-quantizer for W_L. By the isometry I it follows that

$$\rho_{n,L} = \rho_{n,\pi} \stackrel{\triangle}{=} \max\{\|\beta\|_{\mathbb{R}^k}, \beta \in \beta^\pi\}.$$

where $\rho_{n,\pi}$ is the maximal radius of the optimal n-quantizer β^π of a Gaussian vector in $\mathbb{R}^k$. Using Lemma 6.3 it holds

$$\left(\frac{2+k}{k\theta_\star}\right)^{1/2} < \liminf_{n\to+\infty} \frac{\rho_{n,L}}{\sqrt{\ln n}} \leq \limsup_{n\to+\infty} \frac{\rho_{n,L}}{\sqrt{\ln n}} < 2\left(\frac{2+k}{k\theta^\star}\right)^{1/2}.$$

With the expression above the proof is complete.

□

***Remark* 6.6**
It is worth pointing out that we need to find the exact values for both constants $\theta^\star, \theta_\star$ in Theorem 6.6 in order to achieve the correct asymptotic rate for the maximal radius of the optimal n-quantizer for W_L.

6.2.5 Asymptotics for the maximal radius of the Brownian motion

Here we will present some ideas about the maximal radius of the optimal n-quantizer of the Wiener process. We present two different approaches. However we did not reach conclusive results in any case.

Upper bound for the maximal radius

We have proof that if $\beta^\pi = \{\beta_1, \cdots \beta_n\}$ is an optimal n-quantizer for W^π then

$$\rho_{n,\pi} = \max\{\|\beta\|_{\mathbb{R}^k}, \beta \in \beta^\pi\},$$

satisfies

$$\left(\frac{2+k}{k\theta_\star}\right)^{1/2} \leq \liminf_{n\to+\infty} \frac{\rho_{n,\pi}}{(\ln n)^{1/2}} \leq \limsup_{n\to+\infty} \frac{\rho_{n,\pi}}{(\ln n)^{1/2}} \leq 2\left(\frac{2+k}{k\theta^\star}\right)^{1/2}.$$

The main idea was to use the previous equation and Remark 6.5. For the maximal radius for the optimal n-quantizer α of the Wiener process $\rho_n = \max\{\|a\|, a \in \alpha\}$ we define $(a^{(n)})_{n\geq 1}$ as the sequence which satisfies that $a^{(n)} = \arg\max \rho_n$. Hence it follows

$$\rho_n^2 = \|a^{(n)}\|_{\mathcal{L}^2}^2 = \int_0^1 \left(a^{(n)}(t)\right)^2 dt \leq \sup_{t\in[0,1]} \left(a^{(n)}(t)\right)^2,$$

Let us define the sequence $(t_n)_{n\geq 1}$ which satisfies that $t_n = \arg\max \left(a^{(n)}(t)\right)^2$ and $W_k^{(n)} = (W(t_1), \cdots, W(t_{k-1}), W(t_k))$, where $W(t_n) \in W_k^{(n)}$ for all n. If $\beta^{(n)} = \{\beta_1, \cdots \beta_n\}$ is an optimal n-quantizer for $W_k^{(n)}$ and $\rho_{(n)} = \max\{\|\beta\|_{\mathbb{R}^k}, \beta \in \beta^{(n)}\}$ then it follows that

$$\rho_n^2 = \|a^{(n)}\|_{\mathcal{L}^2}^2 \leq \left(a^{(n)}(t_n)\right)^2 = \left(\beta_j^{i_n}\right)^2 \leq C(t_n)\rho_{(n)}^2.$$

However this argument fails because for each n the value of t_n change and we are unable to control de behaviour of $C(t_n)$.

Lower bound for the maximal radius

Let us define the random variable $Z = \|W\|_{\mathcal{L}^2}$ in $\mathbb{R}$ with distribution $\mathbb{P}_Z$. No simple closed formula for this distribution is available. However it is obvious that it has an unbounded support and its tail has an exponential decay:

$$c_1 e^{-c_2 x^2} \leq P(\|W\|_{\mathcal{L}^2} \geq x) \leq c_3 e^{-c_4 x^2}.$$

The previous inequality could be easily checked. We know that $W(t) = \sum_{k\geq 1} \sqrt{\lambda_k}\xi_k\varphi_k(t)$ by the decomposition of Karhunen-Loève. Therefore using the known distribution of $|\xi_1|$ where $\xi_1 \sim N(0,1)$ and

$$\int_x^{+\infty} e^{-u^2/2} du \geq \frac{x}{x^2+1} e^{-x^2/2} \qquad \forall\, x \geq 0,$$

we have that

$$\begin{aligned} P(\|W\|_{\mathcal{L}^2} \geq x) &= P\left(\left(\sum_{i\geq 1} \lambda_i \xi_i^2\right)^{1/2} \geq x\right) \geq P\left(|\xi_1| \geq \frac{x}{\sqrt{\lambda_1}}\right) \\ &\geq \sqrt{\frac{2\lambda_1}{\pi}} \frac{x}{x^2+\lambda_1} e^{-x^2/2\lambda_1}. \end{aligned}$$

The knowing of the large deviation for $W_\infty = \max_{s\leq 1} |W(s)|$, see Dembo and Zeitouni [2010] allow us to conclude that

$$P(\|W\|_{\mathcal{L}^2} \geq x) \quad \leq \quad Ce^{-x^2/2}.$$

Therefore we obtain the decay rate of the distribution of Z.

Actually we do not need to know the closed formula for $\mathbb{P}_Z$. We only need the behaviour of the tail in order to use the general result proposed by Sagna. In other words if the distribution of a random variable X satisfies the classical assumptions stated in Sagna [2008] and its tail has an exponential decay then the maximal radius $\rho_n(X)$ of the optimal quantizer associated satisfies that

$$\rho_n(X) \sim \sqrt{\ln n}.$$

This general idea is the basis for the lower bound of the maximal radius of the optimal n-quantizer of the Wiener process.

