Méthodes de quantification optimale pour le filtrage et applications à la finance

Afef Sellami

Université Paris Dauphine
D.F.R. Mathématiques de la décision

THÈSE
pour l'obtention du titre de
Docteur en Sciences
(Arrêté du 25 avril 2002)
Spécialité : Mathématiques Appliquées

Présentée par
Afef SELLAMI

12 Décembre 2005

Méthodes de quantification optimale pour le filtrage et applications à la finance

JURY

Directeurs de thèse :	**Monsieur Gilles PAGÈS** Professeur à l'Université Paris VI **Monsieur Huyên PHAM** Professeur à l'Université Paris VII
Rapporteurs :	**Madame Valentine GENON-CATALOT** Professeur à l'Université Paris V **Monsieur François LE GLAND** Directeur de recherche à l'IRISA-INRIA Rennes
Suffragants :	**Monsieur Vlad BALLY** Professeur à l'Université Marne-la-Vallée **Monsieur Christian ROBERT** Professeur à l'Université Paris Dauphine
Invitée :	**Madame Agnès SULEM** Directrice de recherche à l'INRIA Rocquencourt

Remerciements

Mes premiers remerciements vont à mes parents et à mes soeurs pour tout leur amour et leur soutien. Au delà des distances, ils ont su me communiquer leur énergie et me redonner la confiance qui me faisait défaut dans les moments de doute. Ce travail leur est dédié avec toute ma gratitude.

Cette thèse n'aurait jamais vu le jour sans le précieux encadrement de mes directeurs de thèse M. Gilles Pagès et M. Huyên Pham qui ont toujours su être disponibles et à l'écoute de mes interrogations. Grâce à leur implication, leur patience et leurs conseils ce travail a pu être mené à bout, je leur en serai à jamais redevable.

Je remercie aussi Mme. Valentine Genon-Catalot et M. François Le Gland d'avoir accepté de rapporter ce travail. Les travaux de M. Vlad Bally ont beaucoup contribué dans l'orientation de mes recherches. Je suis aujourd'hui honorée par sa présence parmi les membres du jury et je lui en suis très reconnaissante. M. Christian Robert, Mme Agnès Sulem et M. Nizar Touzi ont été initiateurs de cette thèse par leurs cours du DEA MASE, que j'ai suivis avec beaucoup d'intérêt, je suis heureuse qu'ils aient accepté d'être dans mon jury et je les en remercie sincèrement.

Enfin, merci aux thésards du bureau 5C9 à Chevaleret pour cette belle ambiance et tous les bons moments partagés autour d'un goûter ou d'une discussion passionnée. Merci à Stéphane Gaiffas pour ses initiations au C++ . Un grand merci également à M. Jacques Portès qui a toujours pu trouver le temps de résoudre mes problèmes informatiques, malgré un emploi du temps très chargé.

Je ne saurai oublier de remercier tous mes amis qui m'ont épaulée durant ces dernières années. Ma pensée va à Marouen, qui a toujours été là pour moi, à Alia, Nessrine, Noura et Walid dont l'amitié m'est très chère, à tous les autres que je ne saurai nommer, qu'ils trouvent dans ces lignes le témoignage de ma profonde gratitude.

Table des matières

Introduction

0.1 Quantification optimale et applications

La quantification est une méthode issue du traitement du signal et de l'information. Elle consiste à approcher un signal à valeurs dans un espace continu par un signal à valeur dans un espace discret. A l'origine, cette technique a été motivée par des raisons pratiques d'économie de transmission [20], et a vu ensuite ses applications s'élargir à différents domaines, notamment, depuis quelques années, les probabilités numériques. Dans cette section nous allons adopter une présentation résolument probabiliste de la quantification en rappelant quelques resultats théoriques utiles et en mettant en évidence ses applications les plus récentes.

0.1.1 Définitions et résultats préliminaires

0.1.1.1 Quantification de variables aléatoires

On se place dans un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, et on se donne une variable aléatoire à valeurs dans $\mathbb{R}^d$, X, de loi $\mathbb{P}_X$ supposée simulable. Un entier $N \in \mathbb{N}^*$ étant fixé, un N-quantifieur est une application borélienne h_N appliquant $\mathbb{R}^d$ dans un ensemble fini $\Gamma = \{x^1, \ldots, x^N\} \subset \mathbb{R}^d$.

Pour définir de manière unique l'application h_N, on a besoin de spécifier en plus une partition borélienne $(A^i)_{1 \leq i \leq N}$ de l'espace $\mathbb{R}^d$ pour avoir :

$$h_N(X) = \sum_{i=1}^{N} x^i \mathbf{1}_{A^i}(X).$$

Le N-quantifieur est donc spécifié par la donnée de :

- $\Gamma = \{x^1, \ldots, x^N\}$ la grille de quantification de taille N appelée aussi ensemble des centres, des points de quantification ou des centroïdes, ou encore N-quantifieur associé à h_N.
- Une partition borélienne $(A^i)_{1 \leq i \leq N}$ de l'espace $\mathbb{R}^d$. A chaque ensemble A^i sera associé un centre x^i que l'on supposera toujours appartenir à A^i.

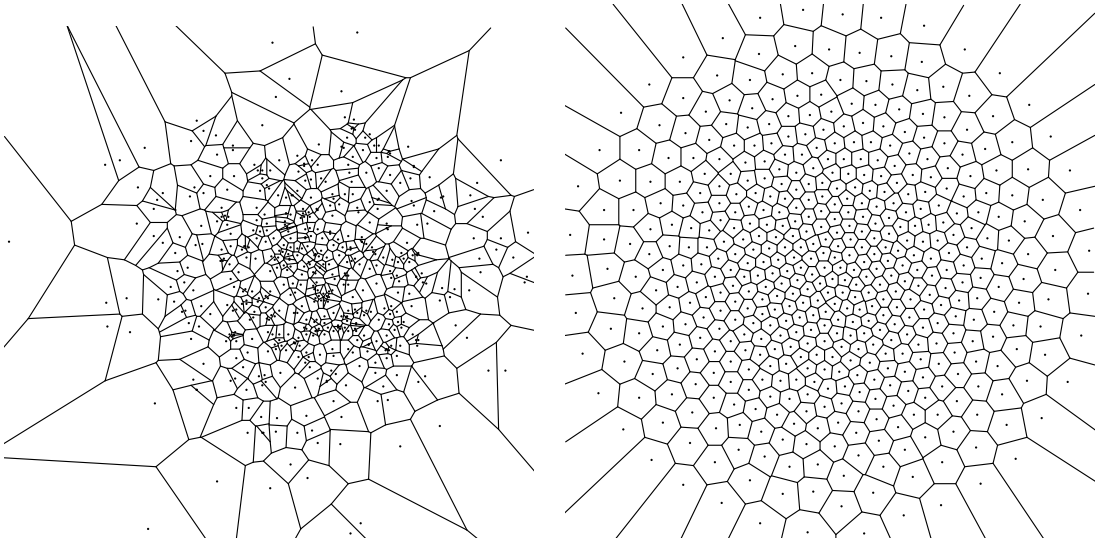

FIG. 1 – Partitions de l'espace associées à une grille de quantification en deux dimensions

Quand $X \in L^p$, on définit un N-quantifieur L^p-optimal de X par l'application h_N^* solution du problème d'optimisation paramétré par la taille de la grille de quantification N :

$$\inf\{\|X - h(X)\|_p^p, h : \mathbb{R}^d \to \mathbb{R}^d, \text{ application borelienne t.q. } |h(\mathbb{R}^d)| \leq N\}. \qquad (0.1.1)$$

D'après les résultats établis par exemple dans [25], ce problème admet une solution h_N^* définie par un N-quantifieur $\Gamma_N^* = \{x^1, \ldots, x^N\}$, vérifiant :

$$\mathbb{E}|X - h_N^*(X)|^p = \mathbb{E} \min_{x \in h_N^*(\mathbb{R}^d)} |X - x|^p,$$

et par la partition $(\mathbf{C}_i(\Gamma^*))_{1 \leq i \leq N}$, dite de Voronoï, associée à cet ensemble, définissant h_N^* comme une projection au plus proche voisin sur l'ensemble des centres $(x^i)_{1 \leq i \leq N}$ (voir Figure 1). Soit :

$$\mathbf{C}_i(\Gamma^*) \subset \{\xi \in \mathbb{R}^d \; t.q. \; |\xi - x^i| = \min_{1 \leq k \leq N} |\xi - x^k|\}.$$

La distorsion[1] s'écrit alors :

$$\underline{\mathcal{D}}_N^{X,p} := \|X - h^*(X)\|_p^p = \| \min_{1 \leq i \leq N} |X - x^i| \|_p^p.$$

Elle converge vers zero quand la taille du quantifieur N tend vers $+\infty$, et sa vitesse de convergence est régie par le théorème de Zador énoncé comme suit :

[1] On notera par ailleurs l'erreur de quantification $\|X - h_N^*(X)\|_p$.

Théorème 0.1.1 (Cf. [25, 4]) *On suppose que* $\int_{\mathbb{R}^d} |\xi|^{p+r} \mathbb{P}_X(d\xi) < +\infty$ *pour* $r > 0$. *Alors,*

$$\lim_N (N^{\frac{p}{d}} \underline{\mathcal{D}}_N^{X,p}) = J_{p,d} \|\varphi\|_{\frac{d}{d+p}},$$

où $\mathbb{P}_X(d\xi) = \varphi(\xi)\lambda_d(d\xi) + \bar{\mu}(d\xi)$, $\bar{\mu} \perp \lambda_d$ (λ_d *mesure de Lebesgue sur* $\mathbb{R}^d$) *et pour tout* $q \in \mathbb{R}_+^*$, $\|g\|_q := (\int |g|^q(u)du)^{\frac{1}{q}}$.

Remarque 0.1.1 $J_{p,d}$ correspond à la limite pour la loi uniforme sur $[0,1]$. On sait que $J_{p,1} = \frac{1}{2^p(p+1)}$ et que $J_{2,2} = \frac{5}{18\sqrt{3}}$. D'une manière générale, on ne connaît pas la valeur de cette constante pour $d > 2$, mais on a $J_{2,d} \sim \frac{d}{2\pi e}$. (Cf. [25]).

Nous pouvons ainsi écrire que $\|X - h_N^*(X)\|_p = O(N^{-\frac{1}{d}})$ au voisinage de $N \to +\infty$.
Par ailleurs, il sera essentiel de noter que les quantifieurs L^2-optimaux vérifient une propriété dite de stationnarité, à savoir que :

$$\mathbb{E}[X|h_N^*(X)] = h_N^*(X). \tag{0.1.2}$$

Cette propriété permet d'utiliser des termes correcteurs de premier ordre dans les différentes applications de la quantification, comme nous allons le voir dans le paragraphe suivant pour l'intégration numérique puis, plus loin, dans les exemples d'évaluation d'options américaines [6] ou de filtrage (Cf. Chapitre 1).

D'un point de vue pratique, définir la fonction h_N^* pour une taille de quantifieur fixée N s'avère être un problème d'optimisation assez délicat à résoudre notamment dans le cas multidimensionnel. En effet, pour plusieurs lois en dimensions 1, il existe des solutions analytiques fermées ou semi-fermées qui peuvent aisément être calculées. On citera par exemple la quantification de la loi normale par la méthode de Newton [44], ou de la loi exponentielle, entre autres, dans [17]. Par ailleurs, dans le cas particulier d'une loi à densité log-concave, la solution au problème (0.1.1) est unique [28], ce qui permet l'utilisation efficace d'algorithmes numériques d'optimisation tels que l'algorithme du gradient ou celui du point fixe appelé aussi méthode de Lloyd I [44, 2].
Ces méthodes déterministes deviennent rapidement difficiles à mettre en œuvre en dimensions supérieures à 1. D'une part parce qu'elles impliquent des calculs d'intégrales trop complexes au delà de la dimension 1, d'autres part, parce que la solution de (0.1.1) n'est pas unique ce qui diminue encore l'efficacité des résolutions numériques. Dans ce cas alors, on préfère utiliser des algorithmes stochastiques d'apprentissage fondés sur la simulation. On citera dans ce contexte l'algorithme *Competitive Learning Vector Quantization* (CLVQ algorithm) aussi appelé algorithme de Kohonen à zéro voisins, les algorithmes génétiques [26] ou l'algorithme de Lloyd I multidimensionnel [2]. Une étude détaillée des méthodes de quantification quadratique optimale est proposée dans [44]. Il est à noter que c'est le cas quadratique ($p = 2$) qui est le plus souvent étudié sur le plan numérique, même si les résultats théoriques sont généralement énoncés pour p quelconque.

Nous ne détaillerons pas plus avant cet aspect de la quantification car il n'est pas au centre de notre travail. Nous nous intéresserons plutôt à l'exploitation de ces quantifieurs optimaux pour différentes applications pratiques. A cet effet, il est important d'introduire tout d'abord une définition de processus quantifié (ou quantification de processus différente de la notion de quantification fonctionnelle de processus. Cf. [33, 45]). Ce sera l'objet du prochain paragraphe.

0.1.1.2 Quantification de processus

Dans les différentes applications de la quantification, il est souvent requis de considérer la quantification d'un processus markovien à temps discret $(X_k)_{k\geq 0}$ dont on connait la dynamique d'évolution de manière à pouvoir en simuler la trajectoire. Une approche possible dans ce cas est de quantifier chaque variable X_k en tenant compte de sa loi marginale, on parle donc de *quantification marginale*. Pour cela, on doit se fixer une taille de grille N_k à chaque pas de temps et un N_k-quantifieur L^p-optimal de $X_k \in L^p$ que l'on notera $\Gamma_k = \{x_k^1, \ldots, x_k^{N_k}\}$.
Par conséquent, on définit la quantification (Voronoï) de X_k par :

$$\hat{X}_k = h^*_{N_k}(X_k) = \sum_{i=1}^{N_k} x_k^i \mathbf{1}_{\mathbf{C}_i(\Gamma_k)}(X_k). \tag{0.1.3}$$

Le processus ainsi quantifié $(\hat{X}_k)_{k\geq 0}$ ne vérifie plus la propriété de Markov. Cependant, une approximation de la probabilité de transition entre différents états à deux dates successives reste possible à travers les paramètres compagnons p_k^{ij}, pour $i \in \{1, \ldots, N_k\}$ et $j \in \{1, \ldots, N_{k+1}\}$ définis par :

$$\begin{aligned} p_k^{ij} &= \mathbb{P}[X_{k+1} \in \mathbf{C}_j(\Gamma_{k+1}) | X_k \in \mathbf{C}_i(\Gamma_k)] \\ &= \mathbb{P}[\hat{X}_{k+1} = x_{k+1}^j | \hat{X}_k = x_k^i]. \end{aligned}$$

D'une manière générale, pour $0 \leq k < n$, $i \in \{1, \ldots, N_k\}$ et f borélienne dans R^d, on notera :

$$\widehat{\mathbf{P}}_k f(x_k^i) = \mathbb{E}[f(\hat{X}_{k+1}) | \hat{X}_k = x_k^i] = \sum_{j=1}^{N_{k+1}} f(x_{k+1}^j) p_k^{ij}.$$

Pour des horizons n raisonnables (≤ 100), il est donc possible de calculer et de stocker dans des tables facilement accessibles les grilles de quantifications et les paramètres compagnons i.e. les (x_k^i) et les (p_k^{ij}) pour $0 \leq k \leq n$, $1 \leq i \leq N_k$ et $1 \leq j \leq N_{k+1}$. Ce prétraitement des données, dit *off-line*, permet de minimiser les calculs d'éventuels estimateurs utilisant la quantification. Un exemple de ces grilles pour la loi normale centrée réduite est diponible et téléchargeable à partir de `http ://www.proba.jussieu.fr/pageperso/pages /quant3.html`.

0.1.2 Application à l'intégration numérique

Une application immédiate de la quantification est le calcul d'approximations numériques d'intégrales par rapport à une mesure donnée. On se pose le problème de l'évaluation de l'intégrale $\mathbb{E}[f(X)]$, pour X de loi $\mathbb{P}_X$ sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Si $\hat{X}$ désigne une N-quantification L^2-optimale de X nous pouvons nous donner comme estimateur : $\mathbb{E}[f(\hat{X})]$. Comme $\hat{X}$ est une variable aléatoire discrète, le calcul de cet estimateur se résumera à une somme pondérée finie, dont les termes sont lus à partir de tables pré-calculées. En reprenant les notations du paragraphe précédent, on pose :

$$\mathbb{E}[f(X)] \approx \mathbb{E}[f(\hat{X})] = \sum_{i=1}^{N} x^i \int \mathbf{1}_{\mathbf{C}_i(\Gamma)}(x)\mathbb{P}_X(dx) = \sum_{i=1}^{N} x^i \hat{\mathbf{p}}^i,$$

où $\hat{\mathbf{p}}^i = \mathbb{P}_X(\mathbf{C}_i(\Gamma)) = \mathbb{P}(X \in \mathbf{C}_i(\Gamma))$.
Les pondérations $\mathbf{p}^i$ sont aussi des paramètres compagnons qui peuvent être calculés en même temps que la grille de quantification Γ et stockés dans des tables accessibles pendant l'estimation. L'erreur d'estimation est contrôlée par l'erreur de quantification $\Delta = X - \hat{X}$. En effet, quand f est continue, dérivable à dérivée bornée, il existe $\xi \in (X, \hat{X})$ tel que :

$$f(X) - f(\hat{X}) = \langle \mathrm{D}f(\xi), \Delta \rangle,$$

où $< ., . >$ désigne le produit scalaire canonique sur $\mathbb{R}^d$.
Ceci donne la majoration d'erreur qu'on appellera *d'ordre zero* :

$$|\mathbb{E}[f(X)] - \mathbb{E}[f(\hat{X})]| \leq C\|\Delta\|_1 \leq C\|\Delta\|_2. \tag{0.1.4}$$

Quand f est continue, 2 fois dérivable, à dérivée seconde bornée, on peut développer f à un ordre supérieur afin d'établir une majoration d'erreur *d'ordre 1*. En effet, il existe $\xi \in (X, \hat{X})$ tel que :

$$f(X) - f(\hat{X}) = \langle \mathrm{D}f(\hat{X}), \Delta \rangle + \frac{1}{2}\Delta' \mathrm{D}^2 f(\xi)\Delta.$$

Ainsi, comme $\hat{X}$ vérifie la propriété de stationnarité (0.1.2), il est possible d'établir :

$$\begin{aligned} |\mathbb{E}[f(X)] - \mathbb{E}[f(\hat{X})]| &\leq \mathbb{E}|\mathbb{E}[f(X) - f(\hat{X})|\hat{X}]| \\ &\leq C\mathbb{E}|\langle \Delta, \Delta \rangle| \leq C\|\Delta\|_2^2. \end{aligned} \tag{0.1.5}$$

En utilisant le théorème de Zador 0.1.1, on obtient un taux de convergence en $O(N^{\frac{-1}{d}})$ dans le cas de l'inégalité (0.1.4), et moyennant l'hypothèse plus restrictive sur la fonction f, on a un taux de convergence deux fois plus rapide $O(N^{\frac{-2}{d}})$ à partir de (0.1.5).

L'intégration numérique par quantification est une méthode qui s'approche dans son principe des méthodes de Monte Carlo [49] : elle s'appuie sur une représentation de la loi

de X par un ensemble discret fini pondéré. L'estimateur de Monte Carlo s'écrit en effet comme la somme équipondérée sur un échantillon iid de taille M :

$$\mathbb{E}[f(X)] \approx \frac{1}{M}\sum_{i=1}^{M} f(X^i) \quad \text{où} \quad (X^1,\dots,X^M) \qquad iid \qquad X^1 \sim \mathbb{P}_X.$$

Mais si le principe reste le même, de grandes différences séparent les deux méthodes :

- Les grilles de quantification ainsi que les pondérations peuvent être calculées offline, et stockées dans des tables accessibles par plusieurs applications à la fois. La complexité du calcul exclut donc la procédure d'optimisation des quantifieurs et compte seulement les opérations élémentaires de somme et de pondération. Au contraire, les méthodes de Monte Carlo utilisent une partie de la capacité de calcul dans la simulation online des échantillons X^i.
- L'estimateur Monte Carlo est un estimateur aléatoire, dont il faudra gérer la variance lors des applications par des procédures de contrôle et de minimisation de variance. A son opposé, l'estimateur par quantification est un estimateur déterministe.
- La vitesse de convergence (en loi) des estimateurs de Monte Carlo est $O(N^{-\frac{1}{2}})$ (TCL) ; il est indépendant de la dimension. La loi du logarithme itéré règle la vitesse de convergence p.s. en $\sqrt{\frac{\log\log N}{N}}$. La convergence des estimateurs par quantification, bien que dépendant de la dimension, est asymptotiquement plus rapide jusqu'à la dimension 2 pour celles du type *ordre 0* et jusqu'à la dimension 4 pour celles du type *ordre 1*. Au delà, la méthode de quantification ne reste pas compétitive lorsque $N \to +\infty$; cependant elle se révèle encore très efficace en pratique notamment en dimension moyenne ($d \leq 10$) lorsque N n'est pas très grand.
  Il est délicat de comparer une méthode déterministe comme la quantification et une méthode fournissant un résultat aléatoire comme la méthode de Monte Carlo. Néanmoins, si l'on se tient à la pratique des utilisations, il paraît naturel de comparer l'erreur de quantification avec la taille d'un intervalle de confiance. Typiquement, avec un choix adéquat d'intervalle de confiance, ceci revient à comparer $\|f(X) - f(\hat{X})\|_2$ et $\frac{2\sigma_{f(X)}}{\sqrt{N}}$ pour $\hat{X}$ une N-quantification L^2-optimale et f Lipshitzienne ; ou bien, plus universellement à comparer $\|X - \hat{X}\|_2$ et $\frac{2\sigma_X}{\sqrt{N}}$ pour les schémas de quantification d'ordre 0 et $\|X - \hat{X}\|_2^2$ et $\frac{2\sigma_X}{\sqrt{N}}$ pour les schémas d'ordre 1. Il est intéressant de voir alors qu'il existe des seuils critiques N_c^0 pour les schémas d'ordre 0 et N_c^1 pour les schémas d'ordre 1, pour lesquels $\forall N \leq N_c^0 \leq N_c^1$ l'erreur par quantification est inférieure à la longueur de l'intervalle de confiance donné par la méthode de Monte Carlo. Ce résultat est détaillé dans [43].

Il est par ailleurs possible d'appliquer un raisonnement d'échantillonnage préférentiel [49] dans l'intégration numérique par quantification.

Définition 0.1.1 ***Echantillonnage Préférentiel :*** *C'est la procédure par laquelle on approche de manière empirique une mesure de probabilité difficile à simuler ν en utilisant*

un échantillon iid $(\xi_1, \ldots, \xi_N)$ *d'une autre mesure de probabilité* μ *dite loi d'importance, plus facile à simuler. S'il existe une fonction* m *et une constante* $\bar{m}$ *vérifiant :*

$$\nu(dx) = m(x)\mu(dx) \qquad et \qquad m(x) \leq \bar{m},$$

l'approximation est alors donnée par $\nu \approx \frac{1}{N}\sum_{i=1}^{N} m(\xi_i)\delta_{\xi_i}$.

Si on désigne par $\mathbf{p}$ la densité de X et par $\mathbf{q}$ une densité d'importance qui vérifiera $f\frac{\mathbf{p}}{\mathbf{q}} \in \mathcal{C}_b$, on pourra utiliser la quantification d'une variable Y de densité q pour estimer $\mathbb{E}[f(X)]$.

$$\begin{aligned} \mathbb{E}[f(X)] = \mathbb{E}[f(Y)\frac{\mathbf{p}(Y)}{\mathbf{q}(Y)}] &\approx \mathbb{E}[f(\hat{Y})\frac{\mathbf{p}(\hat{Y})}{\mathbf{q}(\hat{Y})}], \\ |\mathbb{E}[f(X)] - \mathbb{E}[f(\hat{Y})\frac{\mathbf{p}(\hat{Y})}{\mathbf{q}(\hat{Y})}]| &\leq C\|Y - \hat{Y}\|_2. \end{aligned} \tag{0.1.6}$$

0.1.3 Autres applications

La quantification a historiquement connu plusieurs applications dans le domaine de la théorie de l'information, du traitement du signal et de sa compression. Cependant, la solution qu'elle offre au problème d'intégration numérique permet de l'appliquer dans de nouveaux domaines pour résoudre des problèmes impliquant un calcul numérique d'intégrales, d'espérances ou d'espérances conditionnelles. De tels problèmes se posent en finance dans le cadre de modèles d'évaluation de produits dérivés, où un calcul d'espérance conditionnelle est requis et souvent, ce type de problèmes est reformulé de manière rétrograde en utilisant un principe de programmation dynamique. La quantification, de par son principe de pré-traitement et de calcul off-line des grilles s'adapte bien à ce type d'approche. Parmi les applications qui s'inscrivent dans ce cadre, on cite [7], où est proposé un algorithme d'évaluation d'options américaines et d'estimation de temps d'arrêt optimal, la date d'exercice de l'option, lorsque l'actif sous jacent suit une diffusion brownienne. Comme pour le problème d'intégration numérique, l'utilisation de quantifieurs stationnaires permet dans cette application d'améliorer l'estimation numérique par le passage à un ordre supérieur de convergence (Cf. [6]). Par ailleurs, Pagès et Pham [42] définissent une quantification *markovienne* de processus qui préserve la propriété de Markov du processus original et permet de proposer une solution numérique à un problème de contrôle stochastique apparaissant dans des problématiques financières de gestion de portfeuille. Dans ces applications, la résolution numérique rétrograde est rendue possible grâce aux grilles de quantification précalculées.

Une autre application issue du domaine du traitement du signal est celle de la résolution de problèmes de filtrage non linéaire. Elle a été introduite dans [41], où la construction d'une solution séquentielle rétrograde a permis de voir l'efficacité théorique de la quantification, tout en laissant possible une implémentation numérique classique dite forward. Ici aussi,

comme pour l'intégration numérique, le problème de filtrage non linéaire admet une solution numérique probabiliste de type Monte Carlo, on peut alors se poser les questions suivantes :

- Comment se positionnent les méthodes de filtrage par quantification par rapport au méthodes particulaires du type Monte Carlo ?
- Peut-on définir des procédures de pré-traitement encore plus élaborées pour améliorer la rapidité du calcul on line ? Quel serait l'effet de telles procédures sur le taux de convergence de l'erreur d'estimation ?

Toujours dans le cadre du filtrage, il sera intéressant d'étudier le passage à des schémas numériques de type ordre 1, comme il a été suggéré dans [43] où l'utilisation de la propriété de stationnarité permet de définir des pseudo-schémas récursifs ajoutant un ordre de convergence. Ces schémas ne sont pas implémentables en l'état car ils mettent en jeu des quantités non accessibles numériquement. La question est donc :

- Comment définir des schémas numériques de premier ordre implémentables et présérvant le taux de convergence mis en évidence par les pseudo-schémas ?

Enfin, il est intéressant de relever que le point commun à toutes les applications citées est l'utilisation de formulation *rétrograde* pour établir des majorations de l'erreur sur la valeur objectif, en fonction de l'erreur de quantification. Par le biais du théorème de Zador, il devient ensuite possible de déduire un taux de convergence en fonction de la taille des grilles. En finance, la problématique du filtrage est omniprésente dans les modèles d'actifs à volatilité stochastique : la volatilité est vue comme un signal dont le cours de l'actif est l'observation bruitée. Là encore une question se pose :

- En utilisant le filtre par quantification peut-on proposer des solutions à des problèmes d'évaluation d'options américaines [7, 6] ou d'optimisation de portefeuille [42] dans un cadre à observation partielle, typiquement un modèle à volatilité stochastique ?

Dans notre travail, nous nous sommes intéressés à l'application de la quantification au problème du filtrage, plus particulièrement au calcul du filtre par quantification d'une part et à la quantification du filtre d'une autre. Le problème du filtrage étant au cœur des développements de cette thèse, nous y consacrerons toute la section qui suit.

0.2 Le problème du filtrage

Nous nous intéressons maintenant aux spécifications d'un problème de filtrage et à quelques solutions proposées dans la littérature pour le résoudre. On parle de problème de filtrage quand on est face à un système dont l'évolution en fonction du temps est gérée par un processus caché dont on n'observe que des états bruités. Le filtrage s'inscrit dans une approche bayesienne de reconstitution de la loi conditionnelle du processus caché à un instant donné en s'appuyant sur les observations faites jusqu'à cet instant. Il connait des applications diverses aussi bien dans le domaine de la commande de systèmes physiques que dans le cadre plus récent des marchés financiers. Dans ce qui suit, nous définissons un cadre général à notre problème de filtrage, il constituera à quelques variantes près notre

modèle d'états pour toute la thèse. Par ailleurs, nous passerons en revue quelques méthodes numériques de filtrage qui nous serviront de points de départ à de nouvelles méthodes ou de points de comparaison avec celles-ci.

0.2.1 Le filtrage optimal

On se place dans le cadre d'un problème à temps discret et à horizon fini fixé $n \in \mathbb{N}^*$. On considère les processus signal (X_k) et observation (Y_k) régis par les dynamiques suivantes :

$$\begin{cases} X_k = F_k(X_{k-1}, \varepsilon_k), \quad X_0 \text{ de loi } \mu_0 \text{ connue à priori,} \\ Y_k = G_k(X_k, \eta_k), \quad k \geq 1. \end{cases} \tag{0.2.1}$$

où (ε_k) et (η_k) désignent des suites indépendantes de variables aléatoires iid et indépendantes de X_0. La suite $(\varepsilon_k)_{1\leq k\leq n}$ modélise l'innovation du processus caché, $(\eta_k)_{1\leq k\leq n}$ représente l'imperfection des observations. On se propose de déterminer l'état du système à une date finale n fixée, en s'appuyant sur les observations faites jusqu'à cette date. Au sens de la distance quadratique, l'estimation optimale de X_n sachant les observations $Y_{1:n} = (Y_1, \ldots, Y_n)$ est donnée par l'espérance conditionnelle :

$$\mathbb{E}[X_n | Y_{1:n}].$$

De manière plus générale, l'information la plus riche disponible à travers les observations $Y_{1:n}$ est donnée par la loi conditionnelle $\mathcal{L}(X_n|Y_{1:n})$. Le problème du filtrage consiste donc à calculer cette loi de probabilité, de manière exacte dans le cas du filtrage optimal, ou de manière approchée, on parlera dans ce cas de filtrage sous-optimal.

La formulation explicite de la solution au problème du filtrage n'est généralement pas possible. En effet, dans un cadre général c'est un problème de dimension infinie. Au niveau des applications, on aspire le plus souvent à approcher la densité de la loi conditionnelle à travers un ensemble de fonctions tests qu'on notera de manière générique f. On se donne donc pour objectif de calculer, dans un premier temps :

$$\Pi_{y,n}(dx) = \mathbb{P}[X_n \in dx | Y_1 = y_1, \ldots, Y_n = y_n],$$

ou

$$\Pi_{y,n} f = \mathbb{E}[f(X_n) | Y_1 = y_1, \ldots, Y_n = y_n].$$

Il sera ensuite possible d'envisager le problème d'évaluation du filtre aléatoire :

$$\Pi_{Y,n}(dx) = \mathbb{P}[X_n \in dx | Y_1, \ldots, Y_n] \qquad \text{ou} \qquad \Pi_{Y,n} f = \mathbb{E}[f(X_n) | Y_1, \ldots, Y_n].$$

0.2.1.1 Modèle général d'états

– Le processus signal (X_k) est une chaîne de Markov dont la dynamique est régie par l'équation :

$$X_k = F_k(X_{k-1}, \varepsilon_k),$$

où $F_k : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}^d$, est une fonction borélienne et $(\varepsilon_k)_{1<k\leq n}$ est une suite de variables aléatoires à valeurs dans $\mathbb{R}^q$ de même loi (de densité $\mathbf{p}$), indépendantes entre elles et indépendantes de X_0. La loi μ_0 de X_0 est supposée connue *a priori*. Par ailleurs, on désigne par $\mathbf{P}_k(x, dx')$ le noyau de transition de X_k à X_{k+1}, et on note pour toute fonction f :

$$\mu_0 f = \int f(x)\mu_0(dx) \quad \text{et} \quad \mathbf{P}_k f(x) = \int f(x')\mathbf{P}_k(x, dx').$$

– Le processus des observations (Y_k) obéit à la dynamique suivante :

$$Y_k = G_k(X_k, \eta_k),$$

où $G_k : \mathbb{R}^d \times \mathbb{R}^{q'} \to \mathbb{R}^{d'}$ est une fonction borélienne et (η_k) une suite de variables aléatoires iid à valeurs dans $\mathbb{R}^{q'}$ independantes de $\sigma(X_0, \varepsilon_k, k \geq 1)$. On suppose que la loi conditionnelle $\mathcal{L}(Y_k|X_k)$ est absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}^{d'}$. Soit :

$$\mathbb{P}[Y_k \in dy|X_k = x_k] = g_k(x_k, y)\lambda_{d'}(dy). \tag{0.2.2}$$

La loi initiale du signal étant *a priori* connue, on pourra supposer sans perte de généralité que $Y_0 = y_0$ fixé.