Let $\rho_n = \max\{\|a\|, a \in \alpha\}$ by the maximal radius for the optimal n-quantizer α of the Wiener process W in $[0,1]$ with the usual norm $\mathcal{L}^2$ sur $[0,1]$. Let us define a sequence $(Z_n)_{n\geq 1}$ of random variables in $\mathbb{R}$ defined by

$$Z_n = \|W\|\mathbf{1}_{R_n}(W),$$

where $R_n = \{W : \|W\| \leq \rho_n\}$. For each n let us denote by $\mathbb{P}_Z^{(n)}$ its distribution law. Let us define by $\rho_m^{(n)}$ the maximal radius for the optimal m-quantizer $\alpha^{(n)}$ of Z_n. For $m = n$ let us denote

$$\rho_{(n)} = \rho_n^{(n)}.$$

Following the definition of Z_n it is obvious that

$$\rho_n \geq \rho_{(n)}.$$

It is straightforward that

$$Z_n \overset{a.s}{\to} Z, \qquad \text{when } n \to \infty,$$

and obviously this implies that

$$\mathbb{P}_Z^{(n)} \overset{\mathcal{D}}{\to} \mathbb{P}_Z.$$

If for any $n \in \mathbb{N}$, ρ_n^Z is the maximal radius for the unique optimal n-quantizer α^Z of Z it follows for any fixed m that $\rho_m^{(n)} \to \rho_m^Z$ when $n \to \infty$ and for every $\epsilon_m > 0$ there exists n_m^0 such that

$$\rho_m^{(n)} \in B(\rho_m^Z, \epsilon_m), \quad \text{for all } n \geq n_m^0.$$

Then we can write

$$\rho_m^{(n)} \geq (1 + \epsilon'_m)^{-1}\rho_m^Z.$$

Here we can not assume that exists $n_0 = \sup_{m\geq 1} n_m^0$, $n^0 < +\infty$ such as the previous equation holds for all $n \geq n^0$ and some $\epsilon > 0$ not depending on m.

General conclusions and perspectives

In this thesis we have worked in different directions. We have introduced a new stochastic algorithm for the search of optimal quantizers. We have proposed an estimation method for the Hurst parameter in some fractional process. We have extensively studied the asymptotic behaviour of the maximal radius associated with some quantizers of the Wiener process. We summarize in a brief exposition the most important results obtained. Finally we indicate some possible lines of future work concerning the subjects under discussion in this thesis.

The algorithm proposed in $\mathbb{R}^d$ is a modification of the classic method CLVQ. The *Average Competitive Learning Vector Quantization* offers similar results to those obtained by the CLVQ in the searching of the optimal quantizers in $\mathbb{R}^d$. Despite this new method does not improve the well-known results of the CLVQ, it is a variant that can be interesting from the point of view of application to practical problems in real time. The utilization of an improvement type Lloyd and the generation of set of random vectors instead one allow us to modified more than one quantizer in each iteration. The idea is simple however that it reduces the calculation time of the algorithm. In that sense we think that our proposal gains in interest and importance with respect to the CLVQ. We consider that other simulations studies are necessary for better knowing the behaviour of the calculation time and the link between the dimension and the number of random vectors that are generated in the *competitive phase*. A more complete study would help for a better characterization of the ACLVQ.

On the other hand in the theoretical aspect it was achieved the convergence of the solution of this new method to a stationary quantizer by the Kushner-Clark's theorem. A future work here could be stating some kind of Central Limit Theorem for the solution of the method similar to the results for the classic CLVQ.

The estimation of the Hurst parameter in fractional processes was carried out in the parametric framework. Despite the existence of different methods for that matter our proposal is simple and effective. Our method was based on the knowledge of the behaviour of the eigenvalues of the covariance operator for each process. In addition we avoid some of the numerical problems observed in the calculation of the maximum likelihood estimator, as well as the time of calculation that requires other classic methods.

This new procedure has the drawback that is only implemented for the estimation of the Hurst parameter. In some fractional processes like the Ornstein-Uhlenbeck there exists other parameters that are not considered by our methodology. In that direction it would be interesting to find different contrast functions that allow us to consider these parameters. The lack of knowledge of the asymptotic behaviour of the eigenvalues for some processes is a disadvantage in the theoretical aspect. In simulations it was possible

to observe that the procedure worked correctly even if the theoretical proof for all cases was not achieved.

The study of the asymptotic behaviour of the quantization error for stochastic processes with stretched covariance function $\Gamma(t) = e^{-\theta t^2}$ was another topic we deal with in this work. The utilization of the results proposed by Widom related with the behaviour of the spectral densidad of the process by the eigenvalues of it was crucial to determine the correct order. The general idea was based essentially on Shannon-Kolmogorov's ϵ-entropy and rapidly varying functions at infinity. The result obtained shows the rapid change on the convergence rate for this process in comparison with the fOU(3). This fact confirm the importance of the eigenvalues of the covariance operator to the asymptotics of the quantization error.

The approach to kriging using this function of covariance in the simulation study did not throw good results. The initial idea was that the quantizers could give additional information with respect to the kriging estimation. Our objective never was to replace the estimation by kriging of a curve or surface by an estimation based on quantization. Despite the large number of quantizers calculated for the cases studied (5000 each) we observed that the quantization theory does not seem to be a useful tool in this case. For further works others studies of simulation could be made.

The main results of the last chapter are based on the study of the rate of convergence of the maximal radius for the optimal quantizers of the Wiener process. In that direction we considered several approaches. First extrapolating the results in $\mathbb{R}^d$ due to Sagna for stochastic processes. This methodology despite considering the necessary modifications made was insufficient to obtain the desired results. The known properties of the Wiener measure were not sufficient to achieve our goal.

The second one was based on the construction of certain general sets of quantizers (n-optimal ε-quantizers). Our aim was the inference of asymptotic results in the mentioned set to obtain the results for the optimal quantizers. This approach was not completely satisfactory. The study of the behaviour of these sets although allowed to find certain interesting relations however did not produce useful information for our central subject. We think that a deeper study in these aspects based on the self-similar properties of the processes like the Wiener process possibly contribute to new routes of proofs of well-known asymptotics results.

The approach used for some stationary quantizers confirmed that the asymptotic rate of the maximal radius should be $\ln n$. The results obtained in this subset offer a general idea of the behaviour of the maximal radius.

Attending to the numerical method implemented to calculate the optimal quantizers of the Wiener using a linear interpolation we decide to use this approach to obtain asymptotic results for the maximal radius. In the theoretical framework the linear interpolation of the Wiener fulfills certain requirements that made possible to find the order of convergence for the maximal radius. The evident proximity of the linear optimal quantizers to the optimal ones of the Wiener process was another element that contributed information to the discussion.

In the case of the optimal quantizer of the Wiener process two ideas were used. The first one was for the upper limit related to the approach used for the linear interpolation of the Wiener process. For the lower bound the problems were of another nature: The first

approach was based on the search of a process with smaller norm than the Wiener has and the second one based on the distribution of unbounded support of the $\mathcal{L}^2$-norm the of the Wiener and its rate of decay. However we did not find conclusive results in any case. In this direction other approaches could be used with the aim to find these bounds. In addition in the own course of the research was observed that a possible relation between the maximal radius and the integral number d_n might be exist. It would be interesting for further investigation to go deeply into this topic. The obtention of the relation between both might help to find the real asymptotic behaviour of d_n.

Appendix A

Euclidean distance matrix. A new MATLAB procedure

We propose a new efficient method to compute the Euclidean distance matrix of two sets of vectors in dimension greater than 2. Let us define $Q = (Q_{i,j})_{1\leq i\leq d,\, 1\leq j\leq q}$ the matrix of quantizers and $\xi = (\xi_{i,k})_{1\leq i\leq d,\, 1\leq k\leq S}$ the matrix of random vector with $S \in \mathbb{N}$. Each column of both matrices is a vector of dimension d.

It is obvious that we need to calculate the norm of $q \cdot S$ vectors: $\|Q_{\cdot,j} - \xi_{\cdot,k}\|$ for all j and all k. Using a classical approach for large values of S a large amount of calculations is needed. We propose a method that do the work by using product and sum of matrices.

For example with $d = 2$ we have

$$\|Q_{\cdot,j} - \xi_{\cdot,k}\|^2 = \underbrace{Q_{1,j}^2 + Q_{2,j}^2}_{M^2 Q_{j,k}} \underbrace{-2\xi_{1,k}Q_{1,j} - 2\xi_{2,k}Q_{2,j}}_{M_{DP}Q\xi_{j,k}} + \underbrace{\xi_{1,k}^2 + \xi_{2,k}^2}_{M^2\xi_{j,k}}.$$

The main idea of our proposal is to construct three matrices: M^2Q, $M_{DP}Q\xi$, $M^2\xi$ with dimension $q \times S$. Let us define the structure of each one but first we denote the product $A \star A = A^\star$ as the product of each element of A by himself. If A has dimension $m \times n$ then $A^\star$ has dimension $m \times n$. Therefore

$$Q^\star = \begin{pmatrix} Q_{1,1}^2 & \cdots & Q_{1,q}^2 \\ \vdots & \vdots & \vdots \\ Q_{d,1}^2 & \cdots & Q_{d,q}^2 \end{pmatrix}_{d\times q} = Q \star Q.$$

After that we can construct:

$$Q_+^\star = \begin{pmatrix} \sum_{i=1}^d Q_{i,1}^2 & \cdots & \sum_{i=1}^d Q_{i,q}^2, \end{pmatrix}_{1\times q},$$

then

$$M^2Q = \begin{pmatrix} \sum_{i=1}^d Q_{i,1}^2 & \cdots & \sum_{i=1}^d Q_{i,1}^2 \\ \vdots & \vdots & \vdots \\ \sum_{i=1}^d Q_{i,q}^2 & \cdots & \sum_{i=1}^d Q_{i,q}^2 \end{pmatrix}_{q\times S} = {}^tQ_{+\,q\times 1}^\star \cdot (1, \cdots, 1)_{1\times S}.$$

The second matrix is defined by

$$M_{DP}Q\xi = \begin{pmatrix} -2\sum_{i=1}^{d} Q_{i,1}\xi_{i,1} & \cdots & -2\sum_{i=1}^{d} Q_{i,1}\xi_{i,S} \\ \vdots & \vdots & \vdots \\ -2\sum_{i=1}^{d} Q_{i,q}\xi_{i,1} & \cdots & -2\sum_{i=1}^{d} Q_{i,q}\xi_{i,S} \end{pmatrix}_{q\times S} = -2\,{}^{t}Q_{q\times d}\cdot\xi_{d\times S}.$$

For the last matrix $M^2\xi$ we do

$$\xi^{\star} = \begin{pmatrix} \xi_{1,1}^2 & \cdots & \xi_{1,S}^2 \\ \vdots & \vdots & \vdots \\ \xi_{d,1}^2 & \cdots & \xi_{d,S}^2 \end{pmatrix}_{d\times S} = \xi\star\xi,$$

as in the case of Q

$$\xi_{+}^{\star} = \begin{pmatrix} \sum_{i=1}^{d}\xi_{i,1}^2 & \cdots & \sum_{i=1}^{d}\xi_{i,S}^2 \end{pmatrix}_{1\times S},$$

then

$$M^2\xi = \begin{pmatrix} \sum_{i=1}^{d}\xi_{i,1}^2 & \cdots & \sum_{i=1}^{d}\xi_{i,S}^2 \\ \vdots & \vdots & \vdots \\ \sum_{i=1}^{d}\xi_{i,1}^2 & \cdots & \sum_{i=1}^{d}\xi_{i,S}^2 \end{pmatrix}_{q\times S} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_{q\times 1}\cdot\xi_{+1\times S}^{\star}.$$

Then we define a new matrix:

$$\begin{aligned} N &= M^2Q + M_{DP}Q\xi + M^2\xi \\ &= \begin{pmatrix} \sum_{i=1}^{d} Q_{i,1}^2 - 2Q_{i,1}\xi_{i,1} + \xi_{i,1}^2 & \cdots & \sum_{i=1}^{d} Q_{i,1}^2 - 2Q_{i,1}\xi_{i,S} + \xi_{i,S}^2 \\ \vdots & \vdots & \vdots \\ \sum_{i=1}^{d} Q_{i,q}^2 - 2Q_{i,q}\xi_{i,1} + \xi_{i,1}^2 & \cdots & \sum_{i=1}^{d} Q_{i,q}^2 - 2Q_{i,q}\xi_{i,S} + \xi_{i,S}^2 \end{pmatrix}_{N\times S}. \end{aligned}$$

If $N = (N_{j,k})_{N\times S}$ is easy to verify that

$$\begin{aligned} N_{j,k} &= \sum_{i=1}^{d} Q_{i,j}^2 - 2Q_{i,j}\xi_{i,k} + \xi_{i,k}^2 \\ &= \|Q_{\cdot,j} - \xi_{\cdot,k}\|^2. \end{aligned}$$

A.1 The code in MATLAB

Our method in MATLAB is programmed as follows:

```
% Method for calculate the norm of a set of vector
Quantizer=IniQ; (Q = (Q_{i,j})_{1≤i≤d, 1≤j≤N})
IniQ2=Quantizer.^2; (Q^⋆)
SquareQuant=sum(IniQ2,1)'*ones(1,SizeSample); (M^2 Q)
Double=-2*Quantizer'*sample; (M_{DP} Qξ)
SampleSquare=(sum(sample.^2,1)'*ones(1,N))'; (M^2 ξ)
% Square norm of the vectors
NormVectorT=SampleSquare+Double+SquareQuant; (N = M^2 Q + M_{DP} Qξ + M^2 ξ)
```

Appendix B

ACLVQ. Others numerical simulations

Number of Random Vectors	**Iterations**	**Quantization Error**	**Absolute Error**
20	100000	0,177116	0,001068
20	200000	0,177631	0,000553
20	500000	0,176718	0,001466
40	100000	0,177080	0,001104
40	200000	0,176975	0,001209
40	500000	0,177117	0,001067
60	100000	0,177159	0,001025
60	200000	0,176719	0,001465
60	500000	0,176596	0,001588
80	100000	0,176914	0,001270
80	200000	0,177247	0,000937
80	500000	0,176668	0,001516

Table B.1: Some results for the quantization error, for 20 quantizers.