– Le processus (X_k, Y_k) est une chaine de Markov de transition donnée par :

$$\begin{aligned}\mathbb{P}[(X_k, Y_k) \in (dx, dy)|X_{k-1}, Y_{k-1}] &= \mathbb{P}[Y_k \in dy|X_{k-1}, Y_{k-1}, X_k]\mathbb{P}[X_k \in dx|X_{k-1}, Y_{k-1}], \\ &= g_k(x, y)\mathbf{P}_{k-1}(X_{k-1}, dx)\lambda_{d'}(dy).\end{aligned}$$

grâce à l'indépendance entre η_k et (X_{k-1}, η_{k-1}) et à (0.2.2).
Par conséquent, la loi jointe $\mathcal{L}(X_0, \ldots, X_n, Y_0, \ldots, Y_n)$ s'écrit :

$$\mathcal{L}(X_0, \ldots, X_n, Y_0, \ldots, Y_n) = \mu_0(dx_0)\delta_{y_0} \prod_{k=1}^{n} g_k(x_k, y_k)\mathbf{P}_k(x_{k-1}, dx_k)\lambda_{d'}(dy_k),$$

et par la formule de Bayes pour les vecteurs aléatoires à densité, on en déduit la formule de Kallianpur Striebel [27] :

$$\mathbb{E}[f(X_{0:n})|Y_{1:n} = y_{1:n}] = \frac{\int \ldots \int f(x_{0:n})\mu_0(dx_0)\prod_{k=1}^{n} g_k(x_k, y_k)\mathbf{P}_k(x_{k-1}, dx_k)}{\int \ldots \int \mu_0(dx_0)\prod_{k=1}^{n} g_k(x_k, y_k)\mathbf{P}_k(x_{k-1}, dx_k)}. \tag{0.2.3}$$

Plus particulièrement, on définit le filtre évalué sur la fonction test f par :

$$\Pi_{y,n} f = \mathbb{E}[f(X_n)|Y_{1:n} = y_{1:n}] = \frac{\pi_{y,n} f}{\pi_{y,n}\mathbf{1}},$$

où $\pi_{y,n} f$ est le filtre non normalisé défini par :

$$\pi_{y,n} f = \mathbb{E}[f(X_n) \prod_{k=1}^{n} g_k(X_k, y_k)].$$

– Dans la suite, un échantillon $y_{1:n}$ d'observations étant fixé, on confondra par commodité la densité conditionnelle g_k avec la fonction de vraisemblance associée :

$$g_k(x_k) \overset{\text{Déf}}{=} g_k(x_k, y_k),$$

la dépendance en y_k sera implicite lorsque l'observation est fixée.

0.2.1.2 Formulation récursive

En utilisant la propriété de Markov du signal (X_k), il est possible de décomposer séquentiellement le calcul de $\Pi_{y,n}f$ par un argument de programmation dynamique. En effet, le passage d'un filtre à une date intermédiaire $0 \leq k \leq n-1$ au filtre à la date suivante peut être fait en deux étapes connues sous le nom d'étapes de prédiction et de mise à jour.

$$\Pi_k \overset{\text{Prédiction}}{\longrightarrow} \Pi_{k+1|k} \overset{\text{Mise à jour}}{\longrightarrow} \Pi_{k+1}.$$

Prédiction C'est une étape de transition linéaire qui utilise l'information *a priori* de la transition du signal. On définit alors la prédiction :

$$\Pi_{k+1|k}(dx') = \int \Pi_k(dx)\mathbf{P}_k(x, dx'). \tag{0.2.4}$$

Mise à jour C'est l'étape de correction de la prédiction qui utilise l'information fournie par la nouvelle observation, tombée à la date $k+1$ considérée. Elle est non linéaire car effectue une normalisation issue de la formule de Bayes pour l'espérance conditionnelle. Explicitement, on a :

$$\Pi_{k+1}(dx) = \frac{g_{k+1}(x)\Pi_{k+1|k}(dx)}{\int g_{k+1}(x)\Pi_{k+1|k}(dx)}. \tag{0.2.5}$$

La transition de π_k à π_{k+1} pourra aussi être modélisée par la définition d'un opérateur de transition :

$$H_{y,k}f(x) = \mathbb{E}[f(X_k)g_k(X_k, y_k)|X_{k-1} = x], \quad 1 \leq k \leq n,$$

Ainsi,

$$\begin{array}{rcl} \pi_{y,0}f & = & \mathbb{E}[f(X_0)] \overset{\text{Déf}}{=} H_{y,0}f, \\ \pi_{y,k}f & = & \pi_{y,k-1}H_{y,k}f, \quad 1 \leq k \leq n. \end{array} \tag{0.2.6}$$

Dans la suite, il s'avèrera par ailleurs très utile de voir que cette construction *forward* pourra être inversée en un schéma rétrograde ou *backward* (voir [41]). A cette fin, nous définissons les opérateurs $R_{y,k}$ comme suit :

$$\begin{array}{rcl} R_{y,n}f & = & f, \\ R_{y,k-1}f & = & H_{y,k}R_{y,k}f, \quad 1 \leq k \leq n. \end{array} \tag{0.2.7}$$

et on a l'égalité : $\pi_{y,n} = \mu_0 \circ R_{y,0}$.

0.2.1.3 Le modèle d'états discret

La formulation récursive du calcul du filtre par les équations (0.2.4) et (0.2.5) rend la résolution du problème numériquement aisée dans le cadre d'un signal à espace d'états discret. En effet, si pour tout $0 \leq k \leq n$, $X_k(\Omega) = \{x_k^1, \ldots, x_k^{N_k}\}$ et si $\mathbf{P}_k = (\mathbf{P}_k^{ij})$ désigne la matrice de transition du signal (X_k) entre k et $k+1$, d'après le paragraphe précédent, les opérateurs $H_{y,k}$ s'écrivent simplement :

$$H_{y,k} f(x_{k-1}^i) \quad = \quad \sum_{j=1}^{N_k} f(x_k^j) \mathbf{P}_{k-1}^{ij} g_k(x_k^j, y_k).$$

permettant ainsi un calcul numérique explicite du filtre par le biais de la récursion (0.2.6).

En considérant que $\pi_{y,k} \in \mathcal{M}_{1,N_k}(\mathbb{R})$ et que $H_{y,k} \in \mathcal{M}_{N_{k-1},N_k}(\mathbb{R})$, on aboutit au système récursif matriciel suivant :

$$\begin{aligned} H_{y,k}^{ij} &= \mathbf{P}_{k-1}^{ij} g_k(x_k^j, y_k), \quad 0 < k \leq n, \\ \pi_{y,0} &= \mu_0, \\ \pi_{y,k} &= \pi_{y,k-1} H_{y,k}. \end{aligned}$$

Finalement, en normalisant :

$$\Pi_{y,n}^i = \frac{\pi_{y,k}^i}{\sum_{j=1}^{N_n} \pi_{y,k}^j} \qquad \text{et} \qquad \Pi_{y,n} f = \sum_{i=1}^{N_n} \Pi_{y,n}^i \times f(x_n^i).$$

0.2.1.4 Filtrage de Kalman

Quand on sort du cadre discret précédent, l'évaluation exacte du filtre en utilisant les équations (0.2.4) et (0.2.5) devient plus délicate, car elle implique le calcul successif d'intégrales. Le modèle d'état dit de Kalman-Bucy constitue un des rares modèles à espace d'états continu où une formulation explicite du filtre est possible. On parle de filtre de dimension finie [2]. On considère :

$$\begin{cases} X_k = \rho_k X_{k-1} + \theta_k \varepsilon_{k+1}, \quad X_0 \sim \mathcal{N}(m_0, \Sigma_0), \\ Y_k = X_k + \alpha_k \eta_k, \\ \varepsilon_k \ \ et \ \ \eta_k \ \ iid \ \ \sim \mathcal{N}(0, I_d), \\ \rho_k, \ \ \theta_k, \ \ \alpha_k \ \ \in \mathcal{M}_d(\mathbb{R}). \end{cases} \tag{0.2.8}$$

Dans ce cas particulier, (X_k, Y_k) est une suite gaussienne d'où l'en déduit que le filtre Π_k ainsi que la prédiction $\Pi_{k+1|k}$ sont gaussiens de lois respectives $\mathcal{N}(m_k, \Sigma_k)$ et

[2] Filtres décrivant des lois paramétrées par un ensemble fini de paramètres

$\mathcal{N}(m_{k+1|k}, \Sigma_{k+1|k})$. Les paramètres m_k, Σ_k, $m_{k+1|k}$ et $\Sigma_{k+1|k}$ sont connus récursivement par l'algorithme suivant (Cf. [21]), $k = 1, \ldots, n$:

$$\begin{array}{rcl} m_{k+1|k} & = & \rho_k m_k, \\ \Sigma_{k+1|k} & = & \rho_k \Sigma_k \rho_k' + \theta_k \theta_k', \\ & & \\ m_k & = & m_{k|k-1} + K_k \left(Y_k - m_{k|k-1}\right), \\ \Sigma_k & = & (I - K_k)\Sigma_{k|k-1}, \\ K_k & = & \Sigma_{k|k-1} \left(\Sigma_{k|k-1} + \alpha_k \alpha_k'\right)^{-1}. \end{array} \tag{0.2.9}$$

0.2.1.5 Le filtre explicite *de dimension infinie*

Dans un cadre plus général, le modèle d'états est non linéaire ou non gaussien. Dans ces cas plusieurs travaux ont été élaborés pour définir les conditions permettant d'avoir des filtres de dimension finie. On citera dans ce sens [52, 50] pour les filtres à temps discret, et [13] pour les filtres à temps continu. Les résultats de ces travaux montrent qu'en dehors de quelques cas particuliers ([8, 12, 11, 10]), peu de modèles permettent de définir des filtres de dimension finie.
Dans ce paragraphe nous nous interessons au *filtre explicite à dimension infinie* introduit dans [18, 19, 12] qui sera repris plus tard dans les exemples d'application (voir Chapitre 2). L'idée est de définir une famille paramétrée de lois invariantes par les opérations de prédiction (0.2.4) et de mise à jour (0.2.5), en adoptant des conditions suffisantes sur la transition du signal $\mathbf{P}_k$ et sur la vraisemblance g_k. On introduit des familles $(\mathcal{F}^{i,\theta})_{i\in\mathbb{N},\theta\in\Theta}$ de lois paramétrées par un ensemble fini donné Θ qu'on élargit en une famille $\bar{\mathcal{F}}$ par le moyen de mélanges à coefficients $\alpha = (\alpha_i)_{i\in\mathbb{N}} \in S$.

$$\bar{\mathcal{F}} = \{\nu = \sum_{i\geq 0} \alpha_i \nu_\theta^i,\ \alpha = (\alpha_i)_{i\in\mathbb{N}} \in S,\ \theta \in \Theta,\ \nu_\theta^i \in \mathcal{F}^{i,\theta}\},$$

où $S = \{\alpha = (\alpha_i)_{i\in\mathbb{N}}, \forall i \geq 0,\ \alpha_i \geq 0,\ \sum_{i\geq 0} \alpha_i = 1\}$.
Le cas intéressant en pratique est celui des coefficients de mélange de longueur finie $l \in \mathbb{N}$ vérifiant $\alpha_i = 0$ pour tout $i > l$, qui permettent de définir des lois dépendant d'un nombre fini de paramètres. En partant d'un signal de loi initiale dans $\bar{\mathcal{F}}$ à paramètre de mélange fini, on montre que la loi du filtre et de la prédiction sont aussi dans $\bar{\mathcal{F}}$ et sont à coefficient de mélange de longueur finie. Ces paramètres sont calculables séquentiellement explicitement (Cf. Algorithme 5).

0.2.2 Les méthodes d'approximation

Outre les développements particuliers précédents, des méthodes numériques ont été adoptées pour fournir des estimations fiables dans les cas où aucune solution explicite n'est donnée. Bien que d'approches différentes, ces méthodes s'appuient toutes sur le principe

de trouver une représentation fini-dimensionnelle de la loi objectif Π_k. Dans ce qui suit, nous présentons succintement trois méthodes d'approximation numérique.

0.2.2.1 Filtre de Kalman étendu

Cette méthode est utilisée en cas de modèles gaussiens mais non linéaires. Elle a pour principe de considérer que localement, l'évolution du système peut être approchée par des équations linéaires via des développements de Taylor à l'ordre un. On considère le système non linéaire :

$$\begin{cases} X_{k+1} = F_k(X_k, \varepsilon_{k+1}), \\ Y_k = G_k(X_k) + \alpha_k \eta_k. \end{cases} \tag{0.2.10}$$

Pour ce système, le processus solution (Π_k) n'est pas gaussien, ses moments ne peuvent être calculés de manière simple. Cependant, ce système peut être linéarisé afin de permettre la construction d'un algorithme d'approximation récursif du type (0.2.9). La loi du filtre Π_k ainsi que celle de prédiction $\Pi_{k|k-1}$ sont alors approchées par des lois gaussiennes $\mathcal{N}(m_k, \Sigma_k)$ et $\mathcal{N}(m_{k+1|k}, \Sigma_{k+1|k})$. Soit :

$$\begin{aligned} X_{k+1} &\approx F_k(m_k, 0) + \mathrm{D}_x F_k(m_k, 0)(X_k - m_k) + \mathrm{D}_\varepsilon F_k(m_k, 0)\varepsilon_{k+1}, \\ Y_k &\approx G_k(m_{k|k-1}) + \mathrm{D}G_k(m_{k|k-1})(X_k - m_{k|k-1}) + \alpha_k \eta_k. \end{aligned}$$

Par analogie au modèle linéaire gaussien, on définit alors récursivement :

$$\begin{aligned} m_{k+1|k} &= F_k(m_k, 0), \\ \Sigma_{k+1|k} &= \mathrm{D}_x F_k(m_k, 0)\Sigma_k \mathrm{D}_x F_k(m_k, 0)' + \mathrm{D}_\varepsilon F_k(m_k, 0)\mathrm{D}_\varepsilon F_k(m_k, 0)', \\ \\ m_k &= m_{k|k-1} + K_k \left(Y_k - G_k(m_{k|k-1})\right), \\ \Sigma_k &= \left(I - K_k \mathrm{D}G_k(m_{k|k-1})\right)\Sigma_{k|k-1}, \\ K_k &= \Sigma_{k|k-1}\mathrm{D}G_k(m_{k|k-1})' \left(\mathrm{D}G_k(m_{k|k-1})\Sigma_{k|k-1}\mathrm{D}G_k(m_{k|k-1})' + \alpha_k \alpha_k'\right)^{-1}. \end{aligned} \tag{0.2.11}$$

A notre connaissance, cette méthode n'est pas mathématiquement justifiée dans le cas général même si plusieurs développement ont été faits pour des cas particuliers (par exemple [48] pour des cas d'observations en temps continu). Elle reste toutefois très utilisée dans la pratique. Son efficacité est trés dégradée par l'existence de fortes non linéarités, l'imprécision dans la spécification de la loi initiale, l'instabilité du système [34, 46]...

0.2.2.2 Méthodes de grilles

Cette méthode s'appuie sur la construction de grilles d'approximation de chaque variable X_k par des variables discrètes. Dans les termes introduits en première section, il s'agit de la construction de quantifieurs des variables X_k et de la définition de transitions discrètes d'une date à l'autre de manière à revenir au cas du modèle discret. Revenons aux définitions

de la première section et désignons par les $\hat{X}_k$ les N_k-quantifications des X_k. On notera $(A_k^i)_{1\leq i\leq N_k}$ la partition associée à cette quantification et $\hat{X}_k(\Omega) = \Gamma_k = \{x_k^1, \ldots, x_k^{N_k}\}$. On peut alors approcher le filtre récursivement en définissant l'estimateur $\hat{\pi}_{y,n}$ par l'algorithme inspiré du modèle d'états discret ([1, 41]) :

$$\begin{aligned}
\hat{H}_{y,k}f(x_{k-1}^i) &= \mathbb{E}[f(\hat{X}_k)g_k(\hat{X}_k, y_k)|\hat{X}_{k-1} = x_{k-1}^i], \\
&= \sum_{j=1}^{N_k} f(x_k^j)\hat{\mathbf{P}}_{k-1}^{ij} g_k(x_k^j, y_k), \\
\hat{\pi}_{y,0}f &= \mathbb{E}[f(\hat{X}_0)], \\
\hat{\pi}_{y,k}f &= \hat{\pi}_{y,k-1}\hat{H}_{y,k}f.
\end{aligned}$$

En considérant que $\hat{\pi}_{y,k} \in \mathcal{M}_{1,N_k}(\mathbb{R})$ et que $\hat{H}_{y,k} \in \mathcal{M}_{N_{k-1},N_k}(\mathbb{R})$, on aboutit au système récursif matriciel suivant :

$$\begin{aligned}
\hat{H}_{y,k}^{ij} &= \hat{\mathbf{P}}_{k-1}^{ij} g_k(x_k^j, y_k), \quad 0 < k \leq n, \\
\hat{\pi}_{y,0} &= \hat{\mu}_0 = \hat{H}_{y,0}, \\
\hat{\pi}_{y,k} &= \hat{\pi}_{y,k-1}\hat{H}_{y,k}.
\end{aligned}$$

Le choix du quantifieur, notamment de la grille Γ_k et de la partition associée constituent un point crucial dans la qualité de l'estimation. Comme pour le problème de l'intégration numérique, il est possible de voir cette approche comme une approximation à l'ordre zéro des opérateurs R_k dans (0.2.7). Ils sont séquentiellement approchés par des opérateurs constants par morceaux :

$$\begin{aligned}
\widehat{R}_{y,n}f &= f, \\
\widehat{R}_{y,k-1}f &= \hat{H}_{y,k}\widehat{R}_{y,k}f, \quad 1 \leq k \leq n,
\end{aligned} \tag{0.2.12}$$

pour obtenir de manière équivalente $\hat{\pi}_{y,n} = \widehat{R}_{y,0}$.

Dans [41], cette définition rétrograde (0.2.12) permet d'établir un contrôle de l'erreur pour un choix judicieux du quantifieur et de la fonction test f. En effet, un choix de quantifieur L^2-optimal nous permet d'établir un taux de convergence vers zero par le biais du théorème de Zador. Ceci est possible à travers le contrôle de l'erreur sur le filtre par l'erreur de quantification :

Théorème 0.2.2 [41]
Supposons $\mathbf{P}_k$ est Lipschitzien et f est bornée Lipschitzienne continue, alors il existe une suite positive de constantes $(C_j^n(y))_{0\leq j\leq n}$ telles que :

$$|\pi_n f - \hat{\pi}_n f| \leq \sum_{j=0}^{n} C_j^n(y)\|X_j - \hat{X}_j\|_2.$$

Remarque 0.2.2 La dépendance en les observations des constantes réelles $C_j^n(y)$ peuvent être rendues explicites. (Cf. [41]).

Remarque 0.2.3 Le résultat du Théorème 0.2.2 rend aussi efficace le choix d'un quantifieur L^p-optimal pour établir un taux de convergence de l'erreur vers zéro. Le choix quadratique est justifié par la facilité relative du calcul numérique des grilles de quantfication.

0.2.2.3 Méthodes particulaires

Ce sont des méthodes probabilistes où l'approximation est justifiée par la loi des grands nombres. L'idée est proche des méthodes de grilles, dans le sens où le principe d'approcher la loi par une mesure discrete finie est retenue. Comme pour le problème d'intégration numérique par méthode de Monte Carlo, cette mesure discrète charge les points d'un échantillon aléatoire appelé système de particules. L'algorithme du filtre à particules s'appuie ensuite sur la propagation dans le temps du système de particules initialement issus d'un échantillon de la loi initiale μ_0. L'algorithme le plus élémentaire de filtrage particulaire est le filtre de Monte Carlo pondéré, appelé aussi **SIS** pour *Sequential Importance Sampling* algorithm. Pour chaque date d'observation k, on définit l'estimateur $\Pi_k^M f$ par :

$$\Pi_k^M f = \sum_{i=1}^{M} w_k^i f(X_k^i) \quad \text{où} \quad X_k^i \text{ iid } \sim \mathcal{L}(X_{0:k}) \quad \text{et} \quad w_k^i = \frac{g_k(X_k^i) w_{k-1}^i}{\sum_{i=1}^M g_k(X_k^i) w_{k-1}^i}.$$

Cet algorithme nécessite de savoir simuler la loi jointe $\mathcal{L}(X_{0:k})$ ce qui a l'avantage de pouvoir se faire récursivement grâce à la nature markovienne du signal.
D'un point de vue pratique, l'attrait de cette méthode réside dans la possibilité d'une écriture séquentielle de la solution en utilisant les équations (0.2.4) et (0.2.5). En effet, étant donné un système de particules $(X_{0:k}^i)_{1\leq i\leq M}$ selon $\mathcal{L}(X_{0:k})$, les échantillons $(X_{0:k+1}^i)_{1\leq i\leq M}$ simulés selon la transition $\mathbf{P}_k$ à partir de $(X_{0:k}^i)_{1\leq i\leq M}$ sont iid selon $\mathcal{L}(X_{0:k+1})$ et d'après (0.2.4) on définit la prédiction empirique :

$$\Pi_{k+1|k}^M f = \sum_{i=1}^{M} w_k^i f(X_{k+1}^i).$$

L'étape de correction (0.2.5) intervient sur les pondérations w_k^i par le calcul de la valeur prise par g_k en chaque nouveau point simulé, soit :

$$\Pi_{k+1}^M f = \frac{\sum_{i=1}^M g_{k+1}(X_{k+1}^i) w_k^i f(X_{k+1}^i)}{\sum_{i=1}^M g_{k+1}(X_{k+1}^i) w_k^i} = \sum_{i=1}^{M} w_{k+1}^i f(X_{k+1}^i).$$

Numériquement la méthode souffre d'un problème de dégénérescence des pondérations [1, 39]. Pour y remédier, la solution la plus basique et la plus répandue dans la littérature

propose d'ajouter une étape de rééchantillonnage afin d'améliorer l'exploration de l'espace d'états par les particules [22, 15]. Ce type d'algorithmes séquentiels avec rééchantillonnage multiplie les particules à fortes pondérations et élimine les autres, il est retrouvé sous plusieurs appellations : *filtre particulaire avec interaction* [39], *filtre Bootsrap* [24] ou *filtre Monte Carlo* [39, 36, 30], on le notera **SIR** pour *Sequential Importance Resampling*. Différents types de convergence de ces méthodes sont établis dans [14, 15]. Il est cependant nécessaire de mentionner que la solution du rééchantillonnage peut s'avérer insuffisante dans plusieurs cas pratiques. Il arrive en effet qu'elle apauvrisse la population des particules en concentrant le nuage sur peu de points. La taille effective est ainsi réduite, on parle de dégénérescence des particules du filtre **SIR**. On se reportera à [1, 39] et aux références qu'ils contiennent, pour une revue de quelques variantes permettant de résoudre ce type de problèmes.

0.3 Principaux résultats

Cette thèse présente quelques contributions à la résolution du problème du filtrage en utilisant la méthode de la quantification optimale. Nous nous sommes intéressés aux problèmes théoriques posés par la majoration de l'erreur dans différentes applications du filtrage par quantification et à la vérification numérique de ces résultats via l'implémentation sur machine. Ce travail s'articule en deux parties. La première est consacrée à l'approfondissement des méthodes de filtrage par quantification déjà construites par Pagès et Pham dans [41], en utilisant une approche de développement au premier ordre (Chapitre 1). La comparaison numérique de cette approche par grilles à l'approche Monte Carlo des filtres particulaires fait l'objet du Chapitre 2, dans lequel cette étude se dresse à travers plusieurs modèles d'états. Dans la deuxième partie, nous nous sommes intéressés à l'optimisation de la procédure du filtrage numérique par le moyen de grilles de quantification précalculées soit à travers la quantification des observations d'une part (Chapitre 3) ou d'autre part à travers la quantification du filtre même (Chapitre 4). Le but dans la première approche était d'élaborer un algorithme de calcul plus rapide, tout en établissant une majoration satisfaisante de l'erreur. La deuxième visait à donner une solution numérique au problème d'évaluation d'option américaine dans le cadre d'un marché à volatilité stochastique non observée.

Le Chapitre 1 a été motivé par les travaux de Pagès, Pham et Printems [43], qui suggéraient de construire des schémas dits d'ordre 1 (Cf. paragraphes 0.1.2 et 0.1.3) pour le calcul du filtre par quantification. L'introduction de correcteurs dits d'*ordre 1*, dans les formules récursives (0.2.12), s'inspirant ainsi des résultats de Bally, Pagès et Printems dans [6], a donné naissance à deux difficultés à dépasser pour aboutir à des algorithmes numériques implémentables ayant une convergence de type *ordre 1*. Il fallait proposer des correcteurs d'une part accessibles numériquement, et d'autre part qui permettent d'établir une amélioration du type de convergence. Nous proposons deux résolutions possibles du

problème selon les hypothèses sur le modèle d'état. Typiquement, comme dans [43], la résolution numérique rétrograde du filtre par quantification permet de construire un squelette de schéma dit *schéma générique d'ordre un*, où les termes correcteurs viennent s'ajouter aux approximations d'ordre zéro à la manière d'un développement de Taylor.

$$\left\{\begin{array}{lcl} \widehat{R}_n f & = & f, \\ \widehat{DR}_n f & = & \mathrm{D}f, \\ \widehat{R}_k f & = & \mathbb{E}[g_{k+1}(X_{k+1}, y_{k+1}) \left(\widehat{R}_{k+1} f(X_{k+1} + \langle \widehat{\mathbf{DR}}_{\mathbf{k+1}} \mathbf{f}(\hat{\mathbf{X}}_{\mathbf{k+1}}), \mathbf{\Delta}_{\mathbf{k+1}} \rangle \right) | \hat{X}_k], \\ & & 0 \leq k \leq n-1. \end{array}\right. \tag{0.3.1}$$

Dans les deux algorithmes proposés dans le Chapitre 1, nous définissons l'opérateur $\widehat{DR}_{k+1} f$ comme une approximation d'ordre zéro de $\mathrm{D}R_k f$, en partant de deux écritures différentes de ce dernier. Le premier algorithme Algorithme 1 s'appuie sur une définition récursive donnant $\mathrm{D}R_k$ en fonction de $\mathrm{D}R_{k+1}$, le deuxième Algorithme 2 sur une transformation *à la Malliavin* de $\mathrm{D}R_k$ pour le réécrire en fonction de R_{k+1} et d'une fonction poids que l'on définit (Cf. [6]). Sous les hypothèses permettant ses constructions et d'autres sur la dynamique du signal, nous établissons que les approximations de filtre $\hat{\pi}_{y,n}$ issues de tels schémas améliorent le résultat du Théorème 0.2.2 (Théorèmes 1.3.1 et 1.4.1).

Dans le Chapitre 2, nous dressons une étude comparative de l'implémentation et de la convergence numérique de ces filtres par quantification avec les filtres à particules. Il est intéressant d'y voir le parallèle qui peut être fait conceptuellement entre les deux approches. A partir des différents exemples de modèles d'état qu'on a testés, il se dégage que les méthodes de filtrage par quantification s'avèrent efficaces en termes de convergence même si leur complexité numérique augmente en dimension supérieure à 1. Par rapport aux méthodes à particules, elles ont l'avantage de donner une estimation déterministe au filtre et d'éviter les problèmes posés par la variance des solutions particulaires.

Dans le Chapitre 3, nous étendons la notion de quantification des observations introduite par [38] pour les filtres à modèles d'états discret (Cf. paragraphe 0.2.1.3) au cas des filtres à espace d'états continu (Algorithmes 6 et 7). Ceci est rendu possible par l'adoption des schémas numériques de filtres par quantification. En effet, par son principe de prétraitement offline, le filtrage par quantification permet de stocker au préalable, en plus des quantifieurs du signal, les fonctions de vraisemeblance évaluées sur les produits des grilles de quantification du signal et de l'observation. En projetant l'observation sur sa grille de quantification, il est ainsi possible de construire un algorithme ne faisant pas intervenir le calcul de la vraisemblance online et par conséqeunt plus rapide. L'erreur L^1 sur l'estimation du filtre aléatoire $\Pi_{Y,n}$ est contrôlée par l'erreur de quantification de l'observation dans le cas discret (Théorème 3.2.1), et par celle de l'observation et du signal dans le cas continu (Théorème 3.3.2).
Comme pour le signal, la procédure de prétraitement des observations permet d'accélérer l'éxécution numérique du schéma d'estimation. Elle permet aussi d'envisager un pré-

traitement complet du filtre par quantification. Ceci est étudié dans de plus amples détails dans le Chapitre 4 à travers une application en finance.

Nous nous sommes intéressés au problème d'évaluation d'options américaines dans le cadre d'un modèle de marché à volatilité stochastique. La volatilité X_k est modélisée comme un signal Markovien caché, dont le prix d'actif Y_k constitue une observation bruitée. Nous donnons une résolution numérique du problème de recherche de temps d'arrêt dans ce cadre de marché à information incomplète. Soit :

$$u_0 = \sup_{\tau \in \mathcal{T}_n^Y} \mathbb{E}\left[\sum_{k=0}^{\tau} f(X_k, Y_k) + h(X_\tau, Y_\tau)\right], \tag{0.3.2}$$

où $\mathcal{T}_n^Y$ est l'ensemble des temps d'arrêt adaptés à la filtration des observations $(\mathcal{F}_k^Y) = (\sigma\{Y_0, \ldots, Y_k\})$ à valeurs dans $\{0, \ldots, n\}$. En considérant la variable $Z_k = (Y_k, \Pi_{Y,k})$, ce problème est transformé en un problème de temps d'arrêt à information complète :

$$u_0 = \sup_{\tau \in \mathcal{T}_n^Y} \mathbb{E}\left[\sum_{k=0}^{\tau} \hat{f}(Z_k) + \hat{h}(Z_\tau)\right]. \tag{0.3.3}$$

où Z_k sera une chaîne de Markov par rapport à $(\mathbb{P}, (\mathcal{F}_k^Y))$. La résolution par quantification, à la manière de Bally et Pagès dans [4] est ainsi rendue possible. Une majoration de l'erreur est donnée par le Théorème 4.4.2. Sur le plan numérique, cet exemple est illustré par un problème d'évaluation d'option américaine, il est constaté que la valeur de l'option en observation partielle converge vers celle à observation totale, quand la taille des quantifieurs tend vers $+\infty$.

Le chapitre 4 porte sur un article co-écrit avec H. Pham et W. Runggaldier et publié dans [47]. Les autres chapitres font l'objet de pré-publications du laboratoire de Probabilités et Modèles Aléatoires. Tous les chapitres peuvent être lus indépendamment les uns des autres, le lecteur est prié d'excuser les répétitions inévitables dans les définitions et préliminaires.

Première partie

Filtrage par quantification

Chapter 1

First Order schemes

Prépublication du laboratoire de Probabilités et Modèles Aléatoires [54]

The quantization based filtering method (see [41], [43]) is a grid based approximation method to solve nonlinear filtering problems with discrete time observations. It relies on off-line preprocessing of some signal grids in order to construct fast recursive schemes for filter approximation. We give here an improvement of this method by taking advantage of the stationary quantizer property. The key ingredient is the use of vanishing correction terms to describe schemes based on piecewise linear approximations. Convergence results are given and comparison with sequential Monte Carlo methods is made. Numerical results are presented for both particular cases of linear Gaussian models and stochastic volatility models.

Key words: Quantization, nonlinear filtering, offline preprocessing, stationary quantizer, stochastic volatility models.

1.1 Introduction

In several scientific fields, it is often required to estimate the changing state of a system using noisy observations of its evolution over time. A common manner to do this is the Bayesian approach which constructs the probability density function (pdf) of the state at a given date conditionally to all the available observations till this date.
In the Gaussian linear case, called also the Kalman case (KF) [21, 1], the required pdf is Gaussian and by computing sequentially its two first moments, we can determine it exactly. So in this case an explicit solution is provided. Unfortunately, except in this case, or in a few other cases like the discrete finite state space [1] and some other mixing Gaussian models [18], there is usually no closed expression to the problem solution. So, many numerical estimations have been suggested to represent and recursively produce approximations of the state pdf.
In this context, two different approaches can be mentioned: fisrt, the requierd pdf is represented as a sample which would provide an approximation of the distribution when its size becomes very large [15], this includes for example bootstrap Bayesian method [24] or the interacting particle filter [36, 39]. Second, a quantization of the state space is used in order to come back to the discrete finite case. As the size of the quantizations grows to infinity, it is shown that we can asymptotically approach the continuous infinite state space case. Here, the deal will be in estimating some weights associated to some given *grid* points, which define a finite state discrete distribution. This distribution will approach the continuous space case as the *grid* size gets larger. The weighted Monte Carlo filter [1, 39, 15] using random samples to compute grids and the Kitagawa method [29] for linear non Gaussian models using predefined grids and optimal quantization filtering [41] using off line computations to produce an optimal quantization of the state process are examples of this approach.
The technique of optimal quantization of random vectors is especially useful in problems where many expectations or conditional expectations need to be computed. It appears as an efficient method to transform an integral into a finite weighted sum with a controlled approximation error. We can find some applications of this technique in [43, 4]. In [44], some numerical methods to construct optimal quantization grids for multidimensional Gaussian distributions are given.
Now for the pdf estimation problem we treat here, we use Kallianpur-Striebel formula [27] to derive a dynamic programming formula allowing to estimate the pdf recursively. Like in [41], this approach makes possible the use of quantization at each time step in order to compute conditional expectations. We will call the algorithm introduced in [41] the *zero order scheme*.
In this paper, we are interested by first order approximation using optimal or at least stationary quantizers to estimate the required pdf. This approach was first introduced in [6] for solving optimal stopping time problems, namely multi-asset American option

pricing. It improves the convergence rate of the method. In [43], a first sketch of this idea is presented for pdf estimation but with a pseudo-numerical scheme, which cannot be implemented in practice. Our aim here is to propose operating first order schemes which improve the convergence rate of the zero order schemes from both theoretical and practical viewpoints. We first present them in a backward way; this is the natural manner to devise them and the appropriate formulation to establish error estimates. Then, we show how to derive the forward formulation to be implemented in practice.
The paper is organized as follows: in the second section we give some brief preliminaries on quantization and filtering. The third and fourth sections will deal with the algorithms using first order schemes. Each one presents the approximation procedure, the schemes in their backward and forward formulation and finally convergence theorems. Then, the fifth section is dedicated to summarize the previous results, and enlarge them to the case of normalized filters. Finally, numerical results are presented in the sixth section, including comparison with particle methods for several models.

Notations:
$p \in (1, +\infty)$ is a fixed real number, $|.|$ and $\|.\|_p$ denote respectively Euclidean norm on $\mathbb{R}^d$ and $\mathbf{L}^p$-norm. $\mathcal{C}^1_{b,Lip}$ is the set of continuous differentiable functions $\mathbb{R}^d \to \mathbb{R}$, bounded with bounded Lipschitz continuous derivative and $\mathcal{C}^k_b$ the set of continuous k-times differentiable functions $\mathbb{R}^d \to \mathbb{R}$, bounded with bounded derivatives. We will also define $\|f\|_\infty = \sup_{x\in\mathbb{R}^d} |f(x)|$ and $[f]_{Lip} = \sup_{x\neq x'} \frac{|f(x)-f(x')|}{|x-x'|}$. $\alpha > 0$ denotes a generic constant, $\langle ., . \rangle$ the Euclidean inner product on $\mathbb{R}^d$, A' the transpose of the real matrix A. Finally, $(e_i)_{1\leq i\leq d}$ is the canonical orthonormal basis of $\mathbb{R}^d$.

1.2 Preliminaries

1.2.1 Quantization filtering schemes

We consider a fixed discrete horizon $n \in \mathbb{N}^*$ and some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. A signal process is an $\mathbb{R}^d$-valued discrete time hidden Markov chain $(X_k)_{0\leq k\leq n}$ evolving according to the following signal equation:

$$X_{k+1} = F_{k+1}(X_k, \varepsilon_{k+1}), \quad 0 \leq k \leq n-1, \tag{1.2.1}$$

where $F_k : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}^d$, is a Borel function and $(\varepsilon_k)_{1<k\leq n}$ is a sequence of iid $\mathbb{R}^q$-valued random variables, independent of X_0. The distribution μ_0 of X_0 is supposed to be known. Furthermore, $\mathbf{P}_k(x, dx')$ will denote the probability transition of X_k, and:

$$\mu_0 f = \int f(x)\mu_0(dx) \quad \textit{and} \quad \mathbf{P}_k f(x) = \int f(x')\mathbf{P}_k(x, dx').$$

At each time step k, Y_k an $\mathbb{R}^{d'}$-valued noisy observation of X_k is made. The dynamics of the observation process $(Y_k)_{0\leq k\leq n}$ are driven by Borel functions $G_k : \mathbb{R}^{d'} \times \mathbb{R}^d \times \mathbb{R}^{q'} \to \mathbb{R}^{d'}$

so that:

$$Y_k = G_k(Y_{k-1}, X_k, \eta_k), \quad 1 \leq k \leq n, \tag{1.2.2}$$

where (η_k) is a sequence of iid $\mathbb{R}^{q'}$-valued random variables, independent of $\sigma(X_0, \varepsilon_k, k \geq 1)$. We assume for convenience, that $Y_0 = 0$ and that, for every $1 \leq k \leq n$, the distribution of Y_k given X_k and Y_{k-1} admits a continuous conditional pdf $y \mapsto g_k(Y_{k-1}, X_k, y)$. We suppose in addition that g_k satisfies the following Lipschitz assumption:
$\forall x, x' \in \mathbb{R}^d, \ \ \forall y, y' \in \mathbb{R}^{d'}$,

$$|g_k(y, x, y') - g_k(y, x', y')| \leq [g_k]_{Lip}^{y,y'}|x - x'| \quad \text{and} \quad \max_{0\leq k\leq n} \sup_{x\in\mathbb{R}^d} |g_k(y, x, y')| \leq L^{y,y'} < +\infty.$$

Remark 1.2.1 As the observation process is fixed, we will drop the dependancy of $[g_k]_{Lip}^{y,y'}$ and $L^{y,y'}$ in (y, y') for notational convenience.

The problem we aim to solve is to compute

$$\Pi_n f = \mathbb{E}[f(X_n)|Y_1 = y_1, \ldots, Y_n = y_n],$$

for any reasonable Borel function $f : \mathbb{R}^d \to \mathbb{R}$ and a given observation sequence $y = (y_1, \ldots, y_n)$.

Using Kallianpur-Striebel formula [27], the problem can be reduced to the computation of the unnormalized filter π_n defined by:

$$\pi_n f = \mathbb{E}[f(X_n) \prod_{k=1}^{n} g_k(y_{k-1}, X_k, y_k)].$$

Then, $\Pi_n f = \frac{\pi_n f}{\pi_n \mathbf{1}}$.

Remark 1.2.2 For convenience, the dependency of Π_n and π_n in the observation process has been omitted, as y is fixed. For the same reason, we will denote $g_k(x) := g_k(y_{k-1}, x, y_k)$ for $1 \leq k \leq n$, and $g_0 := \mathbf{1}$.

By introducing the operators $(H_k)_{0\leq k\leq n}$ defined below, a sequential definition of the unnormalized filter π_n can be given.
Namely, if one defines, for every $x \in \mathbb{R}^d$:

$$\begin{cases} H_k f(x) = g_k(x)\mathbb{E}[f(X_{k+1})|X_k = x], \quad 0 \leq k \leq n-1, \\ H_n^n f(x) = g_n(x) f(x), \end{cases} \tag{1.2.3}$$

then we have

$$\pi_n f = \mu_0 \circ H_0 \cdots \circ H_n^n f. \tag{1.2.4}$$

Consequently, we can write sequentially, either in the forward way:

$$U_0 = \mu_0 \circ H_0 \ , \quad U_k = U_{k-1} \circ H_k \ , \quad 1 \leq k \leq n-1, \tag{1.2.5}$$

or in the backward way:

$$R_n = H_n^n \ , \quad R_k = H_k \circ R_{k+1} \ , \quad 0 \leq k \leq n-1, \tag{1.2.6}$$

so that $\pi_n f = \mu_0 R_0 f = U_{n-1} \circ H_n^n f$.

Remark 1.2.3 Note that if G_k depends on X_{k-1} instead of X_k for $1 \leq k \leq n$, we are led to consider the conditional pdf of Y_k, given X_{k-1} and Y_{k-1}. We can then define differently the operators H_k so that $\pi_n f$ still satisfy formally equation (1.2.4).
Namely,

$$\begin{cases} H_k f(x) = g_{k+1}(x)\mathbb{E}[f(X_{k+1})|X_k = x], \quad 0 \leq k \leq n-1, \\ H_n^n f(x) = f(x). \end{cases} \tag{1.2.7}$$

Then, schemes (1.2.5) and (1.2.6), with this new definition of the (H_k) operators, are still valid.

Remark 1.2.4 When G_k depends on both X_{k-1} and X_k, we can also adapt the scheme to the modified $\mathbb{R}^{2d}$-valued signal Markov chain $Z_k = (X_{k-1}, X_k)$ and the same observation process Y_k. In this case we define the new observation dynamics:

$$\bar{G}_k(Y_{k-1}, Z_k, \eta_k) \stackrel{Def}{=} G_k(X_{k-1}, Y_{k-1}, X_k, \eta_k).$$

We succeed then to restore state equations of type (1.2.1) and (1.2.2). The point is that in this case, the signal dimension is twice the original one. This can be numerically constraining, particularly when using grid based approximation methods.

From the recursive definition of either U_k or R_k, it becomes clear that it will be useful to approximate X_k by a random variable $\hat{X}_k$ taking a finite number of values, in order to transform conditional expectations in finite weighted sums. This operation is commonly called *quantization*, and is extensively used in signal processing fields (see [25, 4, 44]).

Temporarily, we suppose that we are able to construct such an approximation $\hat{X}_k$. We define the induced error $\Delta_k := X_k - \hat{X}_k$. Further details about the error modulus $\|\Delta_k\|_p$, $p \geq 1$ will be given in the next paragraph. In [41], these quantizations $\hat{X}_k$ are used to produce a piecewise constant approximation of R_k. So, the *natural* approximation procedure by quantization, as defined in (1.2.8) below appears as a zero order scheme.
It is defined as follows:

$$\begin{cases} \hat{\hat{H}}_k f(\hat{X}_k) = g_k(\hat{X}_k)\mathbb{E}[f(\hat{X}_{k+1})|\hat{X}_k], \quad 0 \leq k \leq n-1, \\ \hat{\hat{H}}_n^n f(\hat{X}_n) = g_n(\hat{X}_n) f(\hat{X}_n). \end{cases} \tag{1.2.8}$$

Defining $\hat{\mu}_0$ the discrete distribution of $\hat{X}_0$, we have respectively the following forward and backward iterative zero order approximation schemes:

$$\hat{\hat{U}}_0 = \hat{\mu}_0 \hat{\hat{H}}_0 \ , \quad \hat{\hat{U}}_k = \hat{\hat{U}}_{k-1} \circ \hat{\hat{H}}_k \ , \quad 1 \leq k \leq n-1, \tag{1.2.9}$$

and

$$\hat{\hat{R}}_n = H_n^n \ , \quad \hat{\hat{R}}_k = \hat{\hat{H}}_k \circ \hat{\hat{R}}_{k+1} \ , \quad 0 \leq k \leq n-1, \tag{1.2.10}$$

so that $\hat{\hat{\pi}}_n f = \hat{\mu}_0 \hat{\hat{R}}_0 f = \hat{\hat{U}}_{n-1} \circ H_n^n f$.
Formally, this scheme is slightly different from that presented in [41] (the definition of H_k operators is different inducing a shifted scheme structure). Nevertheless, the zero order quantization filter estimator itself remains the same. This form of the scheme allows to produce costlessly some error bounds for a wider class of test functions f than in the original theorem established in [41].

Theorem 1.2.1 *Assume that the transition kernels $\mathbf{P}_k$ of the signal Markov chain are K-Lipschitz operators i.e $\forall f : \mathbb{R}^d \to \mathbb{R}$ Lipschitz, $[\mathbf{P}_k f]_{Lip} \leq K[f]_{Lip}$.*
Then, for any f such that $H_n^n f$ is bounded Lipschitz continuous, and $0 \leq k \leq n$, there exists a sequence of positive constants $(C_j^{k,n})_{k \leq j \leq n}$ such that:

$$\|R_k f(X_k) - \hat{\hat{R}}_k f(\hat{X}_k)\|_p \leq \sum_{j=k}^{n} C_j^{k,n} \|\Delta_j\|_p$$

and $C_j^{k,n} \leq \alpha(p,f) L^{n-k} \frac{K^{n-j+1}-1}{K-1}$.

Proof.
The proof of this result is easily adapted from [41] by considering the *shifted* scheme (1.2.10), based on the definition (1.2.3) of the H_k operators. We simply take in consideration that at the last date, we will have $H_n^n f$ instead of f. For that reason, the Lipschitz bounded assumption is made on $H_n^n f$ rather than on f. For a detailed proof, see Appendix 1.6.3. □

Remark 1.2.5 This shifted structure (1.2.10) of the zero order scheme can be useful since regularity and boundedness assumptions have to be satisfied by $H_n^n f$ instead of f (see [41]). This is an advantage, particularly when the conditional pdf g_k goes to zero very fast as $|x| \to +\infty$. For example, if $g_n(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{|y_n - x|^2}{2})$, $H_n^n f$ is Lipschitz continuous and bounded for f bounded Lipschitz continuous as well as for any Lipschitz function f such that $|f(x)| = O(exp(\frac{\alpha |x|^2}{2}))$ for some $0 < \alpha < 1$.

Corollary 1.2.1 *If $\mathbf{P}_k$ is Lipschitz and $H_n^n f$ is bounded Lipschitz continuous, then there exists a sequence of positive constants $(C_j^n)_{0 \leq j \leq n}$ such that:*

$$|\pi_n f - \hat{\hat{\pi}}_n f| \leq \sum_{j=0}^{n} C_j^n \|\Delta_j\|_p.$$

Let us now examine the *quantization error* Δ_k and try to establish some convergence rate toward 0, in which case Corollary 1.2.1 will give a convergence rate of the zero order quantization filter estimation.

1.2.2 Background on quantization and optimal quantization

The aim of quantization is the definition of a random variable taking finite number of values in $\mathbb{R}^d$ as an approximation of an $\mathbb{R}^d$-valued one. In this paragraph, we will present results useful to our work, further details can be found in [25, 44].
Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}^d$ be a random vector and let $\mathbb{P}_X$ denote its probability distribution. A positive integer N being fixed, let $h : \mathbb{R}^d \to \mathbb{R}^d$ be a Borel map such that $|h(\mathbb{R}^d)| \leq N$. We say that $h(X)$ is a N-quantization of X and that $h(\mathbb{R}^d)$ is a N-quantizer. For convenience, the function h itself will be called N-quantizer.

Now, when $X \in L^p(\Omega)$, we aim to construct an L^p-optimal N-quantization of X. That is to determine the function h, if any, which minimizes the L^p-quantization error.
This amounts to solving the optimization problem:

$$\inf\{\|X - h(X)\|_p^p, h : \mathbb{R}^d \to \mathbb{R}^d, \text{ Borel map s.t. } |h(\mathbb{R}^d)| \leq N\}. \tag{1.2.11}$$

This optimization problem has (at least) one solution (see e.g [25]). Any such a solution h^* is called an L^p-optimal N-quantizer (or L^p-optimal N-codebook). Furthermore, one shows that L^p-optimal N-quantizers have full size i.e $|h^*(\mathbb{R}^d)| = N$ and we denote $\Gamma^* := h^*(\mathbb{R}^d) = \{x^1, \ldots, x^N\}$. It is clear that in this case, h^* will necessarily be a projection following the nearest neighbor rule on Γ^*. Namely:

$$h^*(\xi) = \sum_{i=1}^{N} x^i \mathbf{1}_{\mathbf{C}_i(\Gamma^*)}(\xi) \tag{1.2.12}$$

where $(\mathbf{C}_i(\Gamma^*))_{1 \leq i \leq N}$, called the Voronoi diagram of Γ^*, makes up a Borel partition of $\mathbb{R}^d$ satisfying :

$$\mathbf{C}_i(\Gamma^*) \subset \{\xi \in \mathbb{R}^d \ s.t. \ |\xi - x^i| = \min_{1 \leq k \leq N} |\xi - x^k|\}.$$

As a consequence, the induced L^p-mean quantization error (or L^p-distortion) reads:

$$\underline{\mathcal{D}}_N^{X,p} := \|X - h^*(X)\|_p^p = \| \min_{1 \leq i \leq N} |X - x^i| \|_p^p.$$

According to [25, 4], $\underline{\mathcal{D}}_N^{X,p}$ is a (strictly) decreasing sequence converging to 0 when $N \to +\infty$. Furthermore, the rate of convergence of $\underline{\mathcal{D}}_N^{X,p}$ toward 0 is ruled by Zador's Theorem:

Theorem 1.2.2 (see [25, 4]) *Assume that* $\int_{\mathbb{R}^d} |\xi|^{p+\eta} \mathbb{P}_X(d\xi) < +\infty$ *for some* $\eta > 0$. *Then*

$$\lim_N (N^{\frac{p}{d}} \underline{\mathcal{D}}_N^{X,p}) = J_{p,d} \|\varphi\|_{\frac{d}{d+p}}$$

where $\mathbb{P}_X(d\xi) = \phi(\xi)\lambda_d(d\xi) + \bar{\mu}(d\xi)$, $\bar{\mu} \perp \lambda_d$ (λ_d *Lebesgue measure on* $\mathbb{R}^d$) *and for every* $q \in \mathbb{R}_+^*$, $\|g\|_q := (\int |g|^q(u)du)^{\frac{1}{q}}$.

This theorem, combined with Corollary 1.2.1 establishes a convergence rate result for the quantization based zero order scheme (1.2.9).
Now let us introduce an important property of quadratic optimal quantizers:

Proposition 1.2.1 ***(Stationary quantizer property)***
If $\hat{X}$ is a L^2-optimal N-quantization of X, then the stationary quantizer property is verified. Namely,

$$\mathbb{E}[X|\hat{X}] = \hat{X}. \tag{1.2.13}$$

This property is of great help to appreciate the quality of some estimations. This is shown in further details in [44] for numerical integration and in [6] for optimal stopping problems. To illustrate this point by a short example, take the problem of approximating $f(X)$ by $f(\hat{X})$, when $f \in \mathcal{C}_b^2$. We have for some $\xi \in (X, \hat{X})$:

$$f(X) - f(\hat{X}) = \langle \mathrm{D}f(\hat{X}), \Delta \rangle + \frac{1}{2}\Delta' \mathrm{D}^2 f(\xi)\Delta.$$

So, if $\hat{X}$ is a stationary N-quantization of X, we have:

$$\begin{aligned} \mathbb{E}[f(X)|\hat{X}] - f(\hat{X}) &= \langle \mathrm{D}f(\hat{X}), \mathbb{E}[\Delta|\hat{X}] \rangle + \frac{1}{2}\mathbb{E}[\Delta' \mathrm{D}^2 f(\xi)\Delta|\hat{X}] \\ \|\mathbb{E}[f(X)|\hat{X}] - f(\hat{X})\|_p &\leq \frac{1}{2}\|\mathrm{D}^2 f\|_\infty \|\langle \Delta, \Delta \rangle\|_p \leq \frac{1}{2}\|\mathrm{D}^2 f\|_\infty \|\Delta\|_{2p}^2 \end{aligned}$$

We see that, owing to the stationary quantizer property (1.2.13) we succeed to gain one order in estimation costlessly.

Back to our filtering problem, we are interested in quantizing the Markov chain $(X_k)_{0 \leq k \leq n}$. We must settle at each step $0 \leq k \leq n$, a quantizer size N_k and an L^p-optimal N_k-quantizer of X_k denoted $\Gamma_k = \{x_k^1, \ldots, x_k^{N_k}\}$. Consequently, we define $(\hat{X}_k)$ an L^p-optimal (N_k)-quantization of the process (X_k) by:

$$\hat{X}_k = \sum_{i=1}^{N_k} x_k^i \mathbf{1}_{\mathbf{C}_i(\Gamma_k)}(X_k), \quad for \;\; 0 \leq k \leq n. \tag{1.2.14}$$

As the resulting process $(\hat{X}_k)_{0 \leq k \leq n}$ is no longer a Markov chain, this procedure is called *marginal quantization*[1] of the process (X_k).
Nevertheless, an approximation of the transition kernels $\mathbf{P}_k$ of the chain is provided by the following *transition probability* terms:

$$p_k^{ij} = \mathbb{P}[\hat{X}_{k+1} = x_{k+1}^j | \hat{X}_k = x_k^i], \quad i \in \{1, \ldots, N_k\} \;\text{ and }\; j \in \{1, \ldots, N_{k+1}\}.$$

For $0 \leq k < n$ and $i \in \{1, \ldots, N_k\}$, we will denote

$$\widehat{\mathbf{P}}_k f(x_k^i) = \mathbb{E}[f(\hat{X}_{k+1})|\hat{X}_k = x_k^i] = \sum_{j=1}^{N_{k+1}} f(x_{k+1}^j) p_k^{ij}.$$

[1] More details on process quantization are given in [41].

1.2.3 Generic first order scheme

As Theorem 1.2.2 gives a convergence rate of $\underline{\mathcal{D}}_N^{X_k,p}$ toward zero, results such as Corollary 1.2.1 suggest that the quantization filter scheme would lead to better results if we succeed to upper bound the error by higher powers of $\|\Delta_j\|_p$. This leads us to the idea of mimicking first order Taylor expansions in the R_k approximation.
From now on, $(\hat{X}_k)$ denotes a marginal stationary (N_k)-quantization of (X_k), and we denote $\hat{X}_k(\Omega) = \Gamma_k = \{x_k^1, \ldots, x_k^{N_k}\}$. So, $\hat{X}_k = \sum_{i=1}^{N_k} x_k^i \mathbf{1}_{\mathbf{C_i}(\Gamma_k)}$. Since $\hat{X}_k$ is $\sigma(X_k)$-measurable, using the chaining rule for conditional expectation $\mathbb{E}[.|\hat{X}_k] = \mathbb{E}[\mathbb{E}[.|X_k]|\hat{X}_k]$ yields:

$$\mathbb{E}[f(X_{k+1})|\hat{X}_k] = \mathbb{E}[\mathbf{P}_k f(X_k)|\hat{X}_k]. \tag{1.2.15}$$

In view of Proposition 1.2.1, if $\mathrm{D}(\mathbf{P}_k f)$ exists (and is Lipschitz) we can write:

$$\mathbb{E}[f(X_{k+1})|\hat{X}_k] = \mathbf{P}_k f(\hat{X}_k) + \langle \mathrm{D}(\mathbf{P}_k f)(\hat{X}_k), \overbrace{\mathbb{E}[\Delta_k|\hat{X}_k]}^{0} \rangle + O(|\Delta_k|^2). \tag{1.2.16}$$

We can then approach $\mathbb{E}[f(X_{k+1})|\hat{X}_k]$ by $\mathbf{P}_k f(\hat{X}_k)$ with an L^1-estimation error of order $O(\|\Delta_k\|_2^2)$. This is the key idea for constructing first order quantization schemes. For such a purpose, we assume that:

H 1 *For any observation process y, all functions g_k lie in $\mathcal{C}_{b,Lip}^1$ and there exists $L > 0$ such that*

$$\max_{0\le k\le n} \{\|g_k\|_\infty, \|\mathrm{D}g_k\|_\infty, [\mathrm{D}g_k]_{Lip}\} \le L.$$

and that:

H 2 $\mathbf{P}_k$ *is K-Lipschitz and $\forall f \in \mathcal{C}_{b,Lip}^1$:*

$$\mathbf{P}_k f \in \mathcal{C}_{b,Lip}^1 \quad \text{and} \quad [\mathrm{D}\mathbf{P}_k f]_{Lip} \le K(\|\mathrm{D}f\|_\infty \vee [\mathrm{D}f]_{Lip}).$$

Remark 1.2.6 Notice that under assumption **H**2, for $f \in \mathcal{C}_{b,Lip}^1$ we have:

$$\|\mathrm{D}\mathbf{P}_k f\|_\infty = [\mathbf{P}_k f]_{Lip} \le K[f]_{Lip} = K\|\mathrm{D}f\|_\infty$$

Under these assumptions, we can see that $\forall f \in \mathcal{C}_{b,Lip}^1$, $\forall 0 \le k \le n-1$, $R_k f$ defined recursively by (1.2.6), is differentiable and:

$$\mathrm{D}R_k f = \mathrm{D}g_k \mathbf{P}_k R_{k+1} f + g_k \mathrm{D}\mathbf{P}_k R_{k+1} f \tag{1.2.17}$$

So, we can establish the following proposition, using a backward induction:

Proposition 1.2.2 *Assuming* **H1** *and* **H2** *involves:*
$\forall f \in \mathcal{C}^1$ *such that* $H_n^n f \in \mathcal{C}^1_{b,Lip}$, *we have* $\forall 0 \leq k \leq n-1$, $R_k f \in \mathcal{C}^1_{b,Lip}$.
Furthermore:

$$\begin{aligned}
\|R_k f\|_\infty &\leq L^{n-k}\|H_n^n f\|_\infty \\
\|\mathrm{D}R_k f\|_\infty &\leq (LK)^{n-k}\|\mathrm{D}H_n^n f\|_\infty + L^{n-k}\|H_n^n f\|_\infty \frac{K^{n-k}-1}{K-1} \\
u_k &:= \|\mathrm{D}R_k f\|_\infty \vee [\mathrm{D}R_k f]_\infty \\
&\leq (3LK)^{n-k}u_n + L^{n-k}\|H_n^n f\|_\infty \frac{(3K)^{n-k}-1}{3K-1}
\end{aligned}$$

with the convention $\frac{K^m-1}{K-1} = m$ when $K = 1$.

Proof. The proof is based on an induction on k. Suppose for a given $0 \leq k \leq n-1$, $R_{k+1}f \in \mathcal{C}^1_{b,Lip}$.
(Notice that $H_n^n f \in \mathcal{C}^1_{b,Lip}$ by assumption).
By definition, we have $R_k f = g_k \mathbf{P}_k R_{k+1} f$.
According to **H**1 and **H**2, we can establish easily that $R_k f \in \mathcal{C}^1_{b,Lip}$, through a backward induction.
Furthermore,

$$\begin{aligned}
\|R_k f\|_\infty &\leq L\|\mathbf{P}_k R_{k+1} f\|_\infty \\
&\leq L\|R_{k+1} f\|_\infty
\end{aligned} \tag{1.2.18}$$

From (1.2.17) and Remark 1.2.6, we have also:

$$\begin{aligned}
\|\mathrm{D}R_k f\|_\infty &\leq L\|\mathbf{P}_k R_{k+1} f\|_\infty + L\|\mathrm{D}\mathbf{P}_k R_{k+1} f\|_\infty \\
&\leq L\|R_{k+1} f\|_\infty + LK\|\mathrm{D}R_{k+1} f\|_\infty
\end{aligned} \tag{1.2.19}$$

In addition,

$$\begin{aligned}
[\mathrm{D}R_k f]_{Lip} &\leq L\left(\|R_{k+1} f\|_\infty + K\|\mathrm{D}R_{k+1} f\|_\infty \right. \\
&\quad \left. + Ku_{k+1} + K\|\mathrm{D}R_{k+1} f\|_\infty\right)
\end{aligned} \tag{1.2.20}$$

where $u_{k+1} := \|\mathrm{D}R_{k+1} f\|_\infty \vee [\mathrm{D}R_{k+1} f]_{Lip}$.
Noticing from (1.2.19) that also:

$$\|\mathrm{D}R_k f\|_\infty \leq L\left(\|R_{k+1} f\|_\infty + K\|\mathrm{D}R_{k+1} f\|_\infty + Ku_{k+1} + K\|\mathrm{D}R_{k+1} f\|_\infty\right),$$

we have:

$$\begin{aligned}
u_k &\leq L\left(\|R_{k+1} f\|_\infty + K\|\mathrm{D}R_{k+1} f\|_\infty + Ku_{k+1} + K\|\mathrm{D}R_{k+1} f\|_\infty\right) \\
&\leq 3LKu_{k+1} + L\|R_{k+1} f\|_\infty.
\end{aligned} \tag{1.2.21}$$

Recursively we conclude the announced result. □
Now, applying the previous idea (from equations (1.2.15) and (1.2.16)) to the sequential

filter estimation via quantization, when $H_n^n f \in \mathcal{C}^1_{b,Lip}$, a *generic first order scheme* can be designed as follows:

$$\begin{cases} \widehat{R}_n f(\hat{X}_n) &= H_n^n f(\hat{X}_n), \\ \widehat{DR}_n f(\hat{X}_n) &= \mathrm{D}H_n^n f(\hat{X}_n), \\ \widehat{R}_k f(\hat{X}_k) &= g_k(\hat{X}_k)\mathbb{E}[\widehat{R}_{k+1} f(\hat{X}_{k+1}) + \langle \widehat{DR}_{k+1} f(\hat{X}_{k+1}), \Delta_{k+1}\rangle | \hat{X}_k], \\ & \quad 0 \le k \le n-1. \end{cases} \tag{1.2.22}$$

and then, $\hat{\pi}_n f = \hat{\mu}_0 \widehat{R}_0 f$.

In (1.2.22), $\widehat{DR}_k f$ is a quantization based estimate for $\mathrm{D}R_k f$. It needs to be specified to transform the above scheme into an implementable algorithm. In [43], the scheme (1.2.22) is introduced with no computational considerations concerning $\mathrm{D}R_k f$. It is shown that under assumptions **H**2 and **H**1, the quantization based unnormalized filter converges toward $\pi_n f$ at a rate $\sum_{k=1}^n \|\Delta_k\|_2^2$ (instead of $\sum_{k=1}^n \|\Delta_k\|_2$ in the original zero order scheme from [41]).
Our aim is to propose some estimate $\widehat{DR}_k f$ for $\mathrm{D}R_k f$, in order to combine computability skills and convergence rate improvement. In this aim, two methods will be exhibited:

- the first one is based on an induction: at each time step k we evaluate $\{\widehat{DR}_k, \widehat{R}_k\}$ using $\{\widehat{DR}_{k+1}, \widehat{R}_{k+1}\}$. This approach leads to a one step recursive scheme and is investigated in Section 3;
- the second one is based on an integration by parts following an approach developed in [6]: the operator $\widehat{DR}_k$ is defined as a weighted expectation of $\widehat{R}_k$. The scheme constructed by plugging $\widehat{DR}_k f$ expression in (1.2.22) leads to a two step recursive scheme, details are investigated in Section 4.

1.3 One step first order iterative scheme

We introduce for this section the following assumption, in the spirit of **H2**, but in fact a bit more restrictive:

H 2' For each $1 \le k \le n$, F_k admits a bounded, uniformly Lipschitz derivative with respect to its first variable. Namely, $\forall x, x' \in \mathbb{R}^d$, $\forall \varepsilon \in \mathbb{R}^d$:

$$|\partial_x F_k(x,\varepsilon) - \partial_x F_k(x',\varepsilon)| \le [\partial_x F_k]^1_{Lip}|x - x'| \quad \textit{and} \quad \|\partial_x F\|_\infty := \max_{1\le k\le n} \|\partial_x F_k\|_\infty < +\infty.$$

Example 1.3.1 This assumption is e.g. satisfied by dynamics with an *additive noise*, typically for functions $F_k : (x,u) \longmapsto b_k(x) + \sigma_k u$, where b_k is differentiable with bounded Lipschitz continuous derivative and $\sigma_k \in \mathcal{M}(d,q)$, or by dynamics where F_k satisfies: $F_k(x,u) = b_k(x) + \sigma_k(x)u$, b_k, σ_k being differentiable with bounded Lipschitz continuous derivatives, applied to signal innovations ε_k with compactly supported pdf.

1.3.1 Definition of the scheme

In this paragraph, we investigate the recursive approach to estimate $\mathrm{D}R_k$. Under **H2'**, the probability transitions $\mathbf{P}_k$ are K-Lipschitz with $K = \|\partial_x F\|_\infty$. Furthermore, the $\mathbf{P}_k$ are differentiable in the following sense: for every $f \in \mathcal{C}^1_{b,Lip}$,

$$\mathrm{D}\mathbf{P}_k f = Q_k \mathrm{D}f, \quad k = 0, \dots, n-1, \tag{1.3.1}$$

where, for every Borel map $\varphi : \mathbb{R}^d \to \mathbb{R}^d$,

$$Q_k\varphi(x) \;=\; \mathbb{E}[\partial_x F_{k+1}(X_k, \varepsilon_{k+1})'\varphi(X_{k+1})|X_k = x], \text{ for } x \in \mathbb{R}^d. \tag{1.3.2}$$

The quantization based estimate for $\mathrm{D}\mathbf{P}_k f$ is then naturally defined by:

$$\hat{Q}_k \mathrm{D}f(x_k^i) \;=\; \mathbb{E}[\partial_x F_{k+1}(X_k, \varepsilon_{k+1})'\mathrm{D}f(\hat{X}_{k+1})|\hat{X}_k = x_k^i], \text{ for } i = 1, \dots, N_k. \tag{1.3.3}$$

Finally, following equation (1.2.17) one sets:

$$\widehat{DR}_k f(x_k^i) = \mathrm{D}g_k(x_k^i)\widehat{\mathbf{P}}_k \hat{\widehat{R}}_{k+1} f(x_k^i) + g_k(x_k^i)\hat{Q}_k \mathrm{D}f(x_k^i) \tag{1.3.4}$$

as a zero order approximation of $\mathrm{D}R_k f$ defined on $\Gamma_k = \{x_k^1, \dots, x_k^{N_k}\}$, for any $k \in \{1, \dots, n-1\}$.

Remark 1.3.1 From a numerical point of view, it would be more natural to use $\widehat{\mathbf{P}}_k \widehat{R}_{k+1}$ instead of $\widehat{\mathbf{P}}_k \hat{\widehat{R}}_{k+1}$. In fact, the algorithm structure would be less complex. Our choice in (1.3.4) is motivated on one hand by theoretical need to take a zero order approximation for the differential term estimator. On the other hand, using $\widehat{\mathbf{P}}_k \widehat{R}_{k+1}$ will introduce distortion terms in both $\widehat{DR}_k$ and $\widehat{R}_k$ which generates important numerical instability as emphasized by numerical tests in Figure 1.5.