Number of Random Vectors	**Iterations**	**Quantization Error**	**Absolute Error**
50	100000	0,075357	0,000141
50	200000	0,075268	0,000052
50	500000	0,075241	0,000025
100	100000	0,075053	0,000163
100	200000	0,075312	0,000096
100	500000	0,075143	0,000073
150	100000	0,075119	0,000097
150	200000	0,075208	0,000008
150	500000	0,075204	0,000012
200	100000	0,075179	0,000037
200	200000	0,074952	0,000263
200	500000	0,075195	0,000021

Table B.2: Some results for the quantization error, for 50 quantizers.

Appendix C

Convergence rate

***Lemma* 1** *The sequence R_n defined by*

$$R_n = \sum_{j\geq n+1} \lambda_j + n\lambda_n = \sum_{j\geq n+1} j^{-j} + n^{1-n},$$

converge and

$$R_n = \Omega(n^{1-n}) \tag{C.1}$$

$$R_n = O(n^{1-n}) \tag{C.2}$$

Proof

The convergence of R_n it is obvious. For (C.1) it is straightforward that the serie $\sum_{j\geq 1} j^{-j}$ is positive convergent serie hence

$$R_n = \sum_{j\geq n+1} j^{-j} + n^{1-n} \geq n^{1-n},$$

and that implies equation (C.1). Using (C.2) it follows that

$$\begin{aligned} R_n &= \sum_{j\geq n+1} \lambda_j + n\lambda_n = \sum_{j\geq n+1} j^{-j} + n^{1-n} \\ &\leq (n+1)^{-(n+1)} \sum_{j\geq n+1} \left(\frac{j}{n+1}\right)^{-j} + n^{1-n} = (n+1)^{-(n+1)} \sum_{i\geq 0} \left(\frac{n+1+i}{n+1}\right)^{-(n+1+i)} + n^{1-n}. \end{aligned}$$

It is obvious that $(n+1)^{-(n+1)} < n^{1-n}$ for all n. It is easy to check that the serie $\sum_{i\geq 0} \left(\frac{n+1+i}{n+1}\right)^{-(n+1+i)}$ is also convergent for all n and therefore, it follows that

$$\sum_{i\geq 0} \left(\frac{n+1+i}{n+1}\right)^{-(n+1+i)} \leq \sum_{i\geq 0} \left(\frac{2}{i+2}\right)^{i+2} = c_1 < +\infty,$$

then

$$R_n \leq (n+1)^{-(n+1)} \sum_{i\geq 0} \left(\frac{n+1+i}{n+1}\right)^{-(n+1+i)} + n^{1-n} \leq (1+c_1)n^{1-n} = c_2 n^{1-n}.$$

Equation (C.2) follows after and that complete the proof.

□

Appendix D

Convergence rate. Basic definitions

Definition* D.1 *Regularly varying functions
Let us set f a measurable positive function

- *The function f is called regularly varying at infinity with index $\delta \in \mathbb{R}$ if for every $t > 0$,*

$$\lim_{x\to\infty} \frac{f(tx)}{f(x)} = t^{\delta}.$$

When $\delta = 0$ the function f is called slowly varying at infinity.

- *The function f is called rapidly varying at infinity of index ∞ if for every $t > 0$,*

$$\lim_{x\to\infty} \frac{f(tx)}{f(x)} = +\infty, \text{ for } \delta > 1 \quad \text{and} \quad \lim_{x\to\infty} \frac{f(tx)}{f(x)} = 0, \text{ for } 0 < \delta < 1.$$

Definition* D.2 *Small ball probabilities
The small ball probabilities for a process X with distribution $\mathbb{P}$ is defined by

$$b_X(\varepsilon) = -\log \mathbb{P}(\|X\| \leq \varepsilon) \quad as \quad \varepsilon \to 0.$$

Definition* D.3 *Shannon-Kolmogorov's ϵ-entropy
If X is a random vector with probability measure $\mathbb{P}_X$ in H and $\widehat{X}$ is a random vector supported by a discret set and probability mesure $\mathbb{P}_{\widehat{X}}$, then

- *The entropy $\mathbb{H}(\widehat{X})$ for $\widehat{X}$ is*

$$\mathbb{H}(\widehat{X}) = \begin{cases} -\sum_{\widehat{x}} \log(\mathbb{P}_{\widehat{X}}(\widehat{x}))\mathbb{P}_{\widehat{X}}(\widehat{x}) & \text{if } \widehat{X} \text{ is a discret random vector} \\ \infty & \text{otherwise.} \end{cases}$$

- *The Shannon mutual information (or relative entropy) $I(\mathbb{P}_{X,\widehat{X}}, \mathbb{P}_X \otimes \mathbb{P}_{\widehat{X}})$ between X and $\widehat{X}$ is*

$$I(\mathbb{P}_{X,\widehat{X}}, \mathbb{P}_X \otimes \mathbb{P}_{\widehat{X}}) = \begin{cases} \int\limits_{H\times H} \log\left(\frac{d\mathbb{P}_{X,\widehat{X}}}{d\mathbb{P}_X \otimes \mathbb{P}_{\widehat{X}}}\right) d\mathbb{P}_{X,\widehat{X}} & \text{if } \mathbb{P}_{X,\widehat{X}} \ll \mathbb{P}_X \otimes \mathbb{P}_{\widehat{X}} \\ \infty & \text{otherwise.} \end{cases}$$

- *The Shannon-Kolmogorov's ϵ-entropy (or rate distortion function) denoted by $R_X(\epsilon)$ is continuous and decreasing in $\mathbb{R}_+$ and is defined by*

$$\begin{aligned} R_X(\epsilon) &= \inf\{I(\mathbb{P}_{X,Y}, \mathbb{P}_X \otimes \mathbb{P}_Y) : \mathbb{P}_{X,Y} \textit{ probability in } H \times H, \\ &\qquad \textit{with first marginal } \mathbb{P}_X \textit{ and } \int_{H\times H} \|x-y\|^2 d\mathbb{P}_{X,Y} \leq \epsilon\}. \end{aligned}$$

Appendix E

Concatenated quantizers

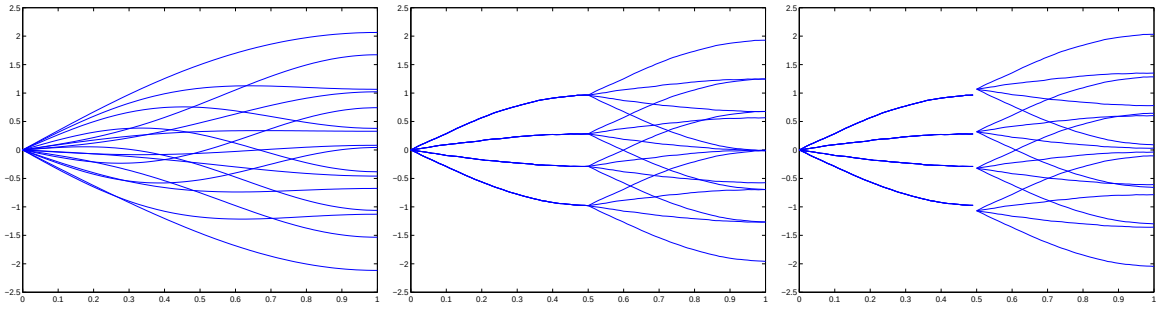

Figure E.1: *From left to right*: 16-optimal quantizer for the Brownian motion W, $16(4^2)$-quantizer $(\alpha^{(1)}_{\otimes k})$ and $16(4^2)$-quantizer $(\alpha^{(2)}_{\otimes k})$.

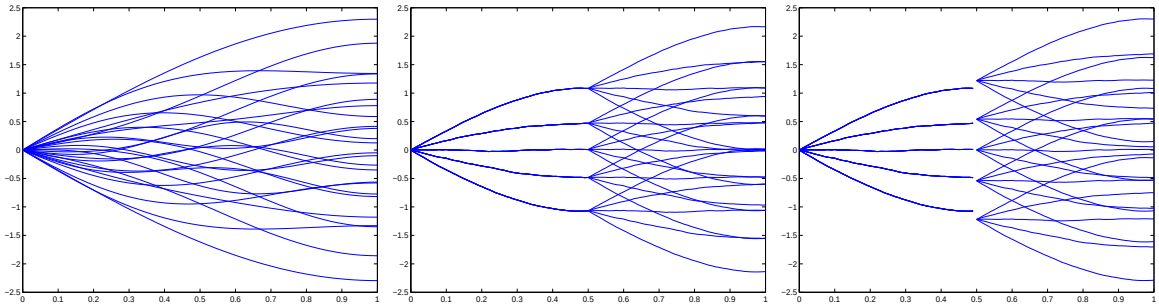

Figure E.2: *From left to right*: 25-optimal quantizer for the Brownian motion W $25(5^2)$-quantizer $(\alpha^{(1)}_{\otimes k})$ and $25(5^2)$-quantizer $(\alpha^{(2)}_{\otimes k})$.

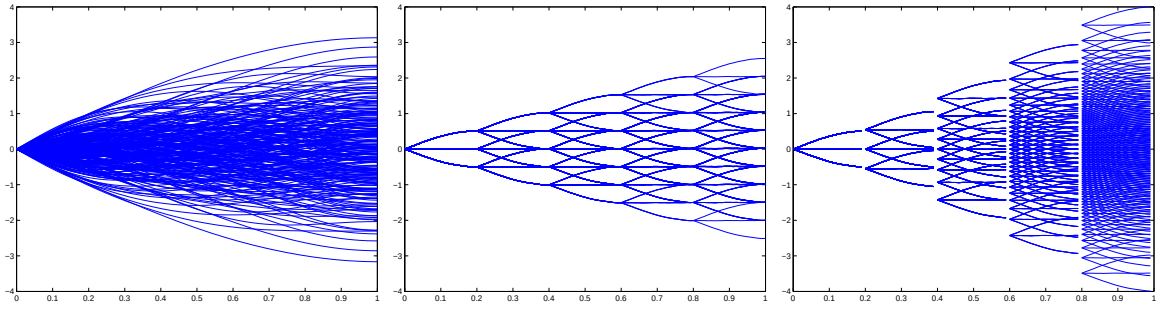

Figure E.3: *From left to right*: 243-optimal quantizer for the Brownian motion W, $243(3^5)$-quantizer $(\alpha^{(1)}_{\otimes k})$ and $243(3^5)$-quantizer $(\alpha^{(2)}_{\otimes k})$.

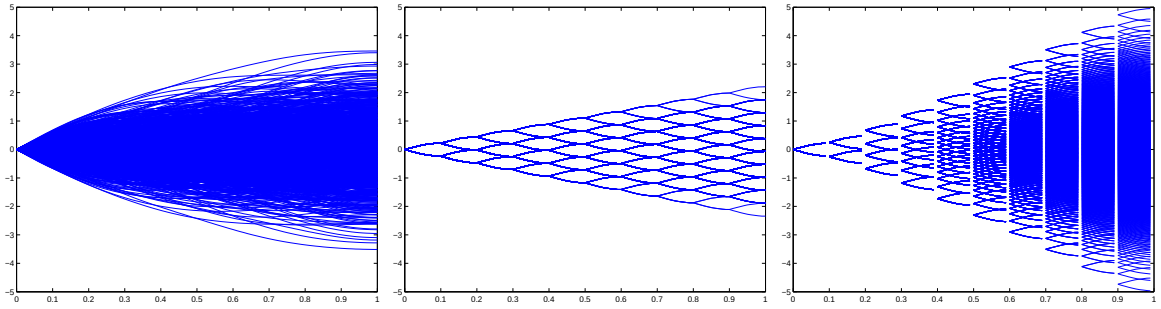

Figure E.4: *From left to right*: 1024-optimal quantizer for the Brownian motion W, $1024(2^{10})$-quantizer $(\alpha^{(1)}_{\otimes k})$ and $1024(2^{10})$-quantizer $(\alpha^{(2)}_{\otimes k})$.