Now, plugging (1.3.4) into the generic first order scheme (1.2.22) yields the following first order scheme:

Scheme B : Backward expression

$$\begin{cases} \widehat{R}_n f(\hat{X}_n) = H_n^n f(\hat{X}_n), \\ \widehat{DR}_n(\hat{X}_n) f = \mathrm{D}H_n^n f(\hat{X}_n), \\ \widehat{R}_k f(\hat{X}_k) = g_k(\hat{X}_k)\mathbb{E}[\widehat{R}_{k+1} f(\hat{X}_{k+1}) + \langle \widehat{DR}_{k+1} f(\hat{X}_{k+1}), \Delta_{k+1}\rangle|\hat{X}_k], \\ \widehat{DR}_k f(\hat{X}_k) = \mathrm{D}g_k(\hat{X}_k)\mathbb{E}[\hat{\widehat{R}}_{k+1} f(\hat{X}_{k+1})|\hat{X}_k] + g_k(\hat{X}_k)\hat{Q}_k \widehat{DR}_{k+1} f(\hat{X}_k) \\ \quad k = 0, \dots, n-1. \end{cases} \tag{1.3.5}$$

Note that this scheme is completely computable, as it can be rewritten easily using finite weighted sums. The quantizers Γ_k and the weights - which we call from now on *companion parameters* - can be computed off line and stored in an accessible codebook, so

that the only on line computation cost will be the calculus of operators $\widehat{R}_k$, $\hat{\widehat{R}}_k$ and $\widehat{DR}_k$. The scheme can be reformulated *in distribution* as follows:

Scheme B

$$\begin{cases} \hat{\widehat{R}}_n f(x_n^i) = H_n^n f(x_n^i), \quad i = 1, \dots, N_n, \\ \widehat{R}_n f(x_n^i) = H_n^n f(x_n^i), \quad i = 1, \dots, N_n, \\ \widehat{DR}_n f(x_n^i) = \mathrm{D}H_n^n f(x_n^i), \quad i = 1, \dots, N_n, \\ \hat{\widehat{R}}_k f(x_k^i) = g_k(x_k^i) \sum_{j=1}^{N_{k+1}} \hat{\widehat{R}}_{k+1} f(x_{k+1}^j) p_k^{ij}, \\ \widehat{R}_k f(x_k^i) = g_k(x_k^i) \sum_{j=1}^{N_{k+1}} \left(\widehat{R}_{k+1} f(x_{k+1}^j) p_k^{ij} + \langle \widehat{DR}_{k+1} f(x_{k+1}^j), \delta_k^{ij} \rangle \right) \\ \widehat{DR}_k f(x_k^i) = \mathrm{D}g_k(x_k^i) \sum_{j=1}^{N_{k+1}} \hat{\widehat{R}}_{k+1} f(x_{k+1}^j) p_k^{ij} + g_k(x_k^i) \sum_{j=1}^{N_{k+1}} \gamma_k^{ij} \widehat{DR}_{k+1} f(x_{k+1}^j), \\ \quad i = 1, \dots, N_k, \quad 0 \le k < n, \end{cases} \tag{1.3.6}$$

where the companion parameters, p_k^{ij}, γ_k^{ij}, and δ_k^{ij} are defined by:

$$\begin{aligned} p_k^{ij} &= \mathbb{E}[\mathbf{1}_{\{\hat{X}_{k+1} = x_{k+1}^j\}} | \hat{X}_k = x_k^i] \in \mathbb{R}, \\ \gamma_k^{ij} &= \mathbb{E}[\partial_x F_k(X_k, \varepsilon_{k+1})' \mathbf{1}_{\{\hat{X}_{k+1} = x_{k+1}^j\}} | \hat{X}_k = x_k^i] \in \mathcal{M}_d(\mathbb{R}) \\ \delta_k^{ij} &= \mathbb{E}[\Delta_{k+1} \mathbf{1}_{\{\hat{X}_{k+1} = x_{k+1}^j\}} | \hat{X}_k = x_k^i], \\ &= \mathbb{E}[(X_{k+1} - x_{k+1}^j) \mathbf{1}_{\{\hat{X}_{k+1} = x_{k+1}^j\}} | \hat{X}_k = x_k^i] \in \mathbb{R}^d. \end{aligned} \tag{1.3.7}$$

Forward expression of scheme **B**

For applications, it is crucial in terms of computational efficiency, to rewrite the scheme in a forward way. This allows us to compute costlessly intermediate estimations of $\pi_k f$, $1 \le k \le n-1$, and to use different test functions f without recomputing the hole scheme. This forward form can be established as follows: one first checks that at each $0 \le k \le n-1$, the vector $\begin{bmatrix} \hat{\widehat{R}}_k \\ \widehat{DR}_k \\ \widehat{R}_k \end{bmatrix}$ satisfies the following one step induction:

$$\begin{bmatrix} \hat{\widehat{R}}_k \\ \widehat{DR}_k \\ \widehat{R}_k \end{bmatrix} = \widehat{\mathcal{H}}_k \begin{bmatrix} \hat{\widehat{R}}_{k+1} \\ \widehat{DR}_{k+1} \\ \widehat{R}_{k+1} \end{bmatrix}, \tag{1.3.8}$$

where $\widehat{\mathcal{H}}_k$ is a lower triangular operator matrix defined by: $\widehat{\mathcal{H}}_k = \begin{pmatrix} \hat{H}_k^1 & 0 & 0 \\ \hat{H}_k^2 & \hat{H}_k^3 & 0 \\ 0 & \hat{H}_k^4 & \hat{H}_k^1 \end{pmatrix}$,

with for $f : \mathbb{R}^d \to \mathbb{R}$ and $\varphi : \mathbb{R}^d \to \mathbb{R}^d$,

$$\begin{aligned} \hat{H}_0^1 f(x) &= \mathbb{E}[f(\hat{X}_1)|\hat{X}_0 = x], \\ \hat{H}_0^2 f(x) &= 0 \in \mathbb{R}^d, \\ \hat{H}_0^3 \varphi(x) &= \mathbb{E}[\partial_x F_1(x, \varepsilon_1)' \varphi(\hat{X}_1)|\hat{X}_0 = x], \\ \hat{H}_0^4 \varphi(x) &= \mathbb{E}[\langle \varphi(\hat{X}_1), \Delta_1 \rangle |\hat{X}_0 = x], \end{aligned}$$

and for every $1 \le k < n$

$$\begin{aligned} \hat{H}_k^1 f(\hat{X}_k) &= g_k(\hat{X}_k)\mathbb{E}[f(\hat{X}_{k+1})|\hat{X}_k], \\ \hat{H}_k^2 f(\hat{X}_k) &= \mathrm{D}g_k(\hat{X}_k)\mathbb{E}[f(\hat{X}_{k+1})|\hat{X}_k], \\ \hat{H}_k^3 \varphi(\hat{X}_k) &= g_k(\hat{X}_k)\mathbb{E}[\partial_x F_{k+1}(X_k, \varepsilon_{k+1})' \varphi(\hat{X}_{k+1})|\hat{X}_k], \\ \hat{H}_k^4 \varphi(\hat{X}_k) &= g_k(\hat{X}_k)\mathbb{E}[\langle \varphi(\hat{X}_{k+1}), \Delta_{k+1} \rangle |\hat{X}_k]. \end{aligned}$$

Notice that here $\hat{H}_k^1 = \hat{\hat{H}}_k$.
Then, one can see from (1.3.8) that:

$$\begin{bmatrix} \hat{\hat{R}}_0 \\ \widehat{DR}_0 \\ \hat{R}_0 \end{bmatrix} = \widehat{\mathcal{H}}_0 \circ \widehat{\mathcal{H}}_1 \circ \cdots \circ \widehat{\mathcal{H}}_{n-1} \begin{bmatrix} H_n^n \\ \mathrm{D}H_n^n \\ H_n^n \end{bmatrix}.$$

Setting $\hat{\mathcal{U}}_k = \hat{\mu}_0 \circ \widehat{\mathcal{H}}_0 \circ \cdots \circ \widehat{\mathcal{H}}_k$, the forward scheme satisfies the following recursive formula:

$$\hat{\mathcal{U}}_0 = \hat{\mu}_0 \widehat{\mathcal{H}}_0 \quad \textit{and} \quad \hat{\mathcal{U}}_k = \hat{\mathcal{U}}_{k-1} \widehat{\mathcal{H}}_k \quad k = 1, \ldots, n-1,$$

so that $\hat{\pi}_n f = \langle \hat{\mathcal{U}}_{n-1} \begin{bmatrix} H_n^n f \\ \mathrm{D}H_n^n f \\ H_n^n f \end{bmatrix}, e_3 \rangle$.

1.3.2 Error bounds

The main result of this section is to establish a convergence result for scheme **B** better than the zero scheme rate. We recall that here, $(\hat{X}_k)$ is a marginal, stationary (N_k)-quantization of (X_k).

Theorem 1.3.1 *Assume* **H1** *and* **H2'** *and let f satisfying $H_n^n f \in \mathcal{C}^1_{b,Lip}$. Then, there exists a sequence of positive real constants $(M_j^n)_{0 \le j \le n}$ such that:*

$$|\pi_n f - \hat{\pi}_n f| \le \sum_{j=0}^{n} M_j^n \|\Delta_j\|_{2p}^2$$

with $M_j^n \le \alpha(p, f) \frac{n+5}{2} L^n \left(\frac{(LK)^{j+1}-1}{LK-1}\right)\left(\frac{(3K)^{n-j+1}-1}{3K-1}\right)\left(\frac{(L)^{j+1}-1}{L-1}\right)$.

The key to prove the above error bound is to rely on the *backward* form of the scheme **B** (see (1.3.6)). The main technical step is to produce some error upper bounds for the differential term approximation, namely $\hat{A}_k = \mathrm{D}R_k f(\hat{X}_k) - \widehat{DR}_k f(\hat{X}_k)$.

The proof of the first lemma below is left to the reader:

Lemma 1.3.1 *For any φ bounded Lipschitz continuous, $Q_k\varphi : \mathbb{R}^d \to \mathbb{R}^d$ is Lipschitz and*

$$[Q_k\varphi]_{Lip} \leq \|\varphi\|_\infty [\partial_x F_{k+1}]^1_{Lip} + \|\partial_x F\|^2_\infty [\varphi]_{Lip}.$$

Then, the error bounds for $\|\hat{A}_k\|_p$ are given in the lemma:

Lemma 1.3.2 *For f satisfying $H^n_n f \in \mathcal{C}^1_{b,Lip}$, there exists a non negative real sequence $(D^{k,n}_j)_{0\leq k\leq j\leq n}$ such that:*

$$\|\hat{A}_k\|_p \leq \sum_{j=k}^{n} D^{k,n}_j \|\Delta_j\|_p$$

where $D^{k,n}_j \leq \alpha(p,f)L^{n-k}(\frac{(LK)^{j-k+1}-1}{LK-1})(\frac{K^{n-j+1}-1}{K-1})$.

Proof.
From equations (1.2.17), (1.3.4) and (1.3.1):

$$\begin{aligned}
\hat{A}_k &= \mathrm{D}g_k(\hat{X}_k)\mathbf{P}_k R_{k+1} f(\hat{X}_k) + g_k(\hat{X}_k) Q_k \mathrm{D}R_{k+1} f(\hat{X}_k) \\
&\quad -\mathrm{D}g_k(\hat{X}_k)\mathbb{E}[\hat{\hat{R}}_{k+1} f(\hat{X}_{k+1})|\hat{X}_k] - g_k(\hat{X}_k)\hat{Q}_k \widehat{DR}_{k+1} f(\hat{X}_k) \\
&= \mathrm{D}g_k(\hat{X}_k)\left(\mathbf{P}_k R_{k+1} f(\hat{X}_k) - \mathbb{E}[\mathbf{P}_k R_{k+1} f(X_k)|\hat{X}_k]\right) \\
&\quad +\mathrm{D}g_k(\hat{X}_k)\left(\mathbb{E}[\mathbf{P}_k R_{k+1} f(X_k)|\hat{X}_k] - \mathbb{E}[\hat{\hat{R}}_{k+1} f(\hat{X}_{k+1})|\hat{X}_k]\right) \\
&\quad +g_k(\hat{X}_k)\left(Q_k \mathrm{D}R_{k+1} f(\hat{X}_k) - \mathbb{E}[Q_k \mathrm{D}R_{k+1} f(X_k)|\hat{X}_k]\right) \\
&\quad +g_k(\hat{X}_k)\left(\mathbb{E}[Q_k \mathrm{D}R_{k+1} f(X_k)|\hat{X}_k] - \hat{Q}_k \widehat{DR}_{k+1} f(\hat{X}_k)\right)
\end{aligned}$$

Then, using **H**1, one gets:

$$\begin{aligned}
\|\hat{A}_k\|_p &\leq L\|\mathbf{P}_k R_{k+1} f(\hat{X}_k) - \mathbb{E}[\mathbf{P}_k R_{k+1} f(X_k)|\hat{X}_k]\|_p \\
&\quad +L\|\mathbb{E}[\mathbf{P}_k R_{k+1} f(X_k)|\hat{X}_k] - \mathbb{E}[\hat{\hat{R}}_{k+1} f(\hat{X}_{k+1})|\hat{X}_k]\|_p \\
&\quad +L\|Q_k \mathrm{D}R_{k+1} f(\hat{X}_k) - \mathbb{E}[Q_k \mathrm{D}R_{k+1} f(X_k)|\hat{X}_k]\|_p \\
&\quad +L\|\mathbb{E}[Q_k \mathrm{D}R_{k+1} f(X_k)|\hat{X}_k] - \hat{Q}_k \widehat{DR}_{k+1} f(\hat{X}_k)\|_p. \qquad (1.3.9)
\end{aligned}$$

Now, the L^p-contraction property of conditional expectation implies that:

$$\begin{aligned}
\|\mathbf{P}_k R_{k+1} f(\hat{X}_k) - \mathbb{E}[\mathbf{P}_k R_{k+1} f(X_k)|\hat{X}_k]\|_p &\leq \|\mathbf{P}_k R_{k+1} f(\hat{X}_k) - \mathbf{P}_k R_{k+1} f(X_k)\|_p \\
&\leq [\mathbf{P}_k R_{k+1} f]_{Lip}\|\Delta_k\|_p \\
&\leq K\|DR_{k+1} f\|_\infty \|\Delta_k\|_p.
\end{aligned}$$

For the second term in the right handside of inequality (1.3.9), we will use on one hand the chaining rule for conditional expectation (see equation (1.2.15)) and on the other hand its L^p-contraction property, to write:

$$\begin{aligned}\|\mathbb{E}[\mathbf{P}_k R_{k+1} f(X_k)|\hat{X}_k] &- \mathbb{E}[\hat{\hat{R}}_{k+1} f(\hat{X}_{k+1})|\hat{X}_k]\|_p \\ &= \|\mathbb{E}[R_{k+1} f(X_{k+1})|\hat{X}_k] - \mathbb{E}[\hat{\hat{R}}_{k+1} f(\hat{X}_{k+1})|\hat{X}_k]\|_p \\ &\leq \|R_{k+1} f(X_{k+1}) - \hat{\hat{R}}_{k+1} f(\hat{X}_{k+1})\|_p\end{aligned}$$

which implies, by Theorem 1.2.1:

$$\|\mathbb{E}[\mathbf{P}_k R_{k+1} f(X_k)|\hat{X}_k] - \mathbb{E}[\hat{\hat{R}}_{k+1} f(\hat{X}_{k+1})|\hat{X}_k]\|_p \leq \sum_{j=k+1}^{n} C_j^{k+1,n} \|\Delta_j\|_p.$$

The same arguments on conditional expectations give:

$$\|Q_k \mathrm{D} R_{k+1} f(\hat{X}_k) - \mathbb{E}[Q_k \mathrm{D} R_{k+1} f(X_k)|\hat{X}_k]\|_p \leq [Q_k \mathrm{D} R_{k+1} f]_{Lip} \|\Delta_k\|_p,$$

which by Lemma 1.3.1 writes:

$$\begin{aligned}\|Q_k \mathrm{D} R_{k+1} f(\hat{X}_k) - \mathbb{E}[Q_k \mathrm{D} R_{k+1} f(X_k)|\hat{X}_k]\|_p \leq \\ \left(\|\mathrm{D} R_{k+1} f\|_\infty [\partial_x F_{k+1}]_{Lip}^1 + \|\partial_x F\|_\infty^2 [\mathrm{D} R_{k+1} f]_{Lip}\right) \|\Delta_k\|_p\end{aligned}$$

since $\mathrm{D} R_{k+1}$ is bounded Lipschitz by Proposition 1.2.2.
Then, using the definition of $\hat{Q}_k$ yields:

$$\begin{aligned}\|\mathbb{E}[Q_k \mathrm{D} R_{k+1} f(X_k)|\hat{X}_k] &- \hat{Q}_k \widehat{DR}_{k+1} f(\hat{X}_k)\|_p \\ &\leq \| \left(\partial_x F_{k+1}(X_k, \varepsilon_{k+1})\right)' \left(\mathrm{D} R_{k+1} f(X_{k+1}) - \widehat{DR}_{k+1} f(\hat{X}_{k+1})\right) \|_p \\ &\leq \|\partial_x F_{k+1}\|_\infty \left(\|\hat{A}_{k+1}\|_p + \|\mathrm{D} R_{k+1} f(X_{k+1}) - \mathrm{D} R_{k+1} f(\hat{X}_{k+1})\|_p\right) \\ &\leq \|\partial_x F_{k+1}\|_\infty \left(\|\hat{A}_{k+1}\|_p + [\mathrm{D} R_{k+1} f]_{Lip} \|\Delta_{k+1}\|_p\right).\end{aligned}$$

Finally, using $\|\partial_x F\|_\infty = K$ and Proposition 1.2.2, we derive:

$$\begin{aligned}\|\hat{A}_k\|_p \leq\ & L\left([\partial_x F_{k+1}]_{Lip}^1 \|\mathrm{D} R_{k+1} f\|_\infty + K^2 [\mathrm{D} R_{k+1} f]_{Lip} + K \|\mathrm{D} R_{k+1} f\|_\infty\right) \|\Delta_k\|_p \\ & + L\left(C_{k+1}^{k+1,n} + K([\mathrm{D} R_{k+1} f]_{Lip}\right) \|\Delta_{k+1}\|_p \\ & + L \sum_{j=k+2}^{n} C_j^{k+1,n} \|\Delta_j\|_p + LK \|\hat{A}_{k+1}\|_p. \qquad (1.3.10)\end{aligned}$$

The required result follows from a backward induction on k. See Appendix 1.6.3 for explicit upper bounds. □

Proof of Theorem 1.3.1.
Let $V_k f$ denote the intermediate estimation error when considering the previous first order approximation scheme $\mathbf{B}$ in its backward form : $V_k f := \mathbb{E}[R_k f(X_k)|\hat{X}_k] - \widehat{R}_k f(\hat{X}_k)$.
Using triangular inequalities, we isolate three error sources in $|V_k f|$. If we set:

$$\begin{aligned}\bar{R}_k f(\hat{X}_k) &= g_k(\hat{X}_k)\mathbb{E}[\mathbf{P}_k R_{k+1} f(X_k)|\hat{X}_k], \\ &= g_k(\hat{X}_k)\mathbb{E}[R_{k+1} f(X_{k+1})|\hat{X}_k],\end{aligned}$$

then we have:

$$\begin{aligned}|V_k f| \leq\ & |\mathbb{E}[R_k f(X_k)|\hat{X}_k] - R_k f(\hat{X}_k)| + |R_k f(\hat{X}_k) - \bar{R}_k f(\hat{X}_k)| \\ & + |\bar{R}_k f(\hat{X}_k) - \widehat{R}_k f(\hat{X}_k)|. \end{aligned} \tag{1.3.11}$$

Using a first order Taylor expansion, there exists $\hat{\zeta}_k^1 \in (\hat{X}_k, X_k)$ such that

$$\begin{aligned}\mathbb{E}[R_k f(X_k)|\hat{X}_k] &= \mathbb{E}[R_k f(\hat{X}_k) + \langle \mathrm{D}R_k f(\hat{\zeta}_k^1), \Delta_k\rangle|\hat{X}_k] \\ &= \mathbb{E}[R_k f(\hat{X}_k) + \langle \mathrm{D}R_k f(\hat{X}_k), \Delta_k\rangle + \langle \mathrm{D}R_k f(\hat{\zeta}_k^1) - \mathrm{D}R_k f(\hat{X}_k), \Delta_k\rangle|\hat{X}_k]\end{aligned}$$

$\hat{X}_k$ being a stationary quantization of X_k, one derives from Proposition 1.2.1 that:

$$\mathbb{E}[\langle \mathrm{D}R_k f(\hat{X}_k), \Delta_k\rangle|\hat{X}_k] = \langle \mathrm{D}R_k f(\hat{X}_k), \mathbb{E}[\Delta_k|\hat{X}_k]\rangle = 0.$$

Then,

$$\begin{aligned}|\mathbb{E}[R_k f(X_k)|\hat{X}_k] - R_k f(\hat{X}_k)| &= |\mathbb{E}[\langle \mathrm{D}R_k f(\hat{\zeta}_k^1) - \mathrm{D}R_k f(\hat{X}_k), \Delta_k\rangle|\hat{X}_k]| \\ &\leq \mathbb{E}[|\mathrm{D}R_k f(\hat{\zeta}_k^1) - \mathrm{D}R_k f(\hat{X}_k)||\Delta_k||\hat{X}_k] \\ &\leq [\mathrm{D}R_k f]_{Lip}\mathbb{E}[|\hat{X}_k - \zeta_k^1||\Delta_k||\hat{X}_k] \\ &\leq [\mathrm{D}R_k f]_{Lip}\mathbb{E}[|\Delta_k|^2|\hat{X}_k]. \end{aligned} \tag{1.3.12}$$

By Taylor expansion of $\mathbf{P}_k R_{k+1} f$, we analogically find $\hat{\zeta}_k^2 \in (\hat{X}_k, X_k)$ such that:

$$\begin{aligned}\bar{R}_k f(\hat{X}_k) &= g_k(\hat{X}_k)\Big(\mathbf{P}_k R_{k+1} f(\hat{X}_k) + \langle \mathrm{D}\mathbf{P}_k R_{k+1} f(\hat{X}_k), \mathbb{E}[\Delta_k|\hat{X}_k]\rangle \\ &\qquad + \mathbb{E}[\langle \mathrm{D}\mathbf{P}_k R_{k+1} f(\hat{\zeta}_k^2) - \mathrm{D}\mathbf{P}_k R_{k+1} f(\hat{X}_k), \Delta_k\rangle|\hat{X}_k]\Big) \\ R_k f(\hat{X}_k) - \bar{R}_k f(\hat{X}_k) &= g_k(\hat{X}_k)\mathbb{E}[\langle \mathrm{D}\mathbf{P}_k R_{k+1} f(\hat{\zeta}_k^2) - \mathrm{D}\mathbf{P}_k R_{k+1} f(\hat{X}_k), \Delta_k\rangle|\hat{X}_k]\end{aligned}$$

$$\begin{aligned}\text{Hence, } |R_k f(\hat{X}_k) &- \bar{R}_k f(\hat{X}_k)| \leq L[\mathrm{D}\mathbf{P}_k R_{k+1} f]_{Lip}\mathbb{E}[|\Delta_k|^2|\hat{X}_k] \\ &\leq LK\left([\mathrm{D}R_{k+1} f]_{Lip} \vee \|\mathrm{D}R_{k+1} f\|_\infty\right)\mathbb{E}[|\Delta_k|^2|\hat{X}_k] \end{aligned} \tag{1.3.13}$$

For the last term in the right handside of inequality (1.3.11), we have:

$$\begin{aligned}
|\bar{R}_k f(\hat{X}_k) - \widehat{R}_k f(\hat{X}_k)| &= \Big| g_k(\hat{X}_k) \Big(\mathbb{E}[R_{k+1} f(X_{k+1}) | \hat{X}_k] - \mathbb{E}[\widehat{R}_{k+1} f(\hat{X}_{k+1}) | \hat{X}_k] \\
&\quad - \mathbb{E}[\langle \widehat{DR}_{k+1} f(\hat{X}_{k+1}), \Delta_{k+1} \rangle | \hat{X}_k] \Big) \Big| \\
&\leq L \Big| \mathbb{E}\Big[R_{k+1} f(X_{k+1}) - \mathbb{E}[R_{k+1} f(X_{k+1}) | \hat{X}_{k+1}] \mid \hat{X}_k \Big] \\
&\quad - \mathbb{E}[\langle \widehat{DR}_{k+1} f(\hat{X}_{k+1}), \Delta_{k+1} \rangle | \hat{X}_k] \Big| \\
&\quad + L \Big| \mathbb{E}\Big[\mathbb{E}[R_{k+1} f(X_{k+1}) | \hat{X}_{k+1}] - \widehat{R}_{k+1} f(\hat{X}_{k+1}) | \hat{X}_k \Big] \Big|
\end{aligned}$$

Furthermore, there exists $\hat{\zeta}^3_{k+1} \in (\hat{X}_{k+1}, X_{k+1})$ such that

$$\begin{aligned}
R_{k+1} f(X_{k+1}) &= R_{k+1} f(\hat{X}_{k+1}) + \langle \mathrm{D} R_{k+1} f(\hat{X}_{k+1}), \Delta_{k+1} \rangle \\
&\quad + \langle \mathrm{D} R_{k+1} f(\hat{\zeta}^3_{k+1}) - \mathrm{D} R_{k+1} f(\hat{X}_{k+1}), \Delta_{k+1} \rangle \\
\mathbb{E}[R_{k+1} f(X_{k+1}) | \hat{X}_{k+1}] &= R_{k+1} f(\hat{X}_{k+1}) + \mathbb{E}[\langle \mathrm{D} R_{k+1} f(\hat{\zeta}^3_{k+1}) - \mathrm{D} R_{k+1} f(\hat{X}_{k+1}), \Delta_{k+1} \rangle | \hat{X}_{k+1}]
\end{aligned}$$

Consequently:

$$\begin{aligned}
|\bar{R}_k f(\hat{X}_k) - \widehat{R}_k f(\hat{X}_k)| &\leq L |\mathbb{E}[V_{k+1} f | \hat{X}_k]| \qquad (1.3.14) \\
&\quad + L |\mathbb{E}[\langle \Big(\mathrm{D} R_{k+1} f(\hat{X}_{k+1}) - \widehat{DR}_{k+1} f(\hat{X}_{k+1}) \Big), \Delta_{k+1} \rangle | \hat{X}_k]| \\
&\quad + L [\mathrm{D} R_{k+1} f]_{Lip} \Big(\mathbb{E}[|\Delta_{k+1}|^2 | \hat{X}_k] + \mathbb{E}\Big[\mathbb{E}[|\Delta_{k+1}|^2 | \hat{X}_{k+1}] \Big| \hat{X}_k \Big] \Big) \qquad (1.3.15)
\end{aligned}$$

Finally combining previous inequalities (1.3.12), (1.3.13), (1.3.14), we obtain by using L^p-contraction property of conditional expectation:

$$\begin{aligned}
\|V_k f\|_p &\leq ([\mathrm{D} R_k f]_{Lip} + L K u_{k+1}) \, \|\Delta_k\|^2_{2p} \\
&\quad + 2L [\mathrm{D} R_{k+1} f]_{Lip} \|\Delta_{k+1}\|^2_{2p} + L \|\langle \hat{A}_{k+1}, \Delta_{k+1} \rangle\|_p \qquad (1.3.16) \\
&\quad + L \|V_{k+1} f\|_p .
\end{aligned}$$

Applying Holder inequality combined to Lemma 1.3.2 to the term $\|\langle \hat{A}_{k+1}, \Delta_{k+1} \rangle\|_p$, we have:

$$\begin{aligned}
\|\langle \hat{A}_{k+1}, \Delta_{k+1} \rangle\|_p &\leq \| |\hat{A}_{k+1}| |\Delta_{k+1}| \|_p \\
&\leq \|\hat{A}_{k+1}\|_{2p} \|\Delta_{k+1}\|_{2p} \\
&\leq \sum_{j=k+1}^{n} D_j^{k+1,n} \|\Delta_j\|_{2p} \|\Delta_{k+1}\|_{2p} \\
&\leq \frac{1}{2} \sum_{j=k+1}^{n} D_j^{k+1,n} \left(\|\Delta_j\|^2_{2p} + \|\Delta_{k+1}\|^2_{2p} \right) \qquad (1.3.17)
\end{aligned}$$

Plugging (1.3.17) into (1.3.16) yields:

$$\begin{aligned}\|V_k f\|_p \;\leq\; & ([\mathrm{D}R_k f]_{Lip} + LKu_{k+1})\,\|\Delta_k\|_{2p}^2 \\ & + L\left(2[\mathrm{D}R_{k+1}f]_{Lip} + \tfrac{1}{2}\textstyle\sum_{j=k+1}^{n} D_j^{k+1,n}\right)\|\Delta_{k+1}\|_{2p}^2 \\ & + \tfrac{1}{2}L\textstyle\sum_{j=k+1}^{n} D_j^{k+1,n}\|\Delta_j\|_{2p}^2 + L\|V_{k+1}f\|_p.\end{aligned} \tag{1.3.18}$$

Then, by induction taking $k=0$ and writing $|\pi_n f - \hat{\pi}_n f| \leq \|V_0 f\|_p$ we derive the required result. See Appendix 1.6.3 for further details. □

Theorem 1.3.1, with $\|\Delta_k\|_{2p}^2 = O(\|\Delta_k\|_{2p})$, shows that the scheme **B** succeeds to embetter the zero-order convergence rate. The forthcoming section has been motivated by our wish to relax **H2'** and to preserve the convergence rate improvement.

1.4 Two step iterative first order scheme

To construct this second first order scheme, the idea is to represent $\mathrm{D}\mathbf{P}_k R_{k+1} f$ as a weighted conditional expectation of $R_{k+1}f$ i.e.

$$\mathrm{D}\mathbf{P}_k R_{k+1} f(x) = \mathbb{E}[R_{k+1}f(X_{k+1}) \times Weight | X_k = x],$$

and then to quantize this representation formula. This is achieved classically by the mean of an integration by parts formula.
Note that in all this section, we will assume $q = d$. Furthermore, F_k will be supposed to be differentiable.

1.4.1 Integration by parts formula

For notational convenience, we will temporarily drop the k indices in the notations of X_k, F_k and $\mathbf{P}_k$. We will also temporarily assume $f \in \mathcal{C}_b^1$.
We start by a transformation of the problem, via differentiation. For that, we need first to assume the following:

H 3 $\forall\ 0 \leq k \leq n$, $\exists\ c_k > 0$ *such that for any* $x \in \mathbb{R}^d$ *and* $\varepsilon \in \mathbb{R}^d$*:*

$$(\partial_\varepsilon F_k(x,\varepsilon))(\partial_\varepsilon F_k(x,\varepsilon))' \geq c_k I_d.$$

We have then, for any $x, \varepsilon \in \mathbb{R}^d$:

$$\begin{aligned}\partial_x (foF)(x,\varepsilon) &= \partial_x F(x,\varepsilon)'(\mathrm{D}f)oF(x,\varepsilon), \\ \partial_\varepsilon (foF)(x,\varepsilon) &= \partial_\varepsilon F(x,\varepsilon)'(\mathrm{D}f)oF(x,\varepsilon).\end{aligned}$$

Assuming **H**3 yields $\partial_x(foF)(x,\varepsilon) = \mathcal{G}_x(\varepsilon)\partial_\varepsilon(foF)(x,\varepsilon)$, where:

$$\begin{aligned}\mathcal{G}_x : \mathbb{R}^d &\to \mathcal{M}_d(\mathbb{R}) \\ \varepsilon &\mapsto \left(\partial_\varepsilon F(x,\varepsilon)^{-1}\partial_x F(x,\varepsilon)\right)'.\end{aligned}$$

Now, in order to allow a differentiation under the integral sign and then apply integration by parts, we will assume the following technical hypothesis:

H 4 Assume that signal innovations ε_k distribution is absolutely continuous toward Lebesgue measure, with a differentiable density $\mathbf{p}$ satisfying for all $x \in \mathbb{R}^d$,

$$\int_{\mathbb{R}^d} |\partial_x F(x,\varepsilon)|\mathbf{p}(\varepsilon)d\varepsilon < +\infty \quad \text{and} \quad \lim_{|\varepsilon|\to+\infty} \mathcal{G}_x(\varepsilon)\mathbf{p}(\varepsilon) = 0.$$

Then, the i-th component of $\mathrm{D}\mathbf{P}f(x)$ for a given $1 \le i \le d$ reads:

$$\frac{\partial \mathbf{P}f}{\partial x^i}(x) = \int_{\mathbb{R}^d} \langle \mathcal{G}_x^i(\varepsilon), \partial_\varepsilon (foF)(x,\varepsilon)\rangle \mathbf{p}(\varepsilon)d\varepsilon \tag{1.4.1}$$

$$\begin{array}{rrcl} \text{where:} & \mathcal{G}_x^i : \mathbb{R}^d & \to & \mathbb{R}^d \\ & \varepsilon & \mapsto & (\mathcal{G}_x(\varepsilon))' e_i. \end{array}$$

Furthermore, performing an integration by parts formula on (1.4.1), and taking in account **H**4 yields:

$$\frac{\partial \mathbf{P}f}{\partial x^i}(x) = -\int_{\mathbb{R}^q} (foF(x,\varepsilon) + C(x))\Psi^i(x,\varepsilon)\mathbf{p}(\varepsilon)d\varepsilon \tag{1.4.2}$$

$$\begin{array}{rrcl} \text{where:} & \Psi^i : \mathbb{R}^d \times \mathbb{R}^d & \to & \mathbb{R} \\ & (x,\varepsilon) & \mapsto & \text{div } \mathcal{G}_x^i(\varepsilon) + \dfrac{1}{\mathbf{p}(\varepsilon)} \langle \mathcal{G}_x^i(\varepsilon), \mathrm{D}\mathbf{p}(\varepsilon)\rangle. \end{array}$$

Finally, defining the weight vector $\Psi(x,\varepsilon) := (\Psi^i(x,\varepsilon))_{0\le i\le d}$, we obtain the generalization of equation (1.4.2): $\mathrm{D}(\mathbf{P}_k f)(x) = -\mathbb{E}[\big(f(F_{k+1}(x,\varepsilon_{k+1})) + C^k(x)\big)\,\Psi_k(x,\varepsilon_{k+1})]$.
In a Monte Carlo method context, the constant C^k is tuned in order to minimize the variance of a probabilistic estimator of $\mathrm{D}(\mathbf{P}_k f)(x)$. In our quantization context, as the variance problem does not occur, a natural value for C^k would be zero. It is at least the choice that minimizes computation cost and provides satisfactory numerical results (see [7] for a discussion about C^k for an American option pricing problem).
From now on, we will take $C^k = 0$.