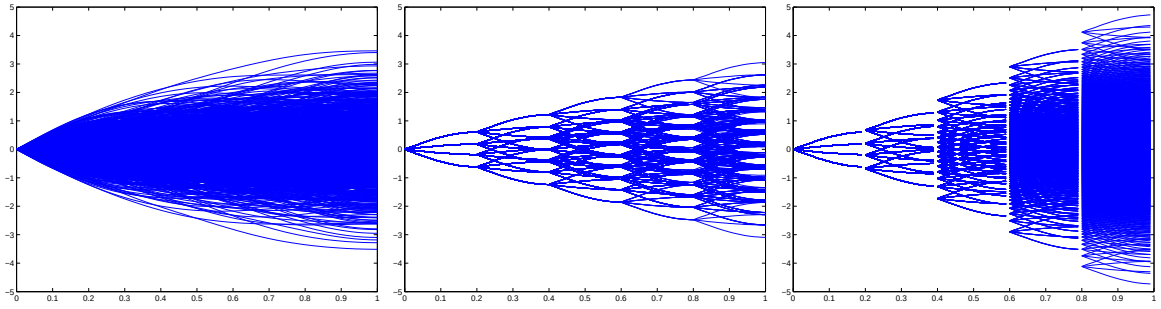

Figure E.5: *From left to right*: 1024-optimal quantizer for the Brownian motion W, $1024(4^5)$-quantizer $(\alpha^{(1)}_{\otimes k})$ and $1024(4^5)$-quantizer $(\alpha^{(2)}_{\otimes k})$.

Bibliography

Søren Asmussen and Peter W. Glynn. *Stochastic simulation: Algorithms and analysis.* Stochastic Modelling and Applied Probability 57. New York, NY: Springer. xiv, 476 p. EUR 49.95/net; SFR 87.00; $ 59.95 , 2007.

V. Bally, G. Pagès, and J. Printems. First-order schemes in the numerical quantization method. *Math. Finance*, 13(1):1–16, 2003.

Vlad Bally and Gilles Pagès. A quantization algorithm for solving multidimensional discrete-time optimal stopping problems. *Bernoulli*, 9(6):1003–1049, 2003.

Vlad Bally, Gilles Pagès, and Jacques Printems. A stochastic quantization method for nonlinear problems. *Monte Carlo Methods Appl.*, 7(1-2):21–33, 2001.

Vlad Bally, Gilles Pagès, and Jacques Printems. A quantization tree method for princing and hedging multidimensional american options. *Math. Finance*, 15(1):119–168, 2005.

Jean-Marc Bardet, Gabriel Lang, Georges Oppenheim, Anne Philippe, and Murad S. Taqqu. Generators of long-range dependent processes: a survey. In *Theory and applications of long-range dependence*, pages 579–623. Birkhäuser Boston, Boston, MA, 2003.

Francesca Biagini, Yaozhong Hu, Bernt Øksendal, and Tusheng Zhang. *Stochastic calculus for fractional Brownian motion and applications.* Probability and its Applications (New York). Springer-Verlag London Ltd., London, 2008a. ISBN 978-1-85233-996-8. doi: 10.1007/978-1-84628-797-8. URL `http://dx.doi.org/10.1007/978-1-84628-797-8`.

Francesca Biagini, Yaozhong Hu, Bernt Øksendal, and Tusheng Zhang. *Stochastic calculus for fractional Brownian motion and applications.* Probability and its Applications (New York). Springer-Verlag London Ltd., London, 2008b. ISBN 978-1-85233-996-8. doi: 10.1007/978-1-84628-797-8. URL `http://dx.doi.org/10.1007/978-1-84628-797-8`.

Vladimir I. Bogachev. *Gaussian measures. Transl. from the Russian by the author.* Mathematical Surveys and Monographs. 62. Providence, RI: American Mathematical Society (AMS). xii, 433 p. $ 95 , 1998.

Catherine Bouton and Gilles Pagès. About the multidimensional competitive learning vector quantization algorithm with constant gain. *Ann. Appl. Probab.*, 7(3):679–710, 1997.

Jared C. Bronski. Small ball constants and tight eigenvalue asymptotics for fractional Brownian motions. *J. Theoret. Probab.*, 16(1):87–100, 2003a. ISSN 0894-9840. doi: 10.1023/A:1022226420564. URL http://dx.doi.org/10.1023/A:1022226420564.

Jared C. Bronski. Asymptotics of Karhunen-Loeve eigenvalues and tight constants for probability distributions of passive scalar transport. *Comm. Math. Phys.*, 238(3):563–582, 2003b. ISSN 0010-3616. doi: 10.1007/s00220-003-0835-3. URL http://dx.doi.org/10.1007/s00220-003-0835-3.

Patrick Cheridito. Arbitrage in fractional Brownian motion models. *Finance Stoch.*, 7(4):533–553, 2003. ISSN 0949-2984. doi: 10.1007/s007800300101. URL https://docadis.ups-tlse.fr:443/http/dx.doi.org/10.1007/s007800300101.

Patrick Cheridito, Hideyuki Kawaguchi, and Makoto Maejima. Fractional Ornstein-Uhlenbeck processes. *Electron. J. Probab.*, 8:no. 3, 14 pp. (electronic), 2003. ISSN 1083-6489. URL http://www.math.washington.edu/~ejpecp/EjpVol8/paper3.abs.html.

Jean-François Coeurjolly. Estimating the parameters of a fractional Brownian motion by discrete variations of its sample paths. *Stat. Inference Stoch. Process.*, 4(2):199–227, 2001. doi: 10.1023/A:1017507306245.

Serge Cohen. Mouvements browniens fractionnaires et multifractionnaires. Available at http://images.math.cnrs.fr/Mouvements-browniens.html, 2004.

Sylvain Corlay and Gilles Pagès. Functional quantization based stratified sampling methods. Preprint available at http://arxiv.org/, 2010.

Noel A. C. Cressie. *Statistics for spatial data*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York, 1993. ISBN 0-471-00255-0. Revised reprint of the 1991 edition, A Wiley-Interscience Publication.

J.A. Cuesta and C. Matran. The strong law of large numbers for k-means and best possible nets of Banach valued random variables. *Probab. Theory Relat. Fields*, 78(4):523–534, 1988.

R. B. Davies and D. S. Harte. Tests for Hurst effect. *Biometrika.*, 74(1):95–101, 1987. doi: 10.1093/biomet/74.1.95.

Lokenath Debnath and Piotr Mikusiński. *Introduction to Hilbert spaces with applications. 3rd ed.* San Diego, CA: Academic Press. xx, 551 p. , 2005.

Sylvain Delattre, Jean-Claude Fort, and Gilles Pagès. Local distortion and μ-mass of the cells of one dimensional asymptotically optimal quantizers. *Commun. Stat., Theory Methods*, 33(5):1087–1117, 2004.

Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications. 2nd ed., corrected 2nd printing.* Stochastic Modelling and Applied Probability 38. Berlin: Springer. xvi, 396 p. EUR 42.75; SFR 62.00 , 2010.

S. Dereich and A. Lifshits. Probabilities of randomly centered small balls and quantization in Banach spaces. *Ann. Probab.*, 33(4):1397–1421, 2005.

S. Dereich and M. Scheutzow. High-resolution quantization and entropy coding for fractional Brownian motion. *Electronic Journal of Probability*, 2006.

S. Dereich, F. Fehringer, A. Matoussi, and M. Scheutzow. On the link between small ball probabilities and the quantization problem for Gaussian measures on Banach spaces. *J. Theor. Probab.*, 16(1):249–265, 2003.

Steffen Dereich. Small ball probabilities around random centers of Gaussian measures and applications to quantization. *J. Theor. Probab.*, 16(2):427–449, 2003.

Steffen Dereich. Asymptotic behavior of the distortion-rate function for Gaussian processes in Banach spaces. *Bull. Sci. Math.*, 129(10):791–803, 2005.

Steffen Dereich. The coding complexity of diffusion processes under $L^p[0,1]$-norm distortion. *Stochastic Processes Appl.*, 118(6):938–951, 2008. doi: 10.1016/j.spa.2007.07.002.

E. Diday and J.C. Simon. Clustering analysis. Digital Pattern Recogn., Commun. Cybern. 10, 47-94 (1976)., 1976.

A. B. Dieker. Simulation of fractional brownian motion. Master's thesis, Vrije Universiteit Amsterdam, 2002. Available in `http://www2.isye.gatech.edu/~adieker3/fbm/index.html`.

P. J. Diggle, J. A. Tawn, and R. A. Moyeed. Model-based geostatistics. *J. Roy. Statist. Soc. Ser. C*, 47(3):299–350, 1998. ISSN 0035-9254. doi: 10.1111/1467-9876.00113. URL `http://dx.doi.org/10.1111/1467-9876.00113`. With discussion and a reply by the authors.

J. L. Doob. The Brownian movement and stochastic equations. *Ann. of Math. (2)*, 43: 351–369, 1942. ISSN 0003-486X.

Paul Doukhan, George Oppenheim, and Murad S. Taqqu, editors. *Theory and applications of long-range dependence*. Birkhäuser Boston Inc., Boston, MA, 2003. ISBN 0-8176-4168-8.

Marie Duflo. *Algorithmes stochastiques*, volume 23 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin, 1996. ISBN 3-540-60699-8.

Marie Duflo. *Random iterative models. Transl. from the French by Stephen S. Wilson.* Applications of Mathematics. 34. Berlin: Springer. xv, 385 p. DM 128.00; öS 934.40; sFr. 113.00 , 1997.

Jean-Claude Fort and Gilles Pagès. Convergence of stochastic algorithms: From the Kushner-Clark theorem to the Lyapounov functional method. *Adv. Appl. Probab.*, 28 (4):1072–1094, 1996.

Jr. Fowler, J.E., M.R. Carbonara, and S.C. Ahalt. Image coding using differential vector quantization image coding using differential vector quantization. *Circuits and Systems for Video Technology, IEEE Transactions on*, 3(5):350–367, Oct 1993. ISSN 1051-8215. doi: 10.1109/76.246087.

Jr. Fowler, J.E., K.C. Adkins, S.B. Bibyk, and S.C. Ahalt. Real-time video compression using differential vector quantization. *Circuits and Systems for Video Technology, IEEE Transactions on*, 5(1):14–24, Feb 1995. ISSN 1051-8215. doi: 10.1109/76.350774.

Alan E. (ed.) Gelfand, Peter J. (ed.) Diggle, Montserrat (ed.) Fuentes, and Peter (ed.) Guttorp. *Handbook of spatial statistics.* Chapman & Hall/CRC Handbooks of Modern Statistical Methods. Boca Raton, FL: CRC Press. xii, 607 p. $ 99.95 , 2010.

A. Gersho. Principles of quantization. *Circuits and Systems, IEEE Transactions on*, 25 (7):427–436, Jul 1978. ISSN 0098-4094.

Allen Gersho and Robert M. Gray. *Vector quantization and signal compression.* Boston, MA: Kluwer Academic Publishers. xxii, 732 p., 1992.

Paul Glasserman. *Monte Carlo methods in financial engineering.* Applications of Mathematics 53. New York, NY: Springer. xiii, 596 p. EUR 69.95/net; $ 69.95; sFr 116.50 , 2004.

Tilmann Gneiting, Hana Sevcikova, and Donald B. Percival. Estimators of fractal dimension: Assessing the roughness of time series and spatial data. Preprint available at `http://arxiv.org/`, 2011.

Israel Gohberg and Seymour Goldberg. *Basic operator theory.* Boston - Basel - Stuttgart: Birkhäuser. XIII, 285 p. SFr. 36.00 , 1981.

Siegfried Graf and Harald Luschgy. *Foundations of Quantization for Probability Distributions.* Lectures Notes in Mathematics. Springer, 2000.

Siegfried Graf, Harald Luschgy, and Gilles Pagès. Functional quantization and small ball probabilities for Gaussian processes. *J. Theor. Probab.*, 16(4):1047–1062, 2003.

E.P. Høg and P.H. Frederiksen. The fractional ornstein- uhlenbeck process: term structure theory and application. Scnadinavian Working Papers in Business Administration, No. F-2006-01., 2006.

J. Hüsler, V. Piterbarg, and O. Seleznjev. On convergence of the uniform norms for Gaussian processes and linear approximation problems. *Ann. Appl. Probab.*, 13(4): 1615–1653, 2003. ISSN 1050-5164. doi: 10.1214/aoap/1069786514. URL `http://dx.doi.org/10.1214/aoap/1069786514`.

Svante Janson. *Gaussian Hilbert spaces. Reprint of the 1997 hardback ed.* Cambridge Tracts in Mathematics 129. Cambridge: Cambridge University Press. x, 340 p., 2008.

Rachid Jennane, Rachid Harba, and Gérard Jacquet. Analysis methods for fractional Brownian motion: Theory and comparative results. (Méthodes d'analyse du mouvement brownien fractionnaire: Théorie et résultats comparatifs.). *Trait. Signal*, 18(5-6):419–436, 2001.

Terhi Kaarakka and Paavo Salminen. On fractional ornstein-uhlenbeck processes. Preprint available at `http://arxiv.org/`, 2007.

Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus. 2nd ed.* Graduate Texts in Mathematics, 113. New York etc.: Springer-Verlag. xxiii, 470 p. DM 68.00/pbk , 1991.