1.4.2 Numerical scheme

Consider now a test function f satisfying $H_n^n f \in \mathcal{C}^1_{b,Lip}$. Then, according to Proposition 1.2.2, $R_k f \in \mathcal{C}^1_{b,Lip}$. Using results of the previous paragraph, we can write, for each $0 \le k \le n-1$: $\mathrm{D}\mathbf{P}_k R_{k+1} f(x) = -\mathbb{E}[R_{k+1}f(X_{k+1})\Psi_{k+1}(x,\varepsilon_{k+1})|X_k = x]$.
So, $(\hat{X}_k)$ still being a stationary marginal (N_k)-quantization of (X_k), an approximation of $\mathrm{D}R_k f$ would be:

$$\widehat{DR}_k f(\hat{X}_k) = \mathrm{D}g_k(\hat{X}_k)\mathbb{E}[\hat{\hat{R}}_{k+1}f(\hat{X}_{k+1})|\hat{X}_k] - g_k(\hat{X}_k)\mathbb{E}[\hat{\hat{R}}_{k+1}f(\hat{X}_{k+1})\Psi_k(X_k,\varepsilon_{k+1})|\hat{X}_k].$$

If one replaces this expression in (1.2.22), it results in the following two step recursive scheme formulated in a backward way:

Scheme A Backward formulation

$$\left\{\begin{array}{lcl}
\widehat{R}_n f(\hat{X}_n) & = & H_n^n f(\hat{X}_n), \\
\widehat{R}_{n-1} f(\hat{X}_{n-1}) & = & g_{n-1}(\hat{X}_{n-1})\mathbb{E}[H_n^n f(\hat{X}_n) + \langle \mathrm{D}H_n^n f(\hat{X}_n), \Delta_n\rangle | \hat{X}_{n-1}], \\
\widehat{R}_k f(\hat{X}_k) & = & g_k(\hat{X}_k)\widehat{\mathbf{P}}_k \widehat{R}_{k+1} f(\hat{X}_k) + g_k(\hat{X}_k)\times \\
 & & \Big(\mathbb{E}[\langle \mathrm{D}g_{k+1}(\hat{X}_{k+1})\widehat{\mathbf{P}}_{k+1}\hat{\widehat{R}}_{k+2} f(\hat{X}_{k+1}), \Delta_{k+1}\rangle | \hat{X}_k] - \mathbb{E}[\,\langle g_{k+1}(\hat{X}_{k+1})\times \\
 & & \mathbb{E}[\hat{\widehat{R}}_{k+2} f(\hat{X}_{k+2})\Psi_{k+1}(X_{k+1}, \varepsilon_{k+2}) | \hat{X}_{k+1}], \Delta_{k+1}\rangle | \hat{X}_k]\Big), \\
 & & 0 \le k \le n-2.
\end{array}\right. \tag{1.4.3}$$

This scheme **A** can be rewritten *in distribution* using finite weighted sums. As for the previous scheme, the weights are to be computed simultaneously with the optimal quantizers. Consequently, the implemented algorithm reads as follows:

Scheme A

$$\left\{\begin{array}{lcl}
\hat{\widehat{R}}_n f(x_n^i) & = & H_n^n f(x_n^i), \quad i = 1, \ldots, N_n, \\
\widehat{R}_n f(x_n^i) & = & H_n^n f(x_n^i), \quad i = 1, \ldots, N_n, \\
\widehat{R}_{n-1} f(x_{n-1}^i) & = & g_{n-1}(x_{n-1}^i) \sum_{j=1}^{N_n} \Big(H_n^n f(x_n^j) p_{n-1}^{ij} + \langle \mathrm{D}H_n^n f(x_n^j), \delta_{n-1}^{ij}\rangle\Big), \\
 & & i = 1, \ldots, N_{n-1}, \\
\hat{\widehat{R}}_k f(x_k^i) & = & g_k(x_k^i) \sum_{j=1}^{N_{k+1}} \hat{\widehat{R}}_{k+1} f(x_{k+1}^j) p_k^{ij}, \quad i = 1, \ldots, N_k, \quad 0 \le k < n, \\
\widehat{R}_k f(x_k^i) & = & g_k(x_k^i) \sum_{j=1}^{N_{k+1}} \widehat{R}_{k+1} f(x_{k+1}^j) p_k^{ij} + g_k(x_k^i)\times \\
 & & \sum_{j=1}^{N_{k+1}} \sum_{l=1}^{N_{k+2}} \Big(\hat{\widehat{R}}_{k+2} f(x_{k+2}^l) p_{k+1}^{jl} \langle \mathrm{D}g_k(x_{k+1}^j), \delta_k^{ij}\rangle \\
 & & - g_{k+1}(x_{k+1}^j) \hat{\widehat{R}}_{k+2} f(x_{k+2}^l) \langle \lambda_{k+1}^{jl}, \delta_k^{ij}\rangle\Big), \\
 & & i = 1, \ldots, N_k, \quad 0 \le k \le n-2
\end{array}\right. \tag{1.4.4}$$

where the companion parameters, p_k^{ij}, λ_k^{ij}, and δ_k^{ij} are defined by:

$$p_k^{ij} = \mathbb{E}[\mathbf{1}_{\{\hat{X}_{k+1} = x_{k+1}^j\}} | \hat{X}_k = x_k^i] \in \mathbb{R}, \tag{1.4.5}$$

$$\lambda_k^{ij} = \mathbb{E}[\Psi_k(X_k, \varepsilon_{k+1})\mathbf{1}_{\{\hat{X}_{k+1} = x_{k+1}^j\}} | \hat{X}_k = x_k^i] \in \mathbb{R}^d, \tag{1.4.6}$$

$$\begin{aligned}
\delta_k^{ij} &= \mathbb{E}[\Delta_{k+1}\mathbf{1}_{\{\hat{X}_{k+1} = x_{k+1}^j\}} | \hat{X}_k = x_k^i], \\
&= \mathbb{E}[(X_{k+1} - x_{k+1}^j)\mathbf{1}_{\{\hat{X}_{k+1} = x_{k+1}^j\}} | \hat{X}_k = x_k^i] \in \mathbb{R}^d.
\end{aligned} \tag{1.4.7}$$

It is important to recall, that the interest of such an approach lies in the possibility of carrying out the computation of all the above companion parameters off line. Once the state equations are fixed and the noise distribution is simulatable, the quantizers and companion parameters can be kept off line. On line computation cost will then be reduced

to the sequential determination of $\hat{\hat{R}}_k$ and $\widehat{R}_k$ on each grid. Compared to the previous case, scheme **A** is more demanding in on line memory capacity, as it involves two step computations, but, it is worth noting that the new companion parameters λ_k^{ij} are of lower dimension, which compensates the two step recursion effect while considering the algorithm complexity or the storage capacity dedicated to codebooks.

Here also, as for scheme **B** , we can see that this backward definition can be rewritten in a forward form. For $0 \leq k \leq n$, let $\hat{H}_k$ be the operator defined on any function $f : \Gamma_{k+2} \to \mathbb{R}$, such that:

$$\begin{aligned}\hat{H}_k f(x_k^i) &= g_k(x_k^i)\mathbb{E}[\langle \mathbb{E}[f(\hat{X}_{k+2})|\hat{X}_{k+1}]\mathrm{D}g_{k+1}(\hat{X}_{k+1}) \\ &\quad -g_{k+1}(\hat{X}_{k+1})\mathbb{E}[f(\hat{X}_{k+2})\Psi_{k+1}(X_{k+1},\varepsilon_{k+2})|\hat{X}_{k+1}],\Delta_{k+1}\rangle|\hat{X}_k = x_k^i].\end{aligned}$$

For a time step $0 \leq k \leq n-2$, we have the following one step transition system:

$$\left\{\begin{array}{lcl}\hat{\hat{R}}_k &=& \hat{\hat{H}}_k\hat{\hat{R}}_{k+1}, \\ \widehat{R}_k &=& \hat{\hat{H}}_k\widehat{R}_{k+1} + \hat{H}_k\hat{\hat{R}}_{k+2}.\end{array}\right. \tag{1.4.8}$$

Introducing $\hat{U}_k$ in addition to $\hat{\hat{U}}_k$ we can define the following forward scheme:

Scheme A: Forward expression

$$\left\{\begin{array}{lcl}\hat{\hat{U}}_0 &=& \hat{\mu}_0 \circ \hat{\hat{H}}_0, \\ \hat{U}_2 &=& \hat{\mu}_0 \circ \hat{H}_0, \\ \text{for any} && 0 \leq k \leq n-3, \\ \hat{\hat{U}}_{k+1} &=& \hat{\hat{U}}_k \circ \hat{\hat{H}}_{k+1}, \\ \hat{U}_{k+3} &=& \hat{U}_{k+2} \circ \hat{\hat{H}}_{k+2} + \hat{\hat{U}}_k \circ \hat{H}_{k+1}.\end{array}\right. \tag{1.4.9}$$

Finally, given the final conditions:

$$\left\{\begin{array}{lcl}\widehat{R}_n f(\hat{X}_n) &=& H_n^n f(\hat{X}_n), \\ \hat{\hat{R}}_n(\hat{X}_n) &=& H_n^n f(\hat{X}_n), \\ \widehat{R}_{n-1}(\hat{X}_{n-1}) &=& g_{n-1}(\hat{X}_{n-1})\mathbb{E}[H_n^n f(\hat{X}_n) + \langle \mathrm{D}H_n^n f(\hat{X}_n), \Delta_n\rangle|\hat{X}_{n-1}],\end{array}\right.$$

we have for any $n > 1$,

$$\left\{\begin{array}{lcll}\hat{\mu}_0\hat{\hat{R}}_0 &=& \hat{\hat{U}}_{n-1} \circ H_n^n &= \hat{\hat{\pi}}_n, \\ \hat{\mu}_0\widehat{R}_0 &=& \hat{\hat{U}}_{n-2}\widehat{R}_{n-1} + \hat{U}_n H_n^n &= \hat{\pi}_n.\end{array}\right.$$

1.4.3 Error bounds

The main result of this paragraph is the following theorem, providing a convergence rate of the unnormalized filter approximation error for the two step recursive scheme **A** .

Theorem 1.4.1 *Let $(\hat{X}_k)$ be a marginal stationary (N_k)-quantization of (X_k), f satisfying $H_n^n f \in \mathcal{C}^1_{b,Lip}$. Assume* **H**1, **H**2, **H**3, **H**4, *$q = d$ and furthermore :*

H 5 *There exists a constant $\psi_p > 0$ and $\bar{s} > 1$ such that:*

$$\max_{0\leq k\leq n-1} \|\Psi_k(X_k, \varepsilon_{k+1})\|_{\bar{s}p} \leq \psi_p < +\infty.$$

Hence, there exists a non negative real sequence of constants $(M_j^n)_{0\leq j\leq n}$ such that:

$$|\pi_n f - \hat{\pi}_n f| \leq \sum_{j=0}^{n} M_j^n \|\Delta_j\|^2_{max\{stp,\bar{t}p,2p\}}$$

where $s = \frac{\bar{s}}{\bar{s}-1}$, $t > 0$, $\frac{1}{t} + \frac{1}{\bar{t}} = 1$ and $M_j^n \leq \alpha(p,f)(n+1)L^n(\frac{(L)^{j+1}-1}{3K-1})(\frac{(3K)^{n-j+1}-1}{3K-1})$.

Example 1.4.2 Assume that $F_k : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ reads:

$$F_k(x, \varepsilon) = b_k(x) + \sigma_k(x)\varepsilon, \tag{1.4.10}$$

where σ_k and b_k are differentiable with bounded derivatives and $\forall x \in \mathbb{R}^d$, $\sigma_k(x) > c_k$. Then:

$$\Psi_k(x, \varepsilon) = \frac{\sigma_k'(x) + \frac{\mathbf{p}'}{\mathbf{p}}(\varepsilon)\big(\varepsilon\sigma_k'(x) + b_k'(x)\big)}{\sigma_k(x)}.$$

- When $\varepsilon_k \sim \mathcal{N}(0,1)$, it is the natural framework to study the Euler scheme of a Brownian diffusion. In this case, the previous hypothesis **H**5 is satisfied.

- When ε_k distribution is centered Laplace of parameter $\lambda > 0$, or $\varepsilon_k + m \sim Gamma(m, 1)$ with $m > 1$, hypothesis **H**5 is also satisfied.

- In a more general case, when $\varepsilon_k \in L^{p+\eta}$ for some $\eta > 0$ the following assumption:
  H 5' There exists a constant $\psi_p > 0$ and $\bar{s}' > 1$ such that $\|\frac{\mathbf{p}'}{\mathbf{p}}(\varepsilon_1)\|_{\bar{s}'p} \leq \psi_p < +\infty$.
  could replace **H**5 and gives more explicit conditions on the signal innovation distribution.

Compared to Example 1.3.1 given for the one step iterative scheme, we see that hypothesis **H**5 (or **H**5') allows to relax the boundedness constraint on $\partial_\varepsilon F_k$ in **H**2' to involve some other constraints on the signal innovations distribution.

The structure of the proof of Theorem 1.4.1 is the same as that of the previous section. We first study the error induced by the differential term estimation. Let us reconsider for $0 \leq k \leq n-1$ and the test function f: $\quad \hat{A}_k := \mathrm{D}R_k f(\hat{X}_k) - \widehat{DR}_k f(\hat{X}_k)$.
The error bounds for $\|\hat{A}_k\|_p$ with the new definition of the differential term approximation $\widehat{DR}_k f$ are given by the following lemma:

Lemma 1.4.1 *With assumption* **H**5 *on the weight function* Ψ_k *and* f *such that* $H_n^n f \in \mathcal{C}^1_{b,Lip}$, *there exists a non negative real sequence* $(D_j^{k,n})_{0\leq k\leq j\leq n}$ *such that:*

$$\|\hat{A}_k\|_p \leq \sum_{j=k}^{n} D_j^{k,n} \|\Delta_j\|_{sp}$$

where $s = \frac{\bar{s}}{\bar{s}-1}$ *and* $D_j^{k,n} \leq \alpha(p,f) L^{n-k} \frac{(3K)^{n-j+1}-1}{3K-1}$.

Proof.
We redefine the operators Q_k and $\hat{Q}_k$ for $f : \mathbb{R}^d \to \mathbb{R}^d$ as follows:

$$\begin{aligned} Q_k f(X_k) &= -\mathbb{E}[f(X_{k+1})\Psi_k(X_k, \varepsilon_{k+1})|X_k], \\ \hat{Q}_k f(\hat{X}_k) &= -\mathbb{E}[f(\hat{X}_{k+1})\Psi_k(X_k, \varepsilon_{k+1})|\hat{X}_k]. \end{aligned}$$

Then $Q_k f = \mathrm{D}\mathbf{P}_k f$, so that:

$$\begin{aligned} \mathrm{D}R_k f(\hat{X}_k) &= \mathrm{D}g_k(\hat{X}_k)\mathbf{P}_k R_{k+1} f(\hat{X}_k) + g_k(\hat{X}_k) Q_k R_{k+1} f(\hat{X}_k), \\ \widehat{DR}_k f(\hat{X}_k) &= \mathrm{D}g_k(\hat{X}_k)\widehat{\mathbf{P}}_k \hat{\hat{R}}_{k+1} f(\hat{X}_k) + g_k(\hat{X}_k) \hat{Q}_k \hat{\hat{R}}_{k+1} f(\hat{X}_k). \end{aligned}$$

Consequently, $\hat{A}_k$ can be written as:

$$\hat{A}_k = \mathrm{D}g_k(\hat{X}_k)\left[\mathbf{P}_k R_{k+1} f(\hat{X}_k) - \widehat{\mathbf{P}}_k \hat{\hat{R}}_{k+1} f(\hat{X}_k)\right] + g_k(\hat{X}_k)\left[Q_k R_{k+1} f(\hat{X}_k) - \hat{Q}_k \hat{\hat{R}}_{k+1} f(\hat{X}_k)\right],$$

so that using **H**1 and that $\widehat{\mathbf{P}}_k \hat{\hat{R}}_{k+1} f(\hat{X}_k) = \mathbb{E}[\hat{\hat{R}}_{k+1} f(\hat{X}_{k+1})|\hat{X}_k]$, we have:

$$\|\hat{A}_k\|_p \leq L\|\mathbf{P}_k R_{k+1} f(\hat{X}_k) - \mathbb{E}[\hat{\hat{R}}_{k+1} f(\hat{X}_{k+1})|\hat{X}_k]\|_p + L\|Q_k R_{k+1} f(\hat{X}_k) - \hat{Q}_k \hat{\hat{R}}_{k+1} f(\hat{X}_k)\|_p. \tag{1.4.11}$$

Since conditional expectation is an L^p-contraction,, the first term on the right hand side of inequality (1.4.11) writes:

$$\begin{aligned} \|\mathbf{P}_k R_{k+1} f(\hat{X}_k) &- \mathbb{E}[\hat{\hat{R}}_{k+1} f(\hat{X}_{k+1})|\hat{X}_k]\|_p \\ &\leq \|\mathbf{P}_k R_{k+1} f(\hat{X}_k) - \mathbb{E}[\mathbf{P}_k R_{k+1} f(X_k)|\hat{X}_k]\|_p \\ &\quad + \|\mathbb{E}[\mathbf{P}_k R_{k+1} f(X_k) - \hat{\hat{R}}_{k+1} f(\hat{X}_{k+1})|\hat{X}_k]\|_p \\ &\leq [\mathbf{P}_k R_{k+1} f]_{Lip}\|\Delta_k\|_p + \|R_{k+1} f(X_{k+1}) - \hat{\hat{R}}_{k+1} f(\hat{X}_{k+1})\|_p. \end{aligned} \tag{1.4.12}$$

It follows from Theorem 1.2.1 that:
$\|\mathbf{P}_k R_{k+1} f(\hat{X}_k) - \widehat{\mathbf{P}}_k \hat{\hat{R}}_{k+1} f(\hat{X}_{k+1})\|_p \leq \sum_{j=k+1}^{n} C_j^{k+1,n} \|\Delta_j\|_p + K \|\mathrm{D} R_{k+1} f\|_\infty \|\Delta_k\|_p$.
Moreover, the second term on the right hand side of inequality (1.4.11) gives:

$$\begin{aligned}
& \|Q_k R_{k+1} f(\hat{X}_k) - \hat{Q}_k \hat{\hat{R}}_{k+1} f(\hat{X}_k)\|_p \\
\leq\ & \|Q_k R_{k+1} f(\hat{X}_k) - \mathbb{E}[Q_k R_{k+1} f(X_k)|\hat{X}_k]\|_p + \|\mathbb{E}[Q_k R_{k+1} f(X_k)|\hat{X}_k] - \hat{Q}_k \hat{\hat{R}}_{k+1} f(\hat{X}_k)\|_p \\
\leq\ & \|Q_k R_{k+1} f(\hat{X}_k) - Q_k R_{k+1} f(X_k)\|_p \\
& + \|\mathbb{E}[\mathbb{E}[R_{k+1} f(X_{k+1}) \Psi_k(X_k, \varepsilon_{k+1})|X_k]|\hat{X}_k] - \mathbb{E}[\hat{\hat{R}}_{k+1} f(\hat{X}_{k+1}) \Psi_k(X_k, \varepsilon_{k+1})|\hat{X}_k]\|_p \\
\leq\ & \|Q_k R_{k+1} f(\hat{X}_k) - Q_k R_{k+1} f(X_k)\|_p + \|\Psi_k(X_k, \varepsilon_{k+1}) \left(R_{k+1} f(X_{k+1}) - \hat{\hat{R}}_{k+1} f(\hat{X}_{k+1}) \right) \|_p.
\end{aligned} \tag{1.4.13}$$

But, $Q_k R_{k+1} f(X_k) = \mathrm{D}\mathbf{P}_k R_{k+1} f(X_k)$, so hypothesis **H**2 on $\mathbf{P}_k$ implies that:

$$[Q_k R_{k+1} f]_{Lip} = [\mathrm{D}\mathbf{P}_k R_{k+1} f]_{Lip} \leq K([\mathrm{D} R_{k+1} f]_{Lip} \vee \|D R_{k+1} f\|_\infty) = K u_{k+1}.$$

Hence,

$$\|Q_k R_{k+1} f(\hat{X}_k) - Q_k R_{k+1} f(X_k)\|_p \leq K u_{k+1} \|\Delta_k\|_p. \tag{1.4.14}$$

Using Holder inequality, with $s = \frac{\bar{s}}{\bar{s}-1} \geq 1$ we get:

$$\begin{aligned}
& \|\Psi_k(X_k, \varepsilon_{k+1}) \left(R_{k+1} f(X_{k+1}) - \hat{\hat{R}}_{k+1} f(\hat{X}_{k+1}) \right) \|_p \\
& \leq \|\Psi_k(X_k, \varepsilon_{k+1})\|_{\bar{s}p} \|R_{k+1} f(X_{k+1}) - \hat{\hat{R}}_{k+1} f(\hat{X}_{k+1})\|_{sp}
\end{aligned} \tag{1.4.15}$$

It follows from Theorem 1.2.1 and hypothesis **H**5 by combining terms (1.4.13), (1.4.14) and (1.4.15) that:

$$\begin{aligned}
\|Q_k R_{k+1} f(\hat{X}_k) \ -\ & \hat{Q}_k \hat{\hat{R}}_{k+1} f(\hat{X}_k)\|_p \\
& \leq \sum_{j=k+1}^{n} \psi_p C_j^{k+1,n} \|\Delta_j\|_{sp} + K u_{k+1} \|\Delta_k\|_p, \qquad (1.4.16) \\
\text{and} \qquad \|\hat{A}_k\|_p \ & \leq L(\psi_p + 1) \sum_{j=k+1}^{n} C_j^{k+1,n} \|\Delta_j\|_{sp} + LK \left(u_{k+1} + \|\mathrm{D} R_{k+1} f\|_\infty \right) \|\Delta_k\|_{sp} \\
& \leq \sum_{j=k}^{n} D_j^{k,n} \|\Delta_j\|_{sp}. \qquad (1.4.17)
\end{aligned}$$

Then, explicit upper bounds for $D_j^{k,n}$ can easily be established. (see Appendix 1.6.3) □

Proof of Theorem 1.4.1. Reconsider $V_k f = \mathbb{E}[R_k f(X_k)|\hat{X}_k] - \hat{R}_k f(\hat{X}_k)$ for $0 \leq k \leq n$. The proof can be carried out as in the previous case of Theorem 1.3.1. The unique difference lies in the term $\hat{A}_k$. Using Lemma 1.4.1 combined with Holder inequality for some $t > 1$ and its conjugate $\bar{t} = \frac{t}{t-1}$ we have:

$$\|\langle \hat{A}_{k+1}, \Delta_{k+1}\rangle\|_p \leq \|\hat{A}_{k+1}\|_{tp}\|\Delta_{k+1}\|_{\bar{t}p} \leq \tfrac{1}{2}\sum_{j=k+1}^{n} D_j^{k+1,n}(\|\Delta_j\|_{stp}^2 + \|\Delta_{k+1}\|_{\bar{t}p}^2).$$

Then inequality (1.3.16) writes:

$$\begin{aligned}
\|V_k f\|_p &\leq ([\mathrm{D}R_k f]_{Lip} + LKu_{k+1})\,\|\Delta_k\|_{2p}^2 + 2L[\mathrm{D}R_{k+1}f]_{Lip}\|\Delta_{k+1}\|_{2p}^2 \\
&\quad + L\frac{1}{2}\sum_{j=k+1}^{n} D_j^{k+1n}(\|\Delta_j\|_{stp}^2 + \|\Delta_{k+1}\|_{\bar{t}p}^2) + L\|V_{k+1}f\|_p \\
&\leq ([\mathrm{D}R_k f]_{Lip} + LKu_{k+1})\,\|\Delta_k\|_{max\{stp,\bar{t}p,2p\}}^2 \\
&\quad + L\left([\mathrm{D}R_{k+1}f]_{Lip} + D_{k+1}^{k+1,n} + \frac{1}{2}\sum_{j=k+2}^{n} D_j^{k+1,n}\right)\|\Delta_{k+1}\|_{max\{stp,\bar{t}p,2p\}}^2 \\
&\quad + \frac{1}{2}L\sum_{j=k+2}^{n} D_j^{k+1,n}\|\Delta_j\|_{max\{stp,\bar{t}p,2p\}}^2 + L\|V_{k+1}f\|_p \qquad (1.4.18)
\end{aligned}$$

By induction, we derive: $\|V_k f\|_p \leq \sum_{j=k}^{n} M_j^{k,n}\|\Delta_j\|_{max\{stp,\bar{t}p,2p\}}^2$.
Taking $k=0$ and writing $|\pi_n f - \hat{\pi}_n f| \leq \|V_0 f\|_p$ we establish the announced result. See Appendix 1.6.3 for a detailed proof of the explicit expressions of $(M_j^{k,n})$. □

1.4.4 The case of regularizing kernels

In this paragraph we deal with an interesting skill of the two step recursive first order scheme, which allows to establish first order schemes for non differentiable test functions f, more precisely with no differentiability assumption on $H_n^n f$.

Proposition 1.4.1 ***H 2"*** Assume $\mathbf{P}_k$ is K-Lipschitz such that for all f bounded Lipschitz continuous, $\mathbf{P}_k f \in \mathcal{C}_{b,Lip}^1$.
If f is a test function such that $H_n^n f$ is bounded Lipschitz continuous, then $R_k f \in \mathcal{C}_{b,Lip}^1$ for all $0 \leq k \leq n-1$.

This proposition is easily proved, using equation (1.2.17) and an induction on k. Furthermore, it allows to define an alternative scheme to scheme **A** , taking into account the non differentiability of $H_n^n f$:

Scheme A'

$$\left\{\begin{array}{lcl}
\widehat{R}_n f(\hat{X}_n) & = & H_n^n f(\hat{X}_n) = \hat{\widehat{R}}_n f(\hat{X}_n), \\
\widehat{R}_{n-1} f(\hat{X}_{n-1}) & = & g_{n-1}(\hat{X}_{n-1})\mathbb{E}[H_n^n f(\hat{X}_n)|\hat{X}_{n-1}] = \hat{\widehat{R}}_{n-1} f(\hat{X}_{n-1}), \\
\widehat{R}_k f(\hat{X}_k) & = & g_k(\hat{X}_k)\widehat{\mathbf{P}}_k\widehat{R}_{k+1}f(\hat{X}_k) + g_k(\hat{X}_k)\times \\
& & \Big(\mathbb{E}[\langle \mathrm{D}g_{k+1}(\hat{X}_{k+1})\widehat{\mathbf{P}}_{k+1}\hat{\widehat{R}}_{k+2}f(\hat{X}_{k+1}), \Delta_{k+1}\rangle|\hat{X}_k] - \mathbb{E}[\,\langle g_{k+1}(\hat{X}_{k+1})\times \\
& & \mathbb{E}[\hat{\widehat{R}}_{k+2}f(\hat{X}_{k+2})\Psi_{k+1}(X_{k+1},\varepsilon_{k+2})|\hat{X}_{k+1}], \Delta_{k+1}\rangle|\hat{X}_k]\Big), \\
& & 0 \leq k \leq n-2.
\end{array}\right. \qquad (1.4.19)$$

We then define the first order unnormalized filter estimator by $\hat{\pi}_n f = \mathbb{E}[\widehat{R}_0 f(\hat{X}_0)]$ generated from scheme **A**'. The error induced by such an estimator introduces additional zero order type terms as we need one single backward iteration to be able to use first order correctors. This can be seen clearly in the the following theorem.

Theorem 1.4.2 *Let $(\hat{X}_k)$ be a stationary (N_k)-quantization of (X_k), f satisfying $H_n^n f$ is bounded Lipschitz continuous. Assume* **H**1, **H**2", **H**3, **H**4, **H**5 *and $q = d$.*
Then, there exists a non negative real sequence of constants $(\bar{M}_j^n)_{0\leq j\leq n}$ such that:

$$|\pi_n f - \hat{\pi}_n f| \leq \sum_{j=0}^{n} \bar{M}_j^n \|\Delta_j\|^2_{max\{stp,\bar{t}p,2p\}} + C^1\|\Delta_{n-1}\|_p + C^2\|\Delta_n\|_p$$

where $s = \frac{\bar{s}}{\bar{s}-1}$, $t > 0$, $\frac{1}{t} + \frac{1}{\bar{t}} = 1$.

In practice, the regularizing effect can be viewed in the case of the Euler scheme of a diffusion implemented with a Gaussian noise (see Example 1.4.2 and Appendix B). This is the case studied in [6] for pricing American options with first order schemes. It is shown that $\mathbf{P}_k$ satisfies **H**2". Nevertehless, a special attention have to be given to the Lipschitz constants dependency in time discretization step, and consequently in our filtering problem, to the observation horizon. Namely, if f is Lipschitz continuous, then according to Proposition 2 in [6] we have $[\mathrm{D}\mathbf{P}_k f]_{Lip} \leq C[f]_{Lip}\sqrt{n}$. This result alters $\bar{M}_j^n$ dependency in n and consequently the filter estimator convergence for high observation horizons.

Remark 1.4.1 For numerical implementation, we can compensate the error bounds deterioration in Theorem 1.4.2 by bigger quantizers in the two last observation dates $n-1$ and n.

1.5 Convergence result for the normalized filter

Let f be such that $H_n^n f \in \mathcal{C}^1_{b,Lip}$. Owing to Theorem 1.3.1 and Theorem 1.4.1, we have seen that the estimation error on the unnormalized filter π_n, using stationary (N_k)-quantizations $(\hat{X}_k)$, can be written as:

$$|\pi_n f - \hat{\pi}_n f| \leq \sum_{j=0}^{n} M_j^n(f,\alpha)\|\Delta_j\|^2_{2\alpha p},$$

where $\alpha = \max\{\frac{st}{2}, \frac{\bar{t}}{2}, 1\} \geq 1$ or $\alpha = 1$ depending on whether we are using scheme **A** or **B** , and $\Delta_j = X_j - \hat{X}_j$.
Now, we derive results on the normalized first order quantization filter estimator $\hat{\Pi}_n$, defined by Kallianpur-Striebel formula as $\hat{\Pi}_n f = \frac{\hat{\pi}_n f}{\hat{\pi}_n \mathbf{1}}$.

Thus, the estimation error will be:

$$\begin{aligned} |\Pi_n f - \hat{\Pi}_n f| &\leq |\frac{\pi_n f}{\pi_n \mathbf{1}} - \frac{\pi_n f}{\hat{\pi}_n \mathbf{1}}| + |\frac{\pi_n f - \hat{\pi}_n f}{\hat{\pi}_n \mathbf{1}}| \\ &\leq \frac{\|H_n^n f\|_\infty \pi_{n-1}\mathbf{1}}{\pi_n \mathbf{1} \hat{\pi}_n \mathbf{1}} |\pi_n \mathbf{1} - \hat{\pi}_n \mathbf{1}| + \frac{1}{\hat{\pi}_n \mathbf{1}} |\pi_n f - \hat{\pi}_n f| \\ &\leq \sum_{j=0}^{n} \frac{M_j^n(f,\alpha) + c^y M_j^n(\mathbf{1},\alpha)\|H_n^n f\|_\infty}{\hat{\pi}_n \mathbf{1}} \|\Delta_j\|_{2\alpha p}^2 \end{aligned} \tag{1.5.1}$$

Since $\alpha = 1$ in Theorem 1.3.1, the convergence rate improvement obtained for the unnormalized filter is preserved by the normalization.
When $\alpha > 1$, which is the case for Theorem 1.4.1, further results are needed to establish a convergence rate improvement. In fact, from inequality (1.5.1) it comes out that we need to describe the $L^{2\alpha p}$-behavior of sequences of L^{2p}-optimal quantizers. In this direction, a rather satisfactory result can be established using Zador Theorem 1.2.2 and Holder inequality. Namely, if $X \in L^{r'}(\mathbb{R}^d)$ for every $r' > 0$, then $\|X - h_N^*(X)\|_s = O(N^{\frac{-\rho}{d}})$ for any $\rho \in (0, \frac{r}{s})$.

This allows to establish the following theorem, for $\hat{\Pi}_n$ obtained from the two step recursive scheme:

Theorem 1.5.1 *Assume that $\bar{s}$ in **H5** satisfies $\bar{s} > \frac{3}{2}$ and that for $0 \leq k \leq n$ and all $r > 0$ $X_k \in L^r(\mathbb{R}^d)$. Let $(\hat{X}_k)$ be an L^2-optimal (N_k)-quantization of (X_k).*
Then, there exists $\rho \in (\frac{1}{2}, 1]$ such that for all $f : \mathbb{R}^d \to \mathbb{R}$ satisfying $H_n^n f \in \mathcal{C}_{b,Lip}^1$ we have:

$$|\Pi_n f - \hat{\Pi}_n f| \leq \sum_{j=0}^{n} c_j(\rho, p, d) \frac{M_j^n(f,\alpha) + c^y M_j^n(\mathbf{1},\alpha)\|H_n^n f\|_\infty}{\hat{\pi}_n \mathbf{1}} N_j^{\frac{-2\rho}{d}}.$$

Proof. If $\bar{s} > \frac{3}{2}$, then $1 < s < 3$ and there exists $\frac{4}{3} < t < \frac{4}{s}$. For such a number $t > 1$ we will have $\bar{t} < 4$ and $st < 4$ so that inequality (1.5.1) is satisfied for $\alpha = \max\{\frac{\bar{t}}{2}, \frac{st}{2}, 1\} \in [1, 2[$.
Hence, for some $\rho \in (\frac{1}{2}, \frac{1}{\alpha}) \subset (0, \frac{1}{\alpha})$, we can write: $\|\Delta_k\|_{2\alpha p} = O(N_k^{\frac{-2\rho}{d}})$.
Consequently from (1.5.1), $|\Pi_n f - \hat{\Pi}_n f| \leq \sum_{j=0}^{n} c_j(\rho, p, d) \frac{M_j^n(f,\alpha) + c^y M_j^n(\mathbf{1},\alpha)\|H_n^n f\|_\infty}{\hat{\pi}_n \mathbf{1}} N_j^{\frac{-2\rho}{d}}$.
□

Remark 1.5.1 A conjecture has been made recently by H. Luschgy and G. Pagès to describe the $L^{s'}$-behavior of sequences of L^r-optimal quantizers of an $\mathbb{R}^d$-valued random vector for some $0 < r < s' < r + d$:
If $X \in L^r(\mathbb{R}^d)$ such that $\mathbb{P}_X(d\xi) = \varphi(\xi)\lambda_d(d\xi)$ and $\int \varphi^{1 - \frac{s'}{r+d}} d\lambda_d < +\infty$, then any sequence (h_N^*) of L^r-optimal N-quantizers most likely satisfies $\|X - h_N^*(X)\|_{s'} = O(N^{-\frac{1}{d}})$.
This allows to establish an equivalent of Theorem 1.5.1 where $\rho = 1$ and $\bar{s}$ is assumed to satisfy $\bar{s} > 1 + \frac{1}{d}$ and then to rise convergence rate order of two step recursive schemes to the one step recursive one.

1.6 Numerical illustrations

Previous filter approximation methods will be applied to estimate $\Pi_n f_1$ and $\Pi_n f_2$, where $f_1(x) = x$ and $f_2(x) = \exp(-|x|)$. Elements of comparison with alternative filter estimation methods will be given, namely particle filtering methods:

SIS Sequential Importance Sampling [1, 15] which is based on a weighted Monte Carlo approach. This method can be considered as close to the quantization method in the sense that it uses weight transformations in the updating step. Unfortunately, it is known to suffer from weights degenerescense.

SIR Sequential Importance Re-sampling [24, 15] which adds a re-sampling step to the previous algorithm in order to avoid weights degenerescense.

We will test estimations for different fixed observation sets and so we denote by $\hat{\Pi}_{y,n}$, the estimation filter associated to the observation process $y = (y_0, \ldots, y_n)$.
In all the following examples, we choose to study stationary signal processes in order to simplify the off line procedure of computing the quantizers. In fact, as we marginally quantize the signal process, we can just expand the grids of the centered reduced corresponding distribution. The obtained quantizers are no longer optimal, some further manipulations are necessary to save the quantizer stationarity property especially in the multidimensional cases.

1.6.1 Kalman filter

Both signal and observation equations are linear with Gaussian independent noises. It is known, that the filter in this case has a Gaussian distribution which parameters (mean and variance) can be computed sequentially via a deterministic algorithm (KF), (see [21]).

$$\text{We set:} \quad \begin{cases} X_k = \rho X_{k-1} + \theta \varepsilon_{k+1}, \\ Y_k = X_k + \alpha \eta_k, \\ \varepsilon_k \ \ and \ \ \eta_k \quad iid \quad \sim \mathcal{N}(0, I_d), \\ \rho, \ \ \theta, \ \ \alpha \quad \in \mathcal{M}_d(\mathbb{R}). \end{cases} \tag{1.6.1}$$

1.6.1.1 One dimensional case: d=1

We choose $-1 < \rho < 1$ and $X_0 \sim \mathcal{N}(0, \frac{\theta^2}{1-\rho^2})$, so that for any $0 \leq k \leq n$, we have $X_k \sim \mathcal{N}(0, \frac{\theta^2}{1-\rho^2})$. In this particular case, we could first compute[2] Γ an L^2-optimal quantizer of the centered reduced Gaussian distribution and the companion parameters for a single transition step. The quantizers Γ_k are then deduced by an expansion $\Gamma_k = \frac{\theta^2}{1-\rho^2} \times \Gamma$.

[2]Optimal quantizers for the Gaussian distribution are downloadable on `http://www.proba.jussieu.fr/pageperso/pages/`

The two first order schemes are compared to the zero order one with $N_k = 200$, $0 \leq k \leq n$. Exact values are computed via the Kalman-Bucy recursive filter algorithm. Particles methods are also tested for the sake of comparison.

(ρ, θ, α)	(0.65,1.0,0.1)		(0.8,1.0,0.1)	
	$\hat{\Pi}_{y,25} f_1$	$\hat{\Pi}_{y,25} f_2$	$\hat{\Pi}_{y,25} f_1$	$\hat{\Pi}_{y,25} f_2$
KF(Ref. Value)	-3.239	0.039	1.754	0.17394
SIS (5000 pts)	-3.244	0.039	1.7487	0.17489
SIR (5000 pts)	-3.2398	0.039	1.7542	0.1739
QF Or0 (200 pts)	-3.2394	0.0393	1.7522	0.17425
QF Or1 1-step (200 pts)	-3.2381	0.039431	1.7524	0.17422
QF Or1 2-step (200 pts)	-3.2381	0.039431	1.7524	0.17422

Table 1.1: One dimensional Kalman filter case.

1.6.1.2 Multidimensional case: d=2

Although the quantization based filter schemes presented previously depend on the signal dimension d, for both complexity and convergence rate, it remains interesting to compute estimations for medium signal dimensions. We reconsider equation (1.6.1) with parameters:

$$\rho = \begin{pmatrix} 0.996 & 0 \\ 0 & 0.996 \end{pmatrix}, \theta = \begin{pmatrix} 0.05 & -0.01 \\ -0.01 & 0.02 \end{pmatrix} \quad \text{and} \quad \alpha = 0.5 I_d.$$

The initial signal distribution is centered, Gaussian with covariance matrix:

$$\Sigma_0 = \begin{pmatrix} 0.325 & -0.087 \\ -0.087 & 0.0626 \end{pmatrix}.$$

The chosen prior distribution is the stationary one. For signal quantization, we take $\Gamma = \{z^1, \ldots, z^N\}$ the L^2-optimal N-quantizer of a centered reduced Gaussian distribution. At $0 \leq k \leq N$, $X_k \sim \mathcal{N}(0, \Sigma_0)$ and we define the marginal stationary (N_k)-quantizer of (X_k) as follows:

$$\hat{X}_k = \sum_{i=1}^{N} \Sigma_0^{\frac{1}{2}} z^i \mathbf{1}_{\{X_k \in \Sigma_0^{\frac{1}{2}} \mathbf{C}_i(\Gamma)\}}$$

Although quantizers are not optimal, we obtain satisfactory convergence results. Convergence errors are represented in Figure 1.4. From the log-log scale representation in Figure 1.4, we can evaluate the convergence rate improvement using a regression. Table 1.6.1.2 summarizes the computed slopes of the regressions.

Or0	Or1 1-step	Or1 2-step
-0.45	-1.1	-1.04

Table 1.2: Regression slopes on the log-log scale representation (d=2)

We observe nearly the expected theoretical results. The convergence rate for the zero order scheme is close of $\frac{1}{d} = 0.5$. For first order schemes, the slope is slightly better than the theoretical one $\frac{2}{d} = 1$.

1.6.2 Canonical stochastic volatility model (SVM)

We introduce now a non linearity in the observation equation. We consider the following state equations in $\mathbb{R}$:

$$\begin{cases} X_k = \beta X_{k-1} + \sigma\varepsilon_{k+1}, \\ Y_k = exp(\frac{X_k}{2})\eta_k, \\ \varepsilon_k \ \ and \ \ \eta_k \ \ \ iid \ \ \sim \mathcal{N}(0,1), \\ -1 < \beta < 1 \ \ and \ \ \ \sigma \in \mathbb{R}_+^*. \end{cases} \tag{1.6.2}$$

Remark 1.6.1 This is the time discretization of a continuous diffusion model introduced in finance as a model of an asset dynamics with stochastic volatility. The stock price S_t and its volatility σ_t solve the following stochastic differential system:

$$\begin{cases} dS_t = \mu_t S_t dt + \sigma_t S_t dW_t, \\ d(ln(\sigma_t^2)) = -\lambda ln(\sigma_t^2)dt + \tau dW_t. \end{cases} \tag{1.6.3}$$

The stock price is supposed to be observable so that the filtering problem corresponds to a volatility estimation problem, given the set of observed past prices. Taking a time discretization step Δ, the Euler scheme writes:

$$\begin{cases} ln(\frac{S_{k+1}}{S_k}) = (\mu_k - \frac{1}{2}\sigma_k^2)\Delta + \sigma_k\sqrt{\Delta}\eta_k, \\ ln(\sigma_{k+1}^2) = (1 - \lambda\Delta)ln(\sigma_k^2) + \tau\sqrt{\Delta}\varepsilon_{k+1}. \end{cases} \tag{1.6.4}$$

Now, taking $Y_k = ln(\frac{S_{k+1}}{S_k})$, $X_k = ln(\sigma_k^2)$, η_k and ε_k iid $\mathcal{N}(0,1)$ conducts to the state equations adopted for the illustration.

Here also we choose $X_0 \sim \mathcal{N}(0, \frac{\sigma^2}{1-\beta^2})$, in order to use the same grid at each time step k. The choice of the triplet (λ, τ, Δ) will determine the discrete time model parameters (β, σ). The exact filter value is not computable for such model, so Figure 1.1 shows the convergence behavior of the quantization filters. The first order schemes clearly converge faster.

Comparison with particle methods is made possible by computing some confidence interval through the 5% and 95% centiles over 4000 realizations of the particle filter estimator. In Figure 1.2 are depicted this interval bounds and one realization of the random estimator as functions of the particle number. For a comparison between the two methods (particles and quantization), we represent in Figure 1.3 quantization based filters in the confidence interval of 10000 particles.

1.6.3 Numerical stability

Two stability aspects have been studied through numerical applications. The implemented state equations are those of the previous section (see equation (1.6.2)) when we model stochastic volatility.
The first point we will be interested in, is degeneration of intuitive scheme devised in Remark 1.3.1. An illustration of such a problem is represented by Figure 1.5.
The second point is the stability of our estimations in time. This is a recurrent problem in filtering methods. Even if we considered a fixed observation horizon all over this work, it is important to study the estimation behaviour when n grows. As the constants M_j^n are exponentially depending of the observation horizon, we have been interested in verifying that this does not alter the numerical performances of our filter estimators. (see Figure 1.6). Note that the chosen state equations and the stationarity assumption give that $K = \beta < 1$.

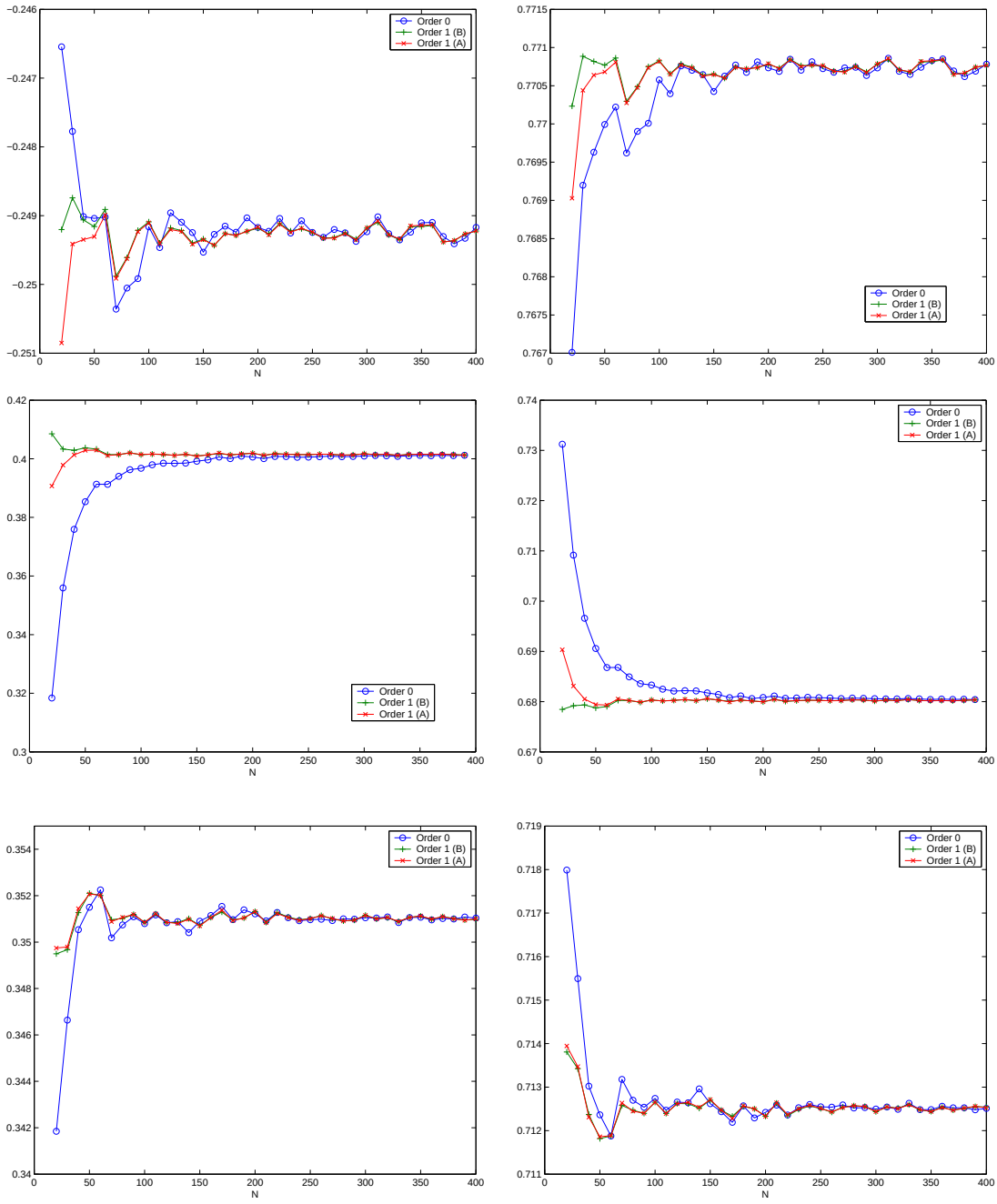

Figure 1.1: Quantization filter approximations for SVM as a function of the quantizer size N_k - three different observation 50-tuples (right: $\Pi_{y,50} f_1$, left: $\Pi_{y,50} f_2$) - $(\beta, \sigma) = (0.996, 0.0316)$.

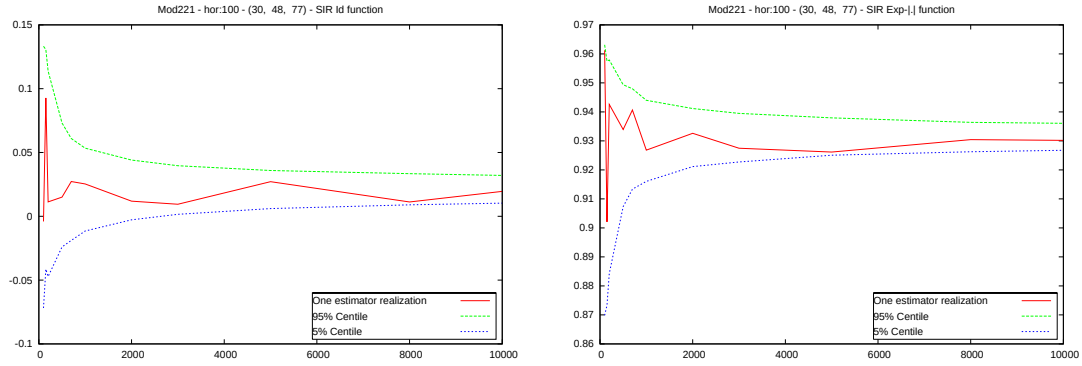

Figure 1.2: Particle filter approximations for SVM as a functions of particle number using SIR algorithm (left: $\Pi_{100} f_1$, right: $\Pi_{100} f_2$) - $(\beta, \sigma) = (0.995, 0.01)$ - Centiles over 4000 realizations.

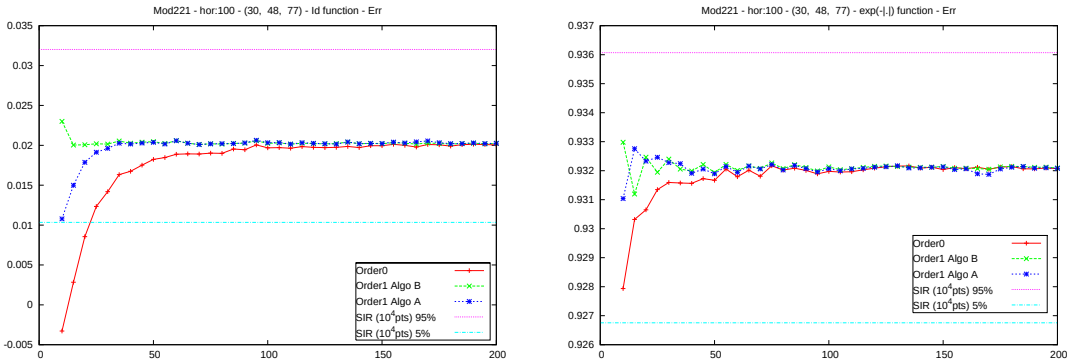

Figure 1.3: Quantization filter estimator as functions of quantizer size, in the SIR confidence interval with 10^4 particles (right: $\hat{\Pi}_{100} f_1$, left: $\hat{\Pi}_{100} f_2$) - $(\beta, \sigma) = (0.995, 0.01)$.

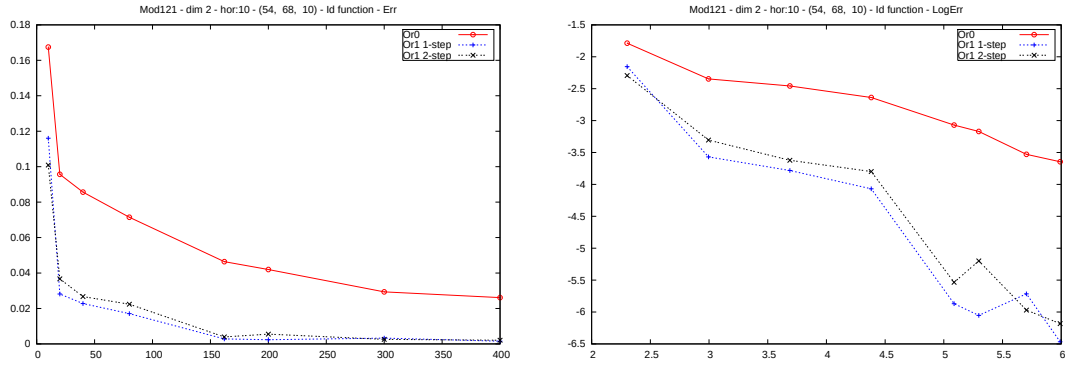

Figure 1.4: Quantization filter estimator errors for 2-dimensional Kalman case as a function of the quantizer size N_k (left: $\|\Pi_{10} f_1 - \hat{\Pi}_{10} f_1\|_2$, right: log-log scale representation).

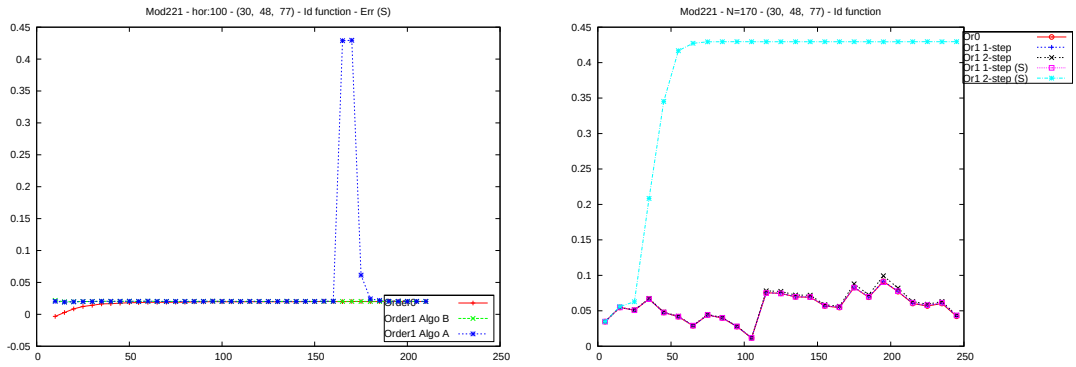

Figure 1.5: Quantization filter estimator for SVM using *intuitive* first order schemes as function of quantizer size (right: $\hat{\Pi}_{100} f_1$ as a function of quantizer size N for $n = 100$, left: $\hat{\Pi}_{\cdot} f_1$ as a function of n for $N = 170$) - $(\beta, \sigma) = (0.995, 0.01)$.

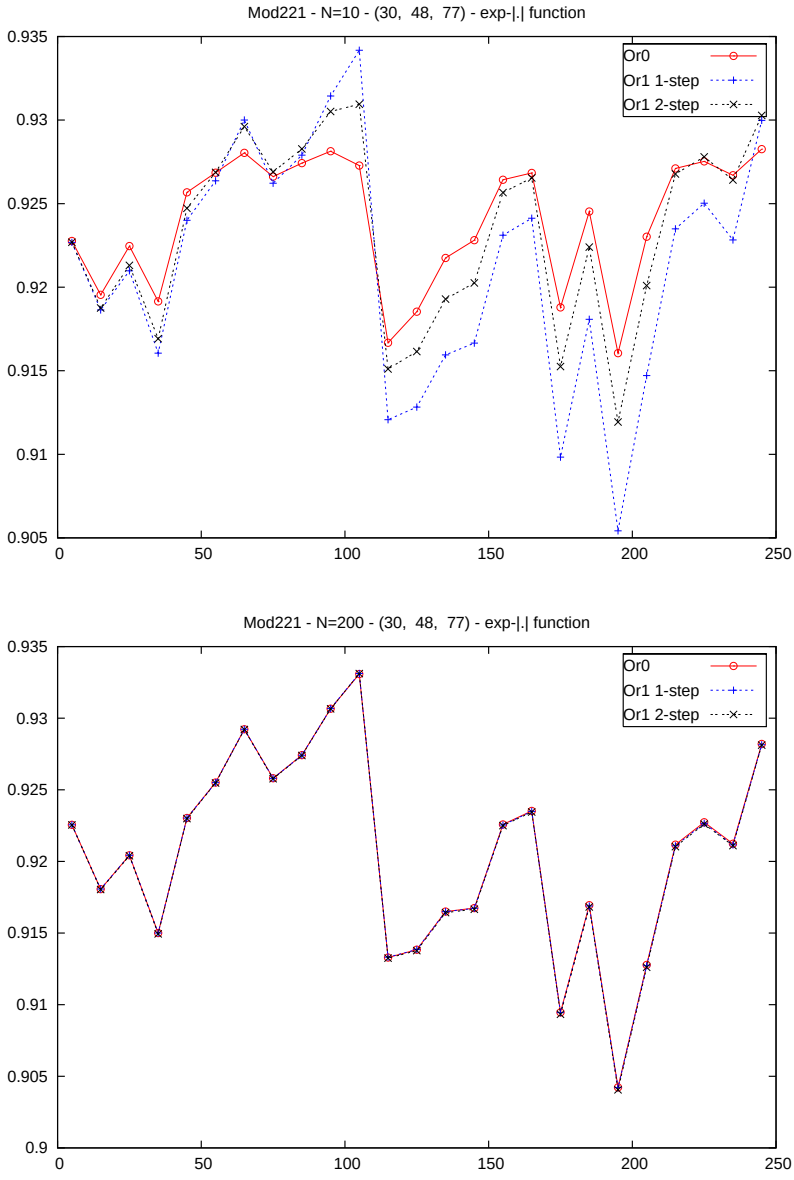

Figure 1.6: Horizon varying effect on quantization based filters for SVM (top: $\hat{\Pi}_{\cdot} f_2$ for $N_k = 10$ as a function of n, bottom: $\hat{\Pi}_{\cdot} f_2$ for $N_k = 200$ as a function of n) - $(\beta, \sigma) = (0.995, 0.01)$.

Appendix A : Explicit error control constants

In all the following, we will denote $G_f^n \stackrel{Def}{=} \max\{\|H_n^n f\|_\infty, \|\mathrm{D}H_n^n f\|_\infty, [\mathrm{D}H_n^n f]_{Lip}\}$. So that, the results of Proposition 1.2.2 write:

$$\begin{array}{rcl} \|R_k f\|_\infty & \leq & G_f^n L^{n-k}, \\ \|\mathrm{D}R_k f\|_\infty & \leq & G_f^n L^{n-k} \frac{K^{n-k+1}-1}{K-1}, \\ u_k & \leq & G_f^n L^{n-k} \frac{(3K)^{n-k+1}-1}{3K-1}. \end{array} \tag{1.6.5}$$

A.1 Zero order scheme constants

Lemma 1.6.1 *Let $\alpha > 0$ and let $(v_k)_{0\leq k\leq n}$, $(u_k)_{0\leq k\leq n}$, $(a_k)_{0\leq k\leq n}$ be some sequences satisfying the following recursive inequalities:*

$$\forall k \in \{0,\ldots,n-1\}, \quad |v_k| \leq |a_k| + \alpha|v_{k+1}| \quad \textit{and} \quad |u_k| \leq \sum_{j=k}^{n} |a_k| + \alpha|u_{k+1}|.$$

Then $\forall k \in \{0,\ldots,n-1\}$:

$$|v_k| \leq \sum_{j=k}^{n-1} \alpha^{j-k}|a_j| + \alpha^{n-k}|v_n|$$

and

$$|u_k| \leq \sum_{j=k}^{n-1} \frac{\alpha^{j-k+1}-1}{\alpha-1}|a_j| + \frac{\alpha^{n-k}-1}{\alpha-1}|a_n| + \alpha^{n-k}|u_n|.$$

A simpler upper bound is also given by:

$$|u_k| \leq \sum_{j=k}^{n} \frac{\alpha^{j-k+1}-1}{\alpha-1}|a_j| + \alpha^{n-k}|u_n|.$$

Compared with the original zero order scheme of [41], the operators defined by system (2.2.1) allow us to enlarge the results of Theorem 1.2.1 and its Corollary, to a slightly larger class of test functions f.

Proposition 1.6.1 *Suppose the test function f for the filter satisfies: $H_n^n f \in \mathcal{C}^1_{b,Lip}$. Then for $k \leq j \leq n$:*

$$C_j^{k,n} \leq G_f^n (2-\delta_{2,p}) L^{n-k} \frac{K^{n-j+1}-1}{K-1}.$$

Proof.

$$\begin{array}{rcl} \|R_k f(X_k) - \hat{\hat{R}}_k f(\hat{X}_k)\|_p & = & \|g_k(X_k)\mathbf{P}_k R_{k+1} f(X_k) - g_k(\hat{X}_k)\widehat{\mathbf{P}}_k \hat{\hat{R}}_{k+1} f(\hat{X}_k)\|_p, \\ & \leq & \|g_k(X_k)\mathbf{P}_k R_{k+1} f(X_k) - g_k(X_k)\mathbb{E}[\mathbf{P}_k R_{k+1} f(X_k)|\hat{X}_k]\|_p \\ & & + \| \left(g_k(X_k) - g_k(\hat{X}_k)\right) \mathbb{E}[\mathbf{P}_k R_{k+1} f(X_k)|\hat{X}_k]\|_p \\ & & + \|g_k(\hat{X}_k) \left(\mathbb{E}[\mathbf{P}_k R_{k+1} f(X_k)|\hat{X}_k] - \mathbb{E}[\hat{\hat{R}}_{k+1} f(\hat{X}_{k+1})|\hat{X}_k]\right) \|_p, \end{array}$$

which gives, using $\mathbb{E}[\mathbf{P}_k R_{k+1} f(X_k)|\hat{X}_k] = \mathbb{E}[R_{k+1} f(X_{k+1})|\hat{X}_k]$ according to (1.2.15):

$$\begin{aligned}\|R_k f(X_k) - \hat{\hat{R}}_k f(\hat{X}_k)\|_p &\leq \|g_k\|_\infty (2-\delta_{2,p})[\mathbf{P}_k R_{k+1} f]_{Lip} \|\Delta_k\|_p \\ &\quad + \|R_{k+1} f\|_\infty [g_k]_{Lip} \|\Delta_k\|_p \\ &\quad + \|g_k\|_\infty \|R_{k+1} f(X_{k+1}) - \hat{\hat{R}}_{k+1} f(\hat{X}_{k+1})\|_p, \\ &\leq L\left((2-\delta_{2,p}) K [R_{k+1} f]_{Lip} + \|R_{k+1} f\|_\infty\right) \|\Delta_k\|_p \\ &\quad + L \|R_{k+1} f(X_{k+1}) - \hat{\hat{R}}_{k+1} f(\hat{X}_{k+1})\|_p.\end{aligned}$$

Now, applying Lemma 1.6.1 we establish:

$$\begin{aligned}\|R_k f(X_k) - \hat{\hat{R}}_k f(\hat{X}_k)\|_p &\leq \sum_{j=k}^{n-1} L^{j-k+1} \left((2-\delta_{2,p}) K [R_{j+1} f]_{Lip} + \|R_{j+1} f\|_\infty\right) \|\Delta_j\|_p \\ &\quad + L^{n-k} \|H_n^n f(X_n) - H_n^n f(\hat{X}_n)\|_p, \\ &\leq \sum_{j=k}^{n} C_j^{k,n} \|\Delta_j\|_p,\end{aligned}$$

where:

$$C_j^{k,n} \leq \begin{cases} L^{j-k+1} \left((2-\delta_{2,p}) K [R_{j+1} f]_{Lip} + \|R_{j+1} f\|_\infty\right), & k \leq j \leq n-1, \\ L^{n-k} [H_n^n f]_{Lip} & j = n. \end{cases} \tag{1.6.6}$$

From equation (1.6.5), and using $[R_k f]_{Lip} = \|\mathrm{D} R_k f\|_\infty$, we derive that for every $k \leq j \leq n-1$:

$$\begin{aligned}C_j^{k,n} &\leq L^{j-k+1} \left((2-\delta_{2,p}) K G_f^n L^{n-j-1} \frac{K^{n-j}-1}{K-1} + 1) + G_f^n L^{n-j-1}\right) \\ &\leq G_f^n (2-\delta_{2,p}) L^{n-k} \left(K \frac{K^{n-j}-1}{K-1} + 1\right) \\ &\leq G_f^n (2-\delta_{2,p}) L^{n-k} \frac{K^{n-j+1}-1}{K-1}.\end{aligned}$$

As we have for the last step:

$$C_n^{k,n} \leq L^{n-k} G_f^n \leq L^{n-k} G_f^n (2-\delta_{2,p}).$$

Then we have for every $k \leq j \leq n$:

$$C_j^{k,n} \leq G_f^n (2-\delta_{2,p}) L^{n-k} \frac{K^{n-j+1}-1}{K-1}.$$

□

A.2 One step iterative first order scheme constants

Proposition 1.6.2 *For $0 \leq k \leq j \leq n$:*

$$D_j^{k,n} \leq \alpha_p G_f^n L^{n-k} \left(\frac{(LK)^{j-k+1}-1}{LK-1}\right)\left(\frac{K^{n-j+1}-1}{K-1}\right). \tag{1.6.7}$$

$$M_j^{k,n} \leq \frac{n+5}{2} L^{n-k} \alpha_p G_f^n \left(\frac{(LK)^{j-k+1}-1}{LK-1}\right)\left(\frac{(3K)^{n-j+1}-1}{3K-1}\right)\left(\frac{(L)^{j-k+1}-1}{L-1}\right). \tag{1.6.8}$$