J.C Kieffer. Uniqueness of locally optimal quantizer for log-concave density and convex error weighting functions. *IEEE Trans. Inf. Theory*, 29:42–47, 1983.

A. N. Kolmogorov. Wienersche Spiralen und einige andere interessante Kurven im Hilbertschen Raum. *C. R. (Doklady) Acad. Sci. URSS (N.S.)*, 26:115–118, 1940.

A.K. Krishnamurthy, S.C. Ahalt, D.E. Melton, and P. Chen. Neural networks for vector quantization of speech and images. *Selected Areas in Communications, IEEE Journal on*, 8(8):1449–1457, Oct 1990. ISSN 0733-8716. doi: 10.1109/49.62823.

Harold J. Kushner and Dean S. Clark. *Stochastic approximation methods for constrained and unconstrained systems*, volume 26 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1978. ISBN 0-387-90341-0.

S.C. Lim and S.V. Muniandy. Generalized Ornstein-Uhlenbeck processes and associated self-similar processes. *J. Phys. A, Math. Gen.*, 36(14):3961–3982, 2003. doi: 10.1088/0305-4470/36/14/303.

Harald Luschgy and Gilles Pagès. Functional quantization of Gaussian processes. *J. Funct. Anal.*, 196(2):486–531, 2002.

Harald Luschgy and Gilles Pagès. Sharp asymptotics of the functional quantization problem for Gaussian processes. *Ann. Probab.*, 32(2):1574–1599, 2004a.

Harald Luschgy and Gilles Pagès. Sharp asymptotics of the Kolmogorov entropy for Gaussian measures. *J. Funct. Anal.*, 212(1):89–120, 2004b.

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

Harald Luschgy, Gilles Pagès, and Benedikt Wilbertz. Asymptotically optimal quantization schemes for Gaussian processes on Hilbert spaces. *ESAIM Probab. Stat.*, 14: 93–116, 2010. ISSN 1292-8100. doi: 10.1051/ps:2008026. URL `http://dx.doi.org/10.1051/ps:2008026`.

Elias Masry, Bede Liu, and Kenneth Steiglitz. Series expansion of wide-sense stationary random processes. *IEEE Trans. Inf. Theory*, 14:792–796, 1968. doi: 10.1109/TIT.1968.1054230.

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

Gilles Pagès and Huyên Pham. Optimal quantization methods for nonlinear filtering with discrete-time observations. *Bernoulli*, 11(5):893–932, 2005.

Gilles Pagés and Jacques Printems. Optimal quadratic quantization for numerics: the Gaussian case. *Monte Carlo Methods Appl.*, 9(2):135–165, 2003.

Gilles Pagès and Jacques Printems. Functional quantization for numerics with an application to option pricing. *Monte Carlo Methods Appl.*, 11(4):407–446, 2005.

Gilles Pagès, Huyên Pham, and Jacques Printems. Optimal quantization methods and applications to numerical problems in finance. Rachev, Svetlozar T. (ed.), Handbook of computational and numerical methods in finance. Boston, MA: Birkhäuser. 253-297., 2004.

M. Rosenblatt. Some results on the asymptotic behavior of eigenvalues for a class of integral equations with translation kernels. *J. Math. Mech.*, 12:619–628, 1963.

Abass Sagna. *Méthodes de quantification optimale avec applications à la Finance.* PhD thesis, L'Université Pierre et Marie Curie, 2008.

Gennady Samorodnitsky and Murad S. Taqqu. *Stable non-Gaussian random processes: stochastic models with infinite variance.* Stochastic Modeling. New York, NY: Chapman & Hall. xviii, 632 p., 1994.

Oleg Seleznjev. Large deviations in the piecewise linear approximation of Gaussian processes with stationary increments. *Adv. in Appl. Probab.*, 28(2):481–499, 1996. ISSN 0001-8678. doi: 10.2307/1428068. URL `http://dx.doi.org/10.2307/1428068`.

Yingcai Su and Stamatis Cambanis. Sampling designs for estimation of a random process. *Stochastic Process. Appl.*, 46(1):47–89, 1993. ISSN 0304-4149. doi: 10.1016/0304-4149(93)90085-I. URL `http://dx.doi.org/10.1016/0304-4149(93)90085-I`.

T Tarpey. Self-consistency: a fundamental concept in statistics. *Statist. Sci.*, 11:229–243, 1996.

T. Tarpey and K.K.J. Kinateder. Clustering functional data. *J. Classification.*, 20:93–114, 2003.

Julius T. Tou and Rafael C. Gonzalez. *Pattern recognition principles.* Applied Mathematics and Computation. 7. Reading, Mass. etc.: Addison- Wesley Publishing Company, Advances Book Program. XXII, pp.377, 1974.

Alexander V. Trushkin. Sufficient conditions for uniqueness of a locally optimal quantizer for a class of convex error weighting functions. *IEEE Trans. Inf. Theory*, 28(2):187–198, 1982.

Joachim Weidmann. *Linear operators in Hilbert spaces. Transl. by Joseph Szücs.* Graduate Texts in Mathematics, Vol. 68. New York - Heidelberg -Berlin: Springer-Verlag. XIII, 402 p. DM 68.00; $ 40.20 , 1980.

Harold Widom. Asymptotic behavior of the eigenvalues of certain integral equations. II. *Arch. Rational Mech. Anal.*, 17:215–229, 1964. ISSN 0003-9527.

B. Wilbertz. Functional quantization of gaussian measures and application to the brownian motion on $l^2([0;t];dt)$ from a numerical point of view. *Diploma thesis*, Universität Trier (Germany), 2005.

Litan Yan, Yunsheng Lu, and Zhiqiang Xu. Some properties of the fractional Ornstein-Uhlenbeck process. *J. Phys. A*, 41(14):145007, 17, 2008. ISSN 1751-8113. doi: 10.1088/1751-8113/41/14/145007. URL `http://dx.doi.org/10.1088/1751-8113/41/14/145007`.

Xu Yong, Yan Guangqun, Chen Hexin, and Dai Yisong. A new competitive learning algorithm for vector quantization based on the neuron winning probability. In *Intelligent Processing Systems, 1997. ICIPS '97. 1997 IEEE International Conference on*, volume 1, pages 485–488 vol.1, Oct 1997. doi: 10.1109/ICIPS.1997.672829.

Paul L. Zador. Asymptotic quantization error of continuous signals and the quantization dimension. *IEEE Trans. Inf. Theory*, 28:139–149, 1982. doi: 10.1109/TIT.1982.1056490.