Proof.
We start by establishing (1.6.7). To this end, we recall inequality (1.3.10):

$$\begin{aligned}
\|\hat{A}_k\|_p &\leq (\|\mathrm{D}R_{k+1}\|_\infty [\partial_x F_{k+1}]_{Lip}^1 + K^2 [\mathrm{D}R_{k+1} f]_{Lip}) \|\Delta_k\|_p \\
&\quad + L\left(C_{k+1}^{k+1,n} + K([\mathrm{D}R_{k+1} f]_{Lip} + \|\mathrm{D}R_{k+1} f\|_\infty)\right) \|\Delta_{k+1}\|_p \\
&\quad + L \sum_{j=k+2}^{n} C_j^{k+1,n} \|\Delta_j\|_p + LK \|\hat{A}_{k+1}\|_p,
\end{aligned}$$

and then apply Lemma 1.6.1 to write $\|\hat{A}_k\|_p \leq \sum_{j=k+1}^n D_j^{k,n} \|\Delta_j\|_p$, with:

$$\begin{aligned}
D_k^{k,n} &\leq L(\|\mathrm{D}R_{k+1} f\|_\infty [\partial_x F_{k+1}]_{Lip}^1 + K^2 [\mathrm{D}R_{k+1} f]_{Lip}), \\
D_{k+1}^{k,n} &\leq (LK+1) L\left(C_{k+1}^{k+1,n} + K([\mathrm{D}R_{k+1} f]_{Lip} + \|\mathrm{D}R_{k+1} f\|_\infty)\right), \\
D_j^{k,n} &\leq L \frac{(LK)^{j-k+1}-1}{LK-1} C_j^{k+1,n}, \quad k+2 \leq j \leq n,
\end{aligned} \tag{1.6.9}$$

which yields, using inequalities (1.6.5):

$$\begin{aligned}
D_k^{k,n} &\leq L([\partial_x F_{k+1}]_{Lip}^1 + K^2) u_{k+1} \\
&\leq L^{n-k}([\partial_x F_{k+1}]_{Lip}^1 + K^2) G_f^n \frac{(3K)^{n-k}-1}{3K-1}, \\
D_{k+1}^{k,n} &\leq L^{n-k}(LK+1) G_f^n \left((2-\delta_{2,p}) \frac{K^{n-k}-1}{K-1} + 2K \frac{(3K)^{n-k}-1}{3K-1}\right) \\
&\leq L^{n-k}(LK+1)(2-\delta_{2,p}+2K) G_f^n \frac{(3K)^{n-k}-1}{3K-1}, \\
D_j^{k,n} &\leq L^{n-k} \frac{(LK)^{j-k+1}-1}{LK-1} G_f^n (2-\delta_{2,p}) \frac{K^{n-j+1}-1}{K-1}, \quad k+2 \leq j \leq n.
\end{aligned}$$

Here we used $\frac{K^{n-j+1}-1}{K-1} \leq \frac{(3K)^{n-j+1}-1}{3K-1}$, which allows to establish inequality (1.6.7). Now for (1.6.8), we apply Lemma 1.6.1 to inequality (1.3.18):

$$\begin{aligned}
\|V_k f\|_p &\leq ([\mathrm{D}R_k f]_{Lip} + LK u_{k+1}) \|\Delta_k\|_{2p}^2 \\
&\quad + L\left(2[\mathrm{D}R_{k+1} f]_{Lip} + \tfrac{1}{2} \textstyle\sum_{j=k+1}^n D_j^{k+1,n}\right) \|\Delta_{k+1}\|_{2p}^2 \\
&\quad + \tfrac{1}{2} L \textstyle\sum_{j=k+1}^n D_j^{k+1,n} \|\Delta_j\|_{2p}^2 \\
&\quad + L \|V_{k+1} f\|_p,
\end{aligned} \tag{1.6.10}$$

which yields:

$$\|V_k f\|_p \leq \sum_{j=k}^{n} M_j^{k,n} \|\Delta\|_{2p}^2,$$

where:

$$\begin{array}{rcl} M_k^{k,n} & \leq & [\mathrm{D}R_k f]_{Lip} + LKu_{k+1}, \\ M_{k+1}^{k,n} & \leq & L(L+1)\left(2[\mathrm{D}R_{k+1} f]_{Lip} + D_{k+1}^{k+1,n} + \frac{1}{2}\sum_{j=k+2}^{n} D_j^{k+1,n}\right), \\ M_j^{k,n} & \leq & \frac{L}{2}\frac{L^{j-k+1}-1}{L-1} D_j^{k+1,n}, \quad k+2 \leq j \leq n-1, \\ M_n^{k,n} & \leq & \frac{L}{2}\frac{L^{n-k}-1}{L-1} D_n^{k+1,n} + L^{n-k}[\mathrm{D}H_n^n f]_{Lip}. \end{array} \tag{1.6.11}$$

For this last term, we use for $\|V_n f\|_p$ the upper bound given by (1.3.12) applied at $k = n$. In particular, we have:

$$\begin{array}{rcl} M_k^{k,n} & \leq & u_k + LKu_{k+1} \\ & \leq & L^{n-k} G_f^n \left(\frac{(3K)^{n-k+1}-1}{3K-1} + K\frac{(3K)^{n-k}-1}{3K-1}\right) \\ & \leq & 2L^{n-k} G_f^n \left(\frac{(3K)^{n-k+1}-1}{3K-1}\right), \\ M_j^{k,n} & \leq & \frac{L}{2}\frac{L^{j-k+1}-1}{L-1}\left(\alpha_p G_f^n L^{n-k-1}\left(\frac{(LK)^{j-k}-1}{(LK)-1}\right)\left(\frac{K^{n-j+1}-1}{K-1}\right)\right) \\ & \leq & \frac{L^{n-k}}{2}\alpha_p G_f^n \left(\frac{L^{j-k+1}-1}{L-1}\right)\left(\frac{(LK)^{j-k}-1}{(LK)-1}\right)\left(\frac{K^{n-j+1}-1}{K-1}\right), \\ & & k+2 \leq j \leq n-1. \end{array}$$

Furthermore,

$$\begin{array}{rcl} \displaystyle\sum_{j=k+2}^{n} D_j^{k+1,n} & \leq & \alpha_p G_f^n L^{n-k-1}\left(\frac{(LK)^{j-k}-1}{LK-1}\right)\left(\frac{K^{n-j+1}-1}{K-1}\right) \\ & \leq & \alpha_p G_f^n L^{n-k-1}\left(\frac{K^{n-k-1}-1}{K-1}\right)\displaystyle\sum_{j=0}^{n-k-2}(n-k-j-1)(LK)^j. \end{array}$$

Then, using inequality (1.6.12), we have:

$$\begin{array}{rcl} M_{k+1}^{k,n} & \leq & L(L+1)\left(2u_{k+1} + \alpha_p G_f^n L^{n-k-1}(LK+1)\frac{K^{n-k}-1}{K-1}\right. \\ & & \left. + \frac{1}{2}\alpha_p G_f^n L^{n-k-1}(n-k-1)\left(\frac{(LK)^{n-k}-1}{LK-1}\right)\left(\frac{K^{n-k-1}-1}{K-1}\right)\right) \\ & \leq & L^{n-k}(L+1)\alpha_p G_f^n \left(2\frac{(3K)^{n-k}-1}{3K-1} + (LK+1)\frac{K^{n-k}-1}{K-1}\right. \\ & & \left. + \frac{1}{2}\left(\frac{K^{n-k-1}-1}{K-1}\right)\displaystyle\sum_{j=0}^{n-k-2}(n-k-j-1)(LK)^j\right), \end{array}$$

$$\begin{aligned} M_{k+1}^{k,n} &\leq L^{n-k}(L+1)\alpha_p G_f^n \left(2\frac{(3K)^{n-k}-1}{3K-1} + (\frac{K^{n-k}-1}{K-1}) \sum_{j=0}^{n-k-2} \frac{n-k-j+1}{2}(LK)^j \right) \\ &\leq L^{n-k}(L+1)\alpha_p G_f^n \frac{n-k+5}{2}(\frac{(3K)^{n-k}-1}{3K-1})(\frac{(LK)^{n-k}-1}{LK-1}), \end{aligned}$$

Finally,

$$\begin{aligned} M_n^{k,n} &\leq \frac{L}{2}\frac{L^{n-k}-1}{L-1}\left(\alpha_p G_f^n L^{n-k-1}(\frac{(LK)^{n-k}-1}{(LK)-1})\right) + G_f^n L^{n-k} \\ &\leq \frac{L^{n-k}}{2} G_f^n \left(\alpha_p \frac{L^{n-k}-1}{L-1}(\frac{(LK)^{n-k}-1}{(LK)-1}) + 2\right). \end{aligned}$$

Then, inequality (1.6.8) is proved. □

A.3 Two step iterative first order scheme constants

Proposition 1.6.3 *For $0 \leq k \leq j \leq n$:*

$$D_j^{k,n} \leq \alpha_{s,p} G_f^n L^{n-k} \frac{(3K)^{n-j+1}-1}{3K-1} \tag{1.6.12}$$

$$M_j^{k,n} \leq \alpha_p G_f^n (n+1) L^{n-k} (\frac{(L)^{j-k+1}-1}{3K-1})(\frac{(3K)^{n-j+1}-1}{3K-1}) \tag{1.6.13}$$

Proof.
The proof is analogue to that of proposition 1.6.2.
We first reconsider inequality (1.4.17):

$$\begin{aligned} \|\hat{A}_{k+1}\|_p &\leq LK(u_{k+2} + \|\mathrm{D}R_{k+2}f\|_\infty)\|\Delta_{k+1}\|_{sp} \\ &\quad + L(\psi_p+1)\sum_{j=k+2}^{n} C_j^{k+2,n}\|\Delta_j\|_{sp}, \end{aligned}$$

to write:

$$\|\hat{A}_{k+1}\|_p \leq \sum_{j=k+1}^{n} D_j^{k+1,n}\|\Delta_j\|_{sp},$$

where:

$$\begin{aligned} D_{k+1}^{k+1,n} &\leq LK(u_{k+2} + \|\mathrm{D}R_{k+2}f\|_\infty), \\ D_j^{k+1,n} &\leq L(\psi_p+1)C_j^{k+2,n}, \quad k+2 \leq j \leq n. \end{aligned}$$

So in more details, using inequalities (1.6.5), we have:

$$\begin{aligned} D_{k+1}^{k+1,n} &\leq 2LKu_{k+2} \leq 2KG_f^n L^{n-k-1}\frac{(3K)^{n-k}-1}{3K-1}, \\ D_j^{k+1,n} &\leq L^{n-k-1}(\psi_p+1)(2-\delta_{2,p})G_f^n\frac{K^{n-j+1}-1}{K-1}, \quad k+2\leq j\leq n. \end{aligned}$$

This leads to result (1.6.12).
Now, according to (1.4.18), $M_j^{k,n}$ will express as previously, the difference is introduced by $D_j^{k,n}$ expression. So plugging (1.6.12) into (1.6.11) yields:

$$\begin{aligned} M_k^{k,n} &\leq 2G_f^n L^{n-k}(\frac{(3K)^{n-k+1}-1}{3K-1}), \\ M_{k+1}^{k,n} &\leq L(L+1)\left(2G_f^n L^{n-k-1}\frac{(3K)^{n-k}-1}{3K-1}+\alpha_p G_f^n L^{n-k-1}\frac{(3K)^{n-k}-1}{3K-1}\right. \\ &\quad \left.+\frac{1}{2}\alpha_p G_f^n(n-k-1)\frac{(3K)^{n-k-1}-1}{3K-1}\right) \\ &\leq \alpha_p G_f^n L^{n-k}(L+1)(n-k+1)\frac{(3K)^{n-k}-1}{3K-1}, \\ M_j^{k,n} &\leq \frac{L}{2}\frac{L^{j-k+1}-1}{L-1}\left(\alpha_p G_f^n L^{n-k-1}(\frac{(3K)^{n-j+1}-1}{(3K)-1})\right) \\ &\leq \frac{L^{n-k}}{2}\alpha_p G_f^n(\frac{L^{j-k+1}-1}{L-1})(\frac{(3K)^{n-j+1}-1}{(LK)-1}), \\ &\quad k+2\leq j\leq n-1, \\ M_n^{k,n} &\leq \frac{L}{2}\frac{L^{n-k}-1}{L-1}\left(\alpha_p G_f^n L^{n-k-1}\right)+G_f^n L^{n-k} \\ &\leq \frac{L^{n-k}}{2}\alpha_p G_f^n(\frac{L^{n-k}-1}{L-1}+1). \end{aligned}$$

□

Appendix B : The particular case of diffusion discretization

We treat the particular case where the signal equation is a diffusion. Namely, we suppose that the signal process has a continuous time evolution, within $t = 0$ and a fixed horizon $T > 0$ which follows the stochastic differential equation:

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \tag{1.6.14}$$

where B_t is a Brownian motion, $b : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : [0,T] \times \mathbb{R}^d \to \mathcal{M}_{d,d}(\mathbb{R})$ are continuously differentiable functions with bounded Lipschitz continuous derivatives with respect to x uniformly in $t \in [0,T]$. In addition σ will be assumed to satisfy a uniform ellipticity condition :
H 3' There exists $c > 0$ such that for any $x \in \mathbb{R}^d$ and $t \in [0,T]$:

$$\sigma(t,x)\sigma'(t,x) \geq cI_d$$

The observation being in discrete time, we proceed to a time dicretization via a Euler scheme of step $\delta = \frac{T}{n}$. In some situations, it is however possible to consider the diffusion itself at times $k\delta$, $k = 0, \ldots, n$. This is the case e.g. when there exists a Borel function $\varphi : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ such that $X_t = \varphi(t, B_t)$.
We define the discrete time process X_k by:

$$X_{k+1} = X_k + b_k(X_k)\delta + \sigma_k(X_k)\sqrt{\delta}\varepsilon_{k+1}, \tag{1.6.15}$$

where $\varepsilon_{k+1} = \frac{B_{(k+1)\delta} - B_{k\delta}}{\sqrt{\delta}}$ are iid, independent of X_0 such that $\varepsilon_k \sim \mathcal{N}(0, I_d)$, and for $k = 0, \ldots, n-1$, $b_k(x) = b(k\delta, x)$, $\sigma_k(x) = \sigma(k\delta, x)$.

Our aim is to establish properties of $\mathbf{P}_k$ required to apply Theorem 1.3.1 and Theorem 1.4.1 to the Euler scheme case. This also specifies the dependency in the time step $\delta = \frac{T}{n}$.

B.1 $\mathbf{P}_k$ properties

With respect to the previous notations, we denote the *Euler functional* by:

$$F_k(x, \varepsilon) = x + b_k(x)\delta + \sigma_k(x)\sqrt{\delta}\varepsilon \tag{1.6.16}$$

and the pdf of $\mathcal{N}(0, I_d)$ by $\mathbf{p}(\varepsilon) = \frac{1}{\sqrt{2\pi}^d} \exp(-\frac{|\varepsilon|^2}{2})$.

In the particular studied case, assumption **H**4 is always satisfied and, for every Lipschitz continuous function f, we have:

$$\mathrm{D}\mathbf{P}_k f(x) = -\mathbb{E}[(f \circ F_{k+1}(x, \varepsilon_{k+1}) + C^k(x))\Psi_k(x, \varepsilon_{k+1})],$$

where $\Psi_k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is such that, for every $x = (x^1, \dots, x^d)$ and $\varepsilon = (\varepsilon^1, \dots, \varepsilon^d)$:

$$\Psi_k^i(x,\varepsilon) \quad = \quad \sum_{l=1}^d \left(\frac{\partial \mathcal{G}_x^{il}(\varepsilon)}{\partial \varepsilon^l} - \varepsilon^l \mathcal{G}_x^{il}(\varepsilon) \right) \tag{1.6.17}$$

We note that, according to equation (1.6.17): $\forall \quad 0 \leq k < n$, $\forall \quad x, \varepsilon \in \mathbb{R}^d$:

$$|\Psi_k(x,\varepsilon)| \leq C(\sqrt{n}|\varepsilon| + |\varepsilon|^2).$$

The following result is adapted from Proposition 2 in [6]:

Proposition 1.6.4 *Assume $f : \mathbb{R}^d \to \mathbb{R}$ Lipschithz, then $\mathbf{P}_k f$ is differentiable and furthermore:*

$$\|\mathrm{D}\mathbf{P}_k f\|_\infty \leq C_{\sigma,b}[f]_{Lip} \quad \textit{and} \quad [\mathrm{D}\mathbf{P}_k f]_{Lip} \leq C_{\sigma,b}\sqrt{n}[f]_{Lip} \tag{1.6.18}$$

Proof.
We will temporarily consider $d = 1$ to alleviate notations. The obtained results can then be easily extended to the multidimensional case.
First, for all $x, x' \in \mathbb{R}$:

$$\begin{aligned} |\mathbf{P}_k f(x) \quad - \quad & \mathbf{P}_k f(x')|^2 \leq [f]_{Lip}^2 \mathbb{E}|F_k(x,\varepsilon_1) - F_k(x',\varepsilon_1)|^2 \\ \leq \quad & [f]_{Lip}^2 \left((1 + [b_k]_{Lip}\delta)^2 |x - x'|^2 + \delta [\sigma_k]_{Lip}^2 |x - x'|^2 \mathbb{E}[\varepsilon_1^2] \right) \\ & + [f]_{Lip}^2 2\sqrt{\delta} \left(x - x' + \delta b_k(x) - \delta b_k(x') \right) \left(\sigma_k(x) - \sigma_k(x') \right) \underbrace{\mathbb{E}[\varepsilon_1]}_{=0}, \end{aligned}$$

Then,

$$\|\mathrm{D}\mathbf{P}_k f\|_\infty \leq [\mathbf{P}_k f]_{Lip} \leq [f]_{Lip} \sqrt{\left(1 + (2[b_k]_{Lip} + [\sigma_k]_{Lip}^2)\delta + [b_k]_{Lip}^2 \delta^2 \right)}. \tag{1.6.19}$$

Since **H**4 is satisfied, we can use equation (1.4.2) to write:

$$\mathrm{D}\mathbf{P}_k f(x) = -\int f \circ F_k(x,\varepsilon) \Psi_k(x,\varepsilon) \mathbf{p}(\varepsilon) d\varepsilon, \tag{1.6.20}$$

where $\Psi_k(x,\varepsilon) = \frac{\mathrm{D}\sigma_k(x)}{\sigma_k(x)}(1 - \varepsilon^2) - \frac{\sqrt{1/\delta} + \mathrm{D}b_k(x)\sqrt{\delta}}{\sigma_k(x)}\varepsilon$.
For all $x \in \mathbb{R}$ we have $\mathbb{E}[\Psi_k(x,\varepsilon_1)] = 0$, $\mathbb{E}[|\Psi_k(x,\varepsilon_1)|] \leq C\sqrt{n}$ and $\mathbb{E}[|\varepsilon_1 \Psi_k(x,\varepsilon_1)|] \leq C\sqrt{n}$.
Consequently:

$$\begin{aligned} |\mathrm{D}\mathbf{P}_k f(x) \quad - \quad & \mathrm{D}\mathbf{P}_k f(x')| \leq \mathbb{E}[|f \circ F_k(x,\varepsilon_1) - f \circ F_k(x',\varepsilon_1)||\Psi_k(x,\varepsilon_1)|] && (1.6.21) \\ + \quad & |\mathbb{E}[\left(f \circ F_k(x',\varepsilon_1) - f \circ F_k(x,0) \right) \left(\Psi_k(x,\varepsilon_1) - \Psi_k(x',\varepsilon_1) \right)]| && (1.6.22) \\ + \quad & |f \circ F_k(x,0) \underbrace{\mathbb{E}[\Psi_k(x,\varepsilon_1) - \Psi_k(x',\varepsilon_1)]}_{=0}|. && (1.6.23) \end{aligned}$$

On one hand, for term (1.6.21) we have:

$$\begin{aligned}\mathbb{E}[|f\circ F_k(x,\varepsilon_1) &- f\circ F_k(x',\varepsilon_1)||\Psi_k(x,\varepsilon_1)|]\\ &\leq [f]_{Lip}\left((1+[b_k]_{Lip}\delta)\mathbb{E}[|\Psi_k(x,\varepsilon_1)|]+[\sigma_k]_{Lip}\sqrt{\delta}\mathbb{E}[|\Psi_k(x,\varepsilon_1)\varepsilon_1|]\right)|x-x'|.\end{aligned}$$

On the other hand, for (1.6.22):

$$\begin{aligned}\mathbb{E}[|\left(F_k(x,\varepsilon_1)\right. &- \left.F_k(x,0)\right)\left(\Psi_k(x,\varepsilon_1)-\Psi_k(x',\varepsilon_1)\right)|]\\ &\leq (1+[b_k]_{Lip}\delta)\mathbb{E}[|\Psi_k(x,\varepsilon_1)-\Psi_k(x',\varepsilon_1)|]|x-x'|\\ &\quad +\mathbb{E}[|\sigma_k(x')\sqrt{\delta}\varepsilon_1\left(\Psi_k(x,\varepsilon_1)-\Psi_k(x',\varepsilon_1)\right)|].\end{aligned}$$

Considering that:

$$\begin{aligned}\mathbb{E}[|\sigma_k(x')\sqrt{\delta}\varepsilon_1\left(\Psi_k(x,\varepsilon_1)-\Psi_k(x',\varepsilon_1)\right)|] \leq\\ \sqrt{\delta}\frac{\|\sigma\|_\infty}{\sqrt{c}}\left([\mathrm{D}\sigma_k]_{Lip}\mathbb{E}|\varepsilon_1(1-\varepsilon_1^2)|+[\mathrm{D}b_k]_{Lip}\sqrt{\delta}\mathbb{E}|\varepsilon_1|^2\right)|x-x'|\\ +\sqrt{\delta}\frac{[\sigma_k]_{Lip}}{\sqrt{c}}\mathbb{E}[|\varepsilon_1\sigma_k(x')\Psi(x',\varepsilon_1)|]|x-x'|,\end{aligned}$$

yields to derive the announced results. □

Remark 1.6.2 From (1.6.19), we note that for $k\in\{0,\dots,n-1\}$, $[\mathbf{P}_k f]^k_{Lip}\leq C$.

B.2 Case of a constant diffusion coefficient

In order to satisfy the assumptions needed for the one step recursive quantization based schemes, we have to consider the following alternative assumption:

H 2" σ is constant and $b\in\mathcal{C}^1_{b,Lip}$.

Hence, we have

$$\partial_x F_{k+1}(x,y)=1+\mathrm{D}b_k(x)\delta,$$

so that, $[\partial_x F_{k+1}]^1_{Lip}=1+\delta[\mathrm{D}b]_{Lip}$ and $\|\partial_x F\|_\infty=1+\|\mathrm{D}b\|_\infty$.
In this very particular case, we can get rid of n dependency in (1.6.18) by writing:

$$\begin{aligned}[\mathrm{D}\mathbf{P}_k f]_{Lip} &\leq [\partial_x F]^1_{Lip}\|\mathrm{D}f\|_\infty+\|\partial_x F\|_\infty[\mathrm{D}f]_{Lip}\\ &\leq \left([\partial_x F]^1_{Lip}+\|\partial_x F\|_\infty\right)\|\mathrm{D}f\|_\infty\vee[\mathrm{D}f]_{Lip}. \qquad (1.6.24)\end{aligned}$$

Chapter 2

Quantization filters and particle filters : a comparison approach

Prépublication du laboratoire de Probabilités et Modèles Aléatoires [53]

We rise a comparative study between two different approaches to construct non linear filter estimators : on one hand grid methods using zero order and first order quantization schemes, on the other hand particle filtering algorithms using sequential importance sampling or resampling. For each method, numerical implementation is explicited in addition to convergence arguments and algorithmic complexity. Numerical examples are then given over three state space models: the Kalman filter case, the canonical stochastic volatility model and the infinite dimension explicit filter introduced in [18].

Key words: Nonlinear filter, quantization schemes, particle filtering, importance sampling, Kalman filter, stochastic volatility, infinite dimension filter.

2.1 Introduction

We consider a fixed discrete horizon $n \in \mathbb{N}^*$ and some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. A signal process is an $\mathbb{R}^d$-valued discrete time hidden Markov chain $(X_k)_{0\leq k\leq n}$ evolving according to the following signal equation:

$$X_{k+1} = F_{k+1}(X_k, \varepsilon_{k+1}), \quad 0 \leq k \leq n-1, \tag{2.1.1}$$

where $F_k : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}^d$, is a Borel function and $(\varepsilon_k)_{1<k\leq n}$ is a sequence of independent identically distributed (iid) $\mathbb{R}^q$-valued random variables, independent of X_0. The distribution μ_0 of X_0 is supposed to be known. Furthermore, $\mathbf{P}_k(x, dx')$ will denote the probability transition of X_k, and:

$$\mu_0 f = \int_{\mathbb{R}^d} f(x)\mu_0(dx) \quad \textit{and} \quad \mathbf{P}_k f(x) = \int_{\mathbb{R}^d} f(x')\mathbf{P}_k(x, dx').$$

At each time step k, an $\mathbb{R}^{d'}$-valued noisy observation Y_k is made. The $(Y_k)_{0\leq k\leq n}$ dynamics are driven by a Borel function $G_k : \mathbb{R}^{d'} \times \mathbb{R}^d \times \mathbb{R}^{q'} \to \mathbb{R}^{d'}$ such that

$$Y_k = G_k(Y_{k-1}, X_k, \eta_k), \quad 1 \leq k \leq n, \tag{2.1.2}$$

where (η_k) is a sequence of iid $\mathbb{R}^{q'}$-valued random variables, independent of $\sigma(X_0, \varepsilon_k, k \geq 1)$. We assume for convenience, that $Y_0 = 0$ and that, for every $1 \leq k \leq n$, the distribution of Y_k given X_k and Y_{k-1} admits a continuous conditional pdf $y \mapsto g_k(Y_{k-1}, X_k, y)$. Furthermore, we assume that g_k satisfies the following Lipschitz assumption:
$\forall x, x' \in \mathbb{R}^d, \ \ \forall y, y' \in \mathbb{R}^{d'}$,

$$|g_k(y, x, y') - g_k(y, x', y')| \leq [g_k]_{Lip}^{y,y'}|x - x'| \quad \text{and} \quad L^{y,y'} = \max_{0\leq k\leq n} \sup_{x\in\mathbb{R}^d} |g_k(y, x, y')| < +\infty.$$

Remark 2.1.1 As the observation process is fixed, we will drop the dependancy of $[g_k]_{Lip}^{y,y'}$ and $L^{y,y'}$ in (y, y') for notational convenience.

The problem we aim to solve is to compute

$$\Pi_n f = \mathbb{E}[f(X_n)|Y_1 = y_1, \ldots, Y_n = y_n],$$

for an appropriate Borel function $f : \mathbb{R}^d \to \mathbb{R}$ and a given observation sequence $y_{1:n} = \{y_1, \ldots, y_n\}$. Througout the paper, we will use capital letters to denote random variables and small letters to designate their realizations. We will also denote by $x \mapsto p(x|y)$ the conditional pdf of a random variable X given that $Y = y$.

2.1.1 Sequential definition

In order to implement some numerical schemes solving the filtering problem, it is important to see that a recursive formulation of the problem is possible owing to the Markov property of both the signal process (X_k) and the pair signal-observation (X_k, Y_k). To establish it we set $p(x_k|y_{1:k-1})$ and $p(x_k|y_{1:k})$ respectively the prediction and the filtering pdf, for $1 \leq k \leq n$. Then for a test function f with sufficient regularity properties, we have:

$$\begin{aligned}\mathbb{E}[f(X_k)|Y_1 = y_1, \ldots, Y_{k-1} = y_{k-1}] &= \int f(x_k)p(x_k|y_{1:k-1})dx_k,\\ \mathbb{E}[f(X_k)|Y_1 = y_1, \ldots, Y_{k-1} = y_k] &= \int f(x_k)p(x_k|y_{1:k})dx_k.\end{aligned}$$

Using the Chapman-Kolmogorov formula and the Bayes formula, we can establish the following recursion:

Prediction

$$p(x_k|y_{1:k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|y_{1:k-1})dx_{k-1}. \tag{2.1.3}$$

Updating/Correction

$$p(x_k|y_{1:k}) = \frac{p(y_k|x_k, y_{1:k-1})p(x_k|y_{1:k-1})}{\int p(y_k|x_k, y_{1:k-1})p(x_k|y_{1:k-1})dx_k} = \frac{g_k(y_{k-1}, x_k, y_k)p(x_k|y_{1:k-1})}{\int g_k(y_{k-1}, x_k, y_k)p(x_k|y_{1:k-1})dx_k}. \tag{2.1.4}$$

Denoting by α a generic constant depending on the fixed observations, we can derive by induction from (2.1.4) that:

$$p(x_n|y_{1:n}) = \alpha \int \prod_{k=1}^{n-1} g_k(y_{k-1}, x_k, y_k)p(x_k|x_{k-1})dx_k\mu_0(dx_0). \tag{2.1.5}$$

Then, it easily derives from the Markov property that:

$$\int f(x_n)p(x_n|y_{1:n})dx_n = \alpha\, \mathbb{E}[\prod_{k=1}^{n-1} g_k(y_{k-1}, X_k, y_k)p(X_k|X_{k-1})].$$

This suggests to consider the so called unnormalized filter π_n defined by:

$$\pi_n f = \mathbb{E}[f(X_n) \prod_{k=1}^{n} g_k(y_{k-1}, X_k, y_k)].$$

Setting $f \equiv 1$ shows that the normalization constant equals $\frac{1}{\pi_n \mathbf{1}}$. Finally one gets

$$\Pi_n f = \frac{\pi_n f}{\pi_n \mathbf{1}}.$$

Remark 2.1.2 For convenience, the dependency of Π_n and π_n in the observation process has been omitted, as it is fixed. For the same reason, we will denote $g_k(x) := g_k(y_{k-1}, x, y_k)$ for $1 \leq k \leq n$, and $g_0 := \mathbf{1}$.

Prediction equation (2.1.3) and update equation (2.1.4) can merge to write a one step transition equation, linking $p(x_k|y_{1:k})$ to $p(x_{k-1}|y_{1:k-1})$ (and consequently the intermediate filters π_k to π_{k-1}).

$$p(x_k|y_{1:k}) \ = \alpha \ p(x_{k-1}|y_{1:k-1})p(y_k|x_k, y_{k-1})p(x_k|x_{k-1}) \tag{2.1.6}$$

By introducing the operators $(H_k)_{0\leq k\leq n}$ defined below, a sequential definition of the unnormalized filter π_n can be given.
Namely, if one defines, for any $x \in \mathbb{R}^d$:

$$\begin{cases} H_k f(x) = g_k(X_k)\mathbb{E}[f(X_{k+1})|X_k = x], \quad 0 \leq k \leq n-1, \\ H_n^n f(x) = g_n(x) f(x), \end{cases} \tag{2.1.7}$$

then we have

$$\pi_n f = \mu_0 \circ H_0 \cdots \circ H_n^n f. \tag{2.1.8}$$

Consequently, we can write sequentially, either in the forward way:

$$U_0 = \mu_0 \circ H_0 \ , \quad U_k = U_{k-1} \circ H_k \ , \quad 1 \leq k \leq n-1, \tag{2.1.9}$$

or in the backward way:

$$R_n = H_n^n \ , \quad R_k = H_k \circ R_{k+1} \ , \quad 0 \leq k \leq n-1, \tag{2.1.10}$$

so that $\pi_n f = \mu_0 R_0 f = U_{n-1} \circ H_n^n f$.

As we can see, the operators involved in the sequential definition of the filter have to be estimated numerically. Both methods we will present and compare below follow two different known approaches to approximate a conditional expectation. The first one, the quantization filtering method, is a grid method to transform an expectation into a finite weighted sum. These grids and their *companion* weights can be pre-computed and stored off line. As concerns filtering, this method has been introduced in [41], where the authors use optimal quantizers to construct what we will call a zero order quantization filter. Further developments of this method can be found in Chapter 1, where the stationary property of optimal quantizers is used to develop some schemes based on first order approximations. In section 3, we recall the construction of these schemes, as well as the most important convergence results. Section 4 deals with the second filtering apporach we will be interested in. It is a Monte Carlo particle method using importance sampling to simulate random grids. This method has been deeply developed by practionners of filtering in different application fields (see [15]). Two algorithms will be focused on, the sequential importance sampling (**SIS**) and the sequential importance resampling (**SIR**).

2.2 Quantization based filters

2.2.1 Zero order scheme

A process quantizer size $(N_k)_{0\leq k\leq n}$ being fixed, and a quantizer Γ_k being precomputed, we set $(\hat{X}_k)_{0\leq k\leq n}$ a marginal quantization of $(X_k)_{0\leq k\leq n}$ defined by:

$$\hat{X}_k = Proj_{\Gamma_k}(X_k).$$

Quantization filters are constructed using recursive schemes. The zero order scheme introduced in [41] uses quantizers to approximate operators H_k by piecewise constant operators defined on the grids Γ_k. Namely, we define:

$$\begin{cases} \hat{\hat{H}}_k f(\hat{X}_k) = g_k(\hat{X}_k)\mathbb{E}[f(\hat{X}_{k+1})|\hat{X}_k], \quad 0 \leq k \leq n-1, \\ \hat{\hat{H}}_n^n f(\hat{X}_n) = H_n^n f(\hat{X}_n) = g_n(\hat{X}_n) f(\hat{X}_n). \end{cases} \tag{2.2.1}$$

So, defining $\hat{\mu}_0$ as the (discrete) distribution of $\hat{X}_0$, we have respectively the following forward and backward iterative approximation schemes :

$$\hat{\hat{U}}_0 = \hat{\mu}_0 \hat{\hat{H}}_0 \ , \quad \hat{\hat{U}}_k = \hat{\hat{U}}_{k-1} \circ \hat{\hat{H}}_k \ , \quad 1 \leq k \leq n-1, \tag{2.2.2}$$

and

$$\hat{\hat{R}}_n = H_n^n \ , \quad \hat{\hat{R}}_k = \hat{\hat{H}}_k \circ \hat{\hat{R}}_{k+1} \ , \quad 0 \leq k \leq n-1, \tag{2.2.3}$$

so that

$$\hat{\hat{\pi}}_n f = \hat{\mu}_0 \hat{\hat{R}}_0 f = \hat{\hat{U}}_{n-1} \circ H_n^n f.$$

This formulation of the zero order scheme provides an implementable solution to the non linear filtering problem. This scheme consists of a recursive procedure based on weighted sums involving the optimal quantization grids and some weights, which only depend on the signal distribution. These quantities are usually computed off line, since they do not depend on the observation process.
To be more specific, we denote $\Gamma_k = \{x_k^1, ..., x_k^{N_k}\}$ and $p_k^{ij} = \mathbb{E}[\mathbf{1}_{\{\hat{X}_{k+1}=x_{k+1}^j\}}|\hat{X}_k = x_k^i] \in \mathbb{R}$, then the implemented forward algorithm writes as Algorithm 0.

Note that $(\hat{\mu}_0^i)_{1\leq i\leq N_0}$ designates the vector of discrete probabilities : $\hat{\mu}_0^i = \mathbb{P}(\hat{X}_0 \in \mathbf{C}_i(\Gamma_0))$. The use of the backward formulation (2.2.3) associated to the L^2-optimal marginal (N_k)-quantizations of (X_k), allows to establish a convergence rate to zero for the error estimate, for a Borel test function f satisfying $H_n^n f$ bounded Lipschitz continuous, when the quantizer size $N = \sum_{k=0}^n N_k$ goes to $+\infty$. Namely,

$$|\pi_n f - \hat{\hat{\pi}}_n f| = O(N^{-1/d}),$$

(see [41]), with a constant depending on the observation vector and on the time horizon n.

Algorithm 0 Zero order quantization based algorithm

$k=0$	$1 \leq i \leq N_0$ and $1 \leq j \leq N_1$
	$\hat{U}_0(j) = \sum_{i=1}^{N_0} \hat{\mu}_0^i p_0^{ij}$
$0 \leq k \leq n-2$	$1 \leq i \leq N_{k+1}$ and $1 \leq j \leq N_{k+2}$
	$\hat{\hat{H}}_{k+1}(i,j) = g_{k+1}(x_{k+1}^j) p_{k+1}^{ij}$
	$\hat{\hat{U}}_{k+1}(j) = \sum_{i=1}^{N_k} \hat{\hat{H}}_{k+1}(i,j) \hat{\hat{U}}_k(i)$
$k=n$	$1 \leq j \leq N_n$
	$H_n^n f(x_n^j) = g_n(x_n^j) f(x_n^j)$
	$\hat{\hat{\pi}}_n f = \sum_{j=1}^{N_n} \hat{\hat{U}}_{n-1}(j) H_n^n f(x_n^j)$

2.2.2 First order schemes

In order to obtain better convergence rates, we introduce first order schemes (see [6, 43] and Chapter 1). The main idea to develop them is to write a piecewise linear approximation of $R_k f$, mimicking some first order Taylor expansion (see [6, 43]) and using differential terms to introduce first order correctors. The generic first order scheme could be written as follows:

$$\begin{cases} \widehat{R}_n f(\hat{X}_n) &= H_n^n f(\hat{X}_n), \\ \widehat{R}_k f(x_k^i) &= g_k(x_k^i)\mathbb{E}[\widehat{R}_{k+1}(\hat{X}_{k+1}) + \langle \widehat{DR}_{k+1} f, \Delta_{k+1} \rangle | \hat{X}_k = x_k^i], \\ & \qquad 0 \leq k \leq n-1. \end{cases} \tag{2.2.4}$$

where $\widehat{DR}_{k+1} f$ is a numerical approximation of $\mathrm{D}R_{k+1} f$ when it exists and $< ., . >$ denotes the Euclidean inner product in $\mathbb{R}^d$.
A first order quantization based unnormalized filter estimator, is then defined by:

$$\hat{\pi}_n f = \hat{\mu}_0 \widehat{R}_0 f.$$

The way the approximation $\widehat{DR}_{k+1} f$ is defined leads to several variants. To establish error estimates, one makes an extensive use of the stationarity property of L^2-optimal quantizers i.e.

$$\mathbb{E}[X_k | \hat{X}_k] = \hat{X}_k. \tag{2.2.5}$$

In the following, we will present the two schemes constructed in Chapter 1: the one step recursive scheme based on a recursive definition of the differential term estimator $\widehat{DR}_k$ and the two step recursive scheme based on an integration by part transformation of conditional expectation derivative $\mathrm{D}\mathbf{P}_k R_{k+1}$.
To establish such first order scheme, we need F_k to be differentiable and the test function f to satisfy $H_n^n f$ differentiable with continous Lipschitz bounded derivatives (in Chapter 1, we see that we can relax this assumption in case of regularizing semi-grroups $\mathbf{P}_k$).

2.2.2.1 One step recursive first order scheme

The recursive definition of the differential term estimator is given in Chapter 1 by:

$$\begin{cases} \widehat{DR}_n(\hat{X}_n)f = \mathrm{D}H_n^n f(\hat{X}_n), \\ \widehat{DR}_k f(\hat{X}_k) = \mathrm{D}g_k(\hat{X}_k)\mathbb{E}[\hat{\hat{R}}_{k+1}f(\hat{X}_{k+1})|\hat{X}_k] + g_k(\hat{X}_k)\mathbb{E}[\partial_x F_k(X_k,\varepsilon_{k+1})'\widehat{DR}_{k+1}f(\hat{X}_{k+1})|\hat{X}_k] \\ \quad k = 0,\dots,n-1. \end{cases} \tag{2.2.6}$$

Combined with (2.2.6), scheme (2.2.4) is completely computable, as it can be rewritten easily using finite weighted sums. The first order corrector terms introduce new companion parameters γ_k^{ij}, and δ_k^{ij} defined by:

$$\begin{aligned} \gamma_k^{ij} &= \mathbb{E}[\partial_x F_k(X_k,\varepsilon_{k+1})'\mathbf{1}_{\{\hat{X}_{k+1}=x_{k+1}^j\}}|\hat{X}_k = x_k^i] \in \mathcal{M}_d(\mathbb{R}), \\ \delta_k^{ij} &= \mathbb{E}[\Delta_{k+1}\mathbf{1}_{\{\hat{X}_{k+1}=x_{k+1}^j\}}|\hat{X}_k = x_k^i], \\ &= \mathbb{E}[(X_{k+1}-x_{k+1}^j)\mathbf{1}_{\{\hat{X}_{k+1}=x_{k+1}^j\}}|\hat{X}_k = x_k^i] \in \mathbb{R}^d, \end{aligned} \tag{2.2.7}$$

which can be computed off-line during the quantization operation. The on line computation cost is reduced to the computation of $\widehat{R}_k$, $\hat{\widehat{R}}_k$ and $\widehat{DR}_k$ operators.
Moreover, the resulting scheme can be reformulated in a forward way, which allows to consider problems with intermediate filtering dates or multiple test functions without adding heavy computations. The implemented algorithm, Algorithm 1, uses a forward formulation based on matricial operators (see Chapter 1 for details).
Forward expression

$$\hat{\mathcal{U}}_0 = \hat{\mu}_0\widehat{\mathcal{H}}_0 \quad \textit{and} \quad \hat{\mathcal{U}}_k = \hat{\mathcal{U}}_{k-1}\widehat{\mathcal{H}}_k \quad k = 1,\dots,n-1,$$

where $\widehat{\mathcal{H}}_k$ is a lower triangular operator matrix defined by $\widehat{\mathcal{H}}_k = \begin{pmatrix} \hat{H}_k^1 & 0 & 0 \\ \hat{H}_k^2 & \hat{H}_k^3 & 0 \\ 0 & \hat{H}_k^4 & \hat{H}_k^1 \end{pmatrix}$.

Using some additionnal assumptions on $\mathbf{P}_k$, g_k and F_k (see Chapter 1), we establish for L^2-optimal N_k-quantizers of (X_k), that:

$$|\pi_n f - \hat{\pi}_n f| = O(N^{\frac{-2}{d}}).$$

2.2.2.2 Two step recursive first order scheme

This method is based on an integration by part formula (see [6, 7, 3]) which leads to write $\mathrm{D}\mathbf{P}_k R_{k+1}f$ as a weighted conditional expectation, namely:

$$\mathrm{D}\mathbf{P}_k R_{k+1}f(x) = \mathbb{E}[R_{k+1}f(X_{k+1})\Psi_k(X_k,\varepsilon_{k+1})|X_k = x],$$

where $\Psi_k(X_k, \varepsilon_{k+1}) = (\Psi^i(x,\varepsilon))_{0\leq i\leq d}$ and

$$\begin{aligned}\Psi^i : \mathbb{R}^d \times \mathbb{R}^d &\to \mathbb{R} \\ (x,\varepsilon) &\mapsto \operatorname{div}\left(\partial_\varepsilon F(x,\varepsilon)^{-1}\partial_x F(x,\varepsilon)\right)^i + \frac{1}{\mathbf{p}(\varepsilon)}\langle\left(\partial_\varepsilon F(x,\varepsilon)^{-1}\partial_x F(x,\varepsilon)\right)^i, \mathrm{D}\mathbf{p}(\varepsilon)\rangle\end{aligned}$$

where $(\partial_\varepsilon F(x,\varepsilon)^{-1}\partial_x F(x,\varepsilon))^i$ designates the i^{th} line of the matrix $(\partial_\varepsilon F(x,\varepsilon)^{-1}\partial_x F(x,\varepsilon))$. Note that some assumptions on both F_k and the signal innovation pdf $\mathbf{p}$ are needed to allow such a transformation (see Chapter 1 for details). The resulting formula for $\widehat{R}_k$, $0 \leq k \leq n-1$, is consequently:

$$\begin{aligned}\widehat{R}_k f(\hat{X}_k) = g_k(\hat{X}_k)\widehat{\mathbf{P}}_k\widehat{R}_{k+1}f(\hat{X}_k) + g_k(\hat{X}_k)\times \\ \Big(\mathbb{E}[\langle \mathrm{D}g_{k+1}(\hat{X}_{k+1})\widehat{\mathbf{P}}_{k+1}\hat{\widehat{R}}_{k+2}f(\hat{X}_{k+1}), \Delta_{k+1}\rangle|\hat{X}_k] - \mathbb{E}[\,\langle g_{k+1}(\hat{X}_{k+1})\times \\ \mathbb{E}[\hat{\widehat{R}}_{k+2}f(\hat{X}_{k+2})\Psi_{k+1}(X_{k+1},\varepsilon_{k+2})|\hat{X}_{k+1}], \Delta_{k+1}\rangle|\hat{X}_k]\Big)\,. \quad (2.2.8)\end{aligned}$$

This formula introduces a new weight λ_k^{ij}:

$$\lambda_k^{ij} = E[\Psi_k(X_k,\varepsilon_{k+1})\mathbf{1}_{\{\hat{X}_{k+1}=x_{k+1}^j\}}|\hat{X}_k = x_k^i] \in \mathbb{R}^d$$

Depending only on the signal parameters, it can be precomputed like p_k^{ij} and δ_k^{ij} and kept off line in accessible tables. As for the previous first order scheme, a forward transcription of the scheme is possible using finite weighted sums which gives the following implemented algorithm, Algorithm 2.

Under some additional assumptions on the weight function Ψ_k, we insure a convergence rate of order $N^{-\frac{2\rho}{d}}$ where $\rho \in]\frac{1}{2}, 1)$ (see Chapter 1).

Algorithm 1 One step recursive 1st order quantization based filtering scheme

$k=0$

$1 \le i \le N_0$ and $1 \le j \le N_1$

$\hat{U}_0^{11}(j) = \sum_{i=1}^{N_0} \mu_0^i p_0^{ij}$ $\quad$ $\hat{U}_0^{21}(j) = 0 \in \mathbb{R}^d$

$\hat{U}_0^{22}(j) = \sum_{i=1}^{N_0} \mu_0^i \gamma_0^{ij}$ $\quad$ $\hat{U}_0^{31}(j) = 0 \in \mathbb{R}$

$\hat{U}_0^{32}(j) = \sum_{i=1}^{N_0} \mu_0^i \delta_0^{ij}$ $\quad$ $\hat{U}_0^{33}(j) = \hat{U}_0^{11}(j)$

$0 \le k \le n-2$

$1 \le i \le N_{k+1}$ and $1 \le j \le N_{k+2}$

$\hat{H}_{k+1}^1(i,j) = g_{k+1}(x_{k+1}^i) p_{k+1}^{ij}$ $\quad$ $\hat{H}_{k+1}^2(i,j) = \mathrm{D}g_{k+1}(x_{k+1}^i) p_{k+1}^{ij}$

$\hat{H}_{k+1}^3(i,j) = g_{k+1}(x_{k+1}^i) \gamma_{k+1}^{ij}$ $\quad$ $\hat{H}_{k+1}^4(i,j) = g_{k+1}(x_{k+1}^i) \delta_{k+1}^{ij}$

$$\begin{aligned}
\hat{U}_{k+1}^{11}(j) &= \textstyle\sum_{i=1}^{N_{k+1}} \hat{U}_k^{11}(i) \hat{H}_{k+1}^1(i,j) \\
\hat{U}_{k+1}^{21}(j) &= \textstyle\sum_{i=1}^{N_{k+1}} \hat{U}_k^{21}(i) \hat{H}_{k+1}^1(i,j) + \hat{U}_k^{22}(i) \hat{H}_{k+1}^2(i,j) \\
\hat{U}_{k+1}^{22}(j) &= \textstyle\sum_{i=1}^{N_{k+1}} \hat{U}_k^{22}(i) \hat{H}_{k+1}^3(i,j) \\
\hat{U}_{k+1}^{31}(j) &= \textstyle\sum_{i=1}^{N_{k+1}} \hat{U}_k^{31}(i) \hat{H}_{k+1}^1(i,j) + \langle \hat{U}_k^{32}(i), \hat{H}_{k+1}^2(i,j) \rangle \\
\hat{U}_{k+1}^{32}(j) &= \textstyle\sum_{i=1}^{N_{k+1}} \hat{H}_{k+1}^3(i,j) \hat{U}_k^{32}(i) + \hat{U}_k^{33}(i) \hat{H}_{k+1}^4(i,j) \\
\hat{U}_{k+1}^{33}(j) &= \textstyle\sum_{i=1}^{N_{k+1}} \hat{U}_k^{33}(i) \hat{H}_{k+1}^1(i,j)
\end{aligned}$$

$k=n$

$H_n^n f(x_n^j) = g_n(x_n^j) f(x_n^j)$ $\quad$ $H_n^n f(x_n^j) = g_n(x_n^j) f(x_n^j)$

$\hat{\pi}_n f = \sum_{j=1}^{N_n} (\hat{U}_{n-1}^{33}(j) + \hat{U}_{n-1}^{31}(j)) H_n^n f(x_n^j) + \langle \hat{U}_{n-1}^{32}(j), \mathrm{D}H_n^n f(x_n^j) \rangle$

2.3 Particle filters

In this section, we provide some background on two classical Monte Carlo particle methods for non linear filters. Based on online simulations, this approach provides, like quantization based methods do, a reformulation of conditional expectations as finite weighted sums. This time the sum terms are random variables. The following subsections deal with two basic particle filtering methods using an importance sampling technique.

2.3.1 Sequential Importance Sampling (SIS)

Suppose we are able to simulate N independent random vectors $(X^i)_{1 \le i \le N}$, distributed according to $\mathcal{L}(X_n | Y_1 = y_1, \ldots, Y_n = y_n)$. An estimator $\Pi_n^N f$ of the filter would be:

$$\Pi_n^N f = \frac{1}{N} \sum_{i=1}^{N} f(X^i).$$

In other words, from a quantization viewpoint, we approximate the *a posteriori* distribution by the set of random samples X^i with weights $\frac{1}{N}$. This estimate is unbiased and its rate of convergence is ruled by central limit theorem (see e.g. [49, 15]).
Unfortunately, the initial assumption that the distribution $\mathcal{L}(X_n | Y_1 = y_1, \ldots, Y_n = y_n)$ is

Algorithm 2 Two step recursive 1st order quantization based filtering scheme

$0 \le k \le n-2$ $\quad 1 \le i \le N_k$, $1 \le j \le N_{k+1}$ and $1 \le l \le N_{k+2}$ $\qquad (g_0 := \mathbf{1})$

$\hat{H}_k^1(i,j) = g_k(x_k^j)p_k^{ij}$

$\hat{H}_k^2(i,l) = g_k(x_n^i)\sum_{j=1}^{N_{k+1}} \langle p_{k+1}^{jl} \mathrm{D}g_{k+1}(x_{k+1}^j) - g_{k+1}(x_{k+1}^j)\lambda_{k+1}^{jl}, \delta_k^{ij} \rangle$

Initial $\quad 1 \le j \le N_1$ and $1 \le l \le N_2$

$\hat{U}_0^1(j) = \sum_{i=0}^{N_0} \hat{\mu}_0^i \hat{H}_0^1(i,j)$

$\hat{U}_2^2(l) = \sum_{i=1}^{N_k} \hat{\mu}_0^i \hat{H}_0^2(i,l)$

Transition $\quad 1 \le k \le n-2$, $1 \le j \le N_{k+1}$ and $1 \le l \le N_{k+2}$

$\hat{U}_{k+2}^2(l) = \sum_{i=1}^{N_k} \hat{H}_k^2(i,l)\hat{U}_k^1(i) + \sum_{j=1}^{N_{k+1}} \hat{H}_{k+1}^1(j,l)\hat{U}_{k+1}^2(j)$

$\hat{U}_k^1(j) = \sum_{i=1}^{N_k} \hat{H}_k^1(i,j)\hat{U}_{k-1}^1(i)$

Final $\quad H_n^n f(x_n^j) = g_n(x_n^j)f(x_n^j) \qquad \mathrm{D}H_n^n f(x_n^j) = f(x_n^j)\mathrm{D}g_n(x_n^j) + g_n(x_n^j)\mathrm{D}f(x_n^j)$

$\hat{\pi}_n f = \sum_{j=1}^{N_n} \hat{U}_n^2(j) H_n^n f(x_n^j) +$

$\sum_{i=1}^{N_{n-1}} \hat{U}_{n-2}^1(i) g_{n-1}(x_{n-1}^i) \sum_{j=1}^{N_n} \left(H_n^n f(x_n^j) p_{n-1}^{ij} + \langle \mathrm{D}H_n^n f(x_n^j), \delta_{n-1}^{ij} \rangle \right)$

simulatable is usually not satisfied. This distribution is unknown as it is the one we aim to approximate by the filtering procedure. In simulation, this problem is often overpassed by applying an importance sampling procedure [49]. Namely, we choose a simulatable distribution $q(x_n|y_{1:n})$, which we will call *importance distribution* and write:

$$\Pi_n f = \int f(x_n)\omega(x_n)q(x_n|y_{0:n})dx_n,$$

where $\omega(x) = \frac{p(x|y_{0:n})}{q(x|y_{1:n})}$ will be called the *importance weight* function. Now, if we designate by $(X^i)_{1\le i\le N}$ N iid samples of distribution q, a new definition of the filter estimator can be given by:

$$\Pi_n^N f = \sum_{i=1}^{N} f(X^i)\omega(X^i).$$

We approximate the *a posteriori* distribution by the set of the N random samples (or particles) X^i and the associated importance weights. In our case, we always know a distribution which can be chosen as an importance distribution: that of the hidden signal (X_k), so we define:

$$q(x_{0:n}|y_{0:n}) = p(x_{1:n}) = \mu_0(x_0)\prod_{k=1}^{n} p(x_k|x_{k-1}). \tag{2.3.1}$$

This choice gives the advantage of making possible a sequential approximation of the particles and weights $\{x_k^i, \omega_k(x_k^i)\}$ via simulation, by the use of the prediction and update

equations (2.1.3) and (2.1.4). At each time step $0 \leq k \leq n$, we simulate an N-sample $(X^i_{0:n})$ having q as a pdf. The filter estimator at each date $1 \leq k \leq N$ is then:

$$\Pi^{q,N}_k f = \sum_{i=1}^{N} f(X^i_k)\omega_k(X^i_{0:k}),$$

with the sequential update equation for the importance weights:

$$\begin{aligned} \omega_{k+1}(x^i_{0:k+1}) &= \alpha \ \frac{p(x^i_{k+1}|y_{1:k+1})}{p(x^i_{0:k+1})} \\ &= \alpha \ \frac{p(x^i_k|y_{1:k})p(y_{k+1}|x^i_{k+1},y_k)p(x^i_{k+1}|x^i_k)}{p(x^i_{k+1}|x^i_k)p(x^i_{0:k})} \\ &= \alpha \ \omega_k(x^i_{0:k})g_{k+1}(y_k,x^i_{k+1},y_{k+1}) \qquad (2.3.2) \end{aligned}$$

$$\text{and} \qquad \omega_{k+1}(x^i_{0:k+1}) = \frac{\omega_k(x^i_{0:k})g_{k+1}(y_k,x^i_{k+1},y_{k+1})}{\sum_{i=1}^N \omega_k(x^i_k)g_{k+1}(y_k,x^i_{k+1},y_{k+1})} \qquad (2.3.3)$$

Hence, we can implement sequentially Algorithm 3.

Algorithm 3 Algorithm SIS

- Simulate from the signal initial distribution $\{(x^i_0)_{1\leq i\leq N}, \frac{1}{N}\}$
- $*$ At a given date $0 \leq k \leq n-1$, we have $\{(x^i_k)_{1\leq i\leq N}, (\omega_k)_{1\leq i\leq N}\}$
- Simulate particles $x^i_{k+1} \sim p(x_{k+1}|x^i_k)$
- Update weights using equation (2.3.2) and (2.3.3), with the observation y_{k+1}
- Go to $*$ for date $k+1$

The convergence rate of the method is independent of the signal dimension, since it is of Monte Carlo type. For each bounded continuous test function f, we have:

$$\mathbb{E}[|\Pi_{y,n}f - \Pi^{q,N}_{y,n}f|] \leq \frac{m_n}{\sqrt{N}}\|f\|_\infty. \qquad (2.3.4)$$

where m_n is a constant depending on n and on the observation.

2.3.2 Sequential Importance Resampling or Bootstrap filter (SIR)

The method described previously suffers from the problem of weight degeneracy, as the sequential definition implies to multiply repeatedly the likelihood terms. This produces sometimes particles with very small weights, that we carry out throughout the estimation although their contribution to the distribution description is negligible. This problem occurs for example when the importance distribution is badly adjusted to the filter distribution.

One solution to the degeneracy problem is introduced in [24, 1]. It suggests to diffuse

sequentially equally weighted particles by adding a resampling step to the previous algorithm. This new method called *Sequential Importance Resampling* (**SIR**) or Bootstrap filter uses interaction between particles to eliminate weakly weighted ones. As an intereaction phase is introduced, samples are no longer independent and Monte Carlo arguments cannot be used to establish a convergence rate like in the previous case. Nevertheless, in [15, 36, 14], it is shown that by the means of convenient resampling procedures, for example the multinomial sampling procedure, a convergence rate of type (2.3.4) can be established.

Algorithm 4 Algorithm SIR

- Simulate from the signal initial distribution $(x_0^i)_{1\leq i\leq N}$
- $*$ At a given date $0 \leq k \leq n-1$, we have $(x_k^i)_{1\leq i\leq N}$
- Simulate particles $x_{k+1}^i \sim p(x_{k+1}|x_k^i)$
- Update weights using equations (2.3.2) and (2.3.3), with the observation y_{k+1}
- Resampling step:
  Simulate N samples i_{k+1}^j from the discrete multinomial distribution
  with parameters $(\omega_{k+1}^i)_{1\leq i\leq N}$
- Go to $*$ with date $k+1$ and modified particles $(x_{k+1}^{i_{k+1}^j})_{1\leq j\leq N}$

The choice of resampling procedures, also called branching methods, can be conditionned by variance reduction criteria, numerical complexity or convergence rate preserving [14, 15]. The common point they share is that they aim to obtain an *unweighted* empirical distribution to sequentially approximate intermediate filtering pdf. Namely, they achieve that $\mathbb{E}[\mathbf{1}_{\{i_k^j=i\}}] = \omega_k^i$.

2.3.3 Elements for a comparison

As mentioned above, the underlying principles of numerical methods in sequential non linear filtering are the same (see [34]). Both particle and quantization based methods use the approximation of the objective distribution by a finite state one, so that the final expression of the filter estimator appears as a finite weighted sum. The difference lies in the construction of such an approximation. For quantization filters, we use off line precomputed marginal distribution quantizers. For particle filters, the grids are random samples of the same distributions, that need to be computed on line.
Following this remark, we see that both approaches are similar, and we expect that they behave the same way when treating comparable state models and observations. However some differences deserve to be pointed out.

As a probabilistic method, particle filters give a random solution to the filtering problem. This is a point we do not have to manage when treating with quantization methods.

Algotithm 0	C_0N^2
Algotithm 1	$C_1N^2d^3$
Algorithm 2	C_2N^3d
Algorithm 3	C_3N
Algorithm 4	C_4N

Table 2.1: Comparison of complexity degrees for different numerical filtering algorithms

Conversly, being based on Monte Carlo convergence arguments, particle methods do not suffer from dimension dependency when considering their theoretical convergence rate, whereas quantization based methods do depend on the dimension of the state space. Considering the theoretical convergence results, quantization methods are still competitive till dimension 2 for zero order schemes and till dimension 4 for first order ones.

From an algorithmic viewpoint, some more differences deserve to be mentionned. It is about the complexity of each algorithm, summerized in Table 2.1.

Owing to the off line computations of the quantizer grids, constants C_0, C_1 and C_3 for quantization filters represent elementary operations computation cost. For particle filters, C_3 and C_4 include simulation cost, and could be dependent of N which results in more complex algorithms. This occurs for **SIR** algorithm with some particular resampling algorithms.

Finally, we should remark that quantization methods need smaller grid sizes than Monte Carlo methods to attain convergence regions. This will be pointed out in numerical results below. This fact, in some cases, compensates the relatively high complexity range of quantization methods, particularily in low signal dimensions.

2.4 State Equations

We aim to compare numerical performances of the two approaches. In this paragraph, we briefly present three models chosen to make up the benchmark. As a first step, we consider the Kalman filter, a typical case where an exact explicit solution can be computed. Then, we will consider a canonical stochastic volatility model, issued from the discretization of a diffusion process representing an asset price dynamics on a financial market. We will finally examine the nonlinear filter case introduced in [18] by V. Genon Catalot, for which some semi-closed forms of the solution are available. It can also be adapted to treat asset prices.

2.4.1 Kalman filter (KF)

For this model, both signal and observation equations are linear with Gaussian independent noises. It is well known that the filter in this case is a Gaussian process which parameters (the two first moments) can be computed sequentially by a deterministic algorithm (KF), (see e.g. [21] for details). We set:

$$\begin{cases} X_k = \rho X_{k-1} + \theta\varepsilon_{k+1}, \quad X_0 \sim \mathcal{N}(m_0, \Sigma_0'\Sigma_0) \\ Y_k = X_k + \alpha\eta_k, \\ \varepsilon_k \ \ and \ \ \eta_k \quad iid \quad \sim \mathcal{N}(0, I_d), \\ \rho, \ \ \theta, \ \ \alpha \quad \in \mathcal{M}_d(\mathbb{R}). \end{cases} \tag{2.4.1}$$

2.4.2 Canonical stochastic volatility model (SVM)

This is the time discretization of a continuous diffusion model commonly used in finance. The stock price S_t and its volatility σ_t solve the following stochastic differential system:

$$\begin{cases} dS_t = \frac{1}{2}\sigma_t^2 S_t dt + \sigma_t S_t dW_t^1, \\ d(log(\sigma_t^2)) = -\lambda log(\sigma_t^2)dt + \tau dW_t^2. \end{cases} \tag{2.4.2}$$

So, the Euler scheme with time step Δ writes:

$$\begin{cases} \log(\frac{S_{k+1}}{S_k}) = \sigma_k\sqrt{\Delta}\eta_k, \\ \log(\sigma_{k+1}^2) - \log(\sigma_k^2) = -\lambda\log(\sigma_k^2)\Delta + \tau\sqrt{\Delta}\varepsilon_{k+1}. \end{cases} \tag{2.4.3}$$

where η_k and ε_k are iid $\mathcal{N}(0,1)$.
Now, setting $Y_k = log(\frac{S_{k+1}}{S_k})$ and $X_k = log(\sigma_k^2)$ leads to the following discrete time state equations.

$$\begin{cases} X_k = \rho X_{k-1} + \theta\varepsilon_{k+1}, \\ Y_k = exp(\frac{X_k}{2})\eta_k, \\ \varepsilon_k \ \ and \ \ \eta_k \quad iid \quad \sim \mathcal{N}(0,1), \\ \rho = (1-\lambda\Delta) \ \ and \quad \theta = \tau\sqrt{\Delta} \in \mathbb{R}_+^*. \end{cases} \tag{2.4.4}$$

When $\Delta \in (0, \frac{2}{\lambda})$ then $\rho = (1-\lambda\Delta) \in (-1,1)$ and (X_k) is a positively recurrent Markov chain which converges geometrically to its invariant distribution $\mathcal{N}(0, \frac{\theta^2}{1-\rho^2})$ (with respect to the total variation metric). The non linearity introduced in the observation equation makes impossible to determine exactly the filtering distribution. This is a case where numerical methods are necessary to solve the filtering problem.

2.4.3 Explicit non linear filter [18]

In this example, we treat a non linear non Gaussian state equation introduced by V. Genon Catalot in [18]. When both the noise distribution and the initial signal distribution are

specified in an appropriate way, it is shown how to construct an infinite dimensional *explicit* non-linear filter, in the sense that all parameters of the a posteriori pdf can be determined by recursive explicit schemes.
For that we introduce the family of the so called *Serial Gaussian* distributions $SG(\sigma^2, (\alpha_i)_{i\geq 0})$ (see [18]) which probability density functions are defined by:

$$u \mapsto \left(\sum_{i\geq 0} \alpha_i \frac{u^{2i}}{\sigma^{2i} C_{2i}} \right) \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{u^2}{2\sigma^2}),$$

where:

- $C_{2i} = \frac{(2i)!}{2^i i!}$ is the $2i^{th}$ moment of a $\mathcal{N}(0,1)$ distribution.
- $\forall i \geq 0$, $\alpha_i \geq 0$ and $\sum_{i\geq 0} \alpha_i = 1$. We will denote $\alpha = (\alpha_i)_{i\geq 0}$.
- $\sigma > 0$.

$SG(\sigma^2, \alpha)$ can be seen as a mixture distribution with a scale parameter σ and a mixture parameter α. In a filtering context, considering the following state equations:

$$\begin{cases} X_k = \rho X_{k-1} + \theta \varepsilon_{k+1} \\ Y_k = X_k \eta_k \end{cases} \quad \text{where} \quad \begin{cases} \varepsilon_k & iid & \sim \mathcal{N}(0,1), \\ \eta_k & iid & \sim \mathcal{B}(\pm 1, \frac{1}{2}) \times \sqrt{\frac{1}{\mathcal{E}(\lambda)}}, \\ X_0 & & \sim SG(\sigma_0^2, (\alpha_i^0)_{i\geq 0}), \end{cases}$$

it is established that the filter distribution as well as the prediction one will be $SG(\sigma^2, (\alpha_i)_{i\geq 0})$, with parameters that can be evaluated sequentially. We will denote for $0 \leq k \leq n$:

$$\begin{aligned} \mathcal{L}(X_k | Y_{1:k}) &\sim SG(\hat{\sigma}_k^2, (\hat{\alpha}_i^k)_{i\geq 0}), \\ \mathcal{L}(X_{k+1} | Y_{1:k}) &\sim SG(\sigma_{k+1}^2, (\alpha_i^{k+1})_{i\geq 0}), \end{aligned}$$

where scale and mixture parameters are defined by Algorithm 5:

Algorithm 5 Explicit filter

Initial step : $\mathcal{L}(X_0) \sim SG(\sigma_0^2, (\alpha_i^0)_{i\geq 0})$
Transition from date k to $k+1$:

Update:	$\hat{\sigma}_k(Y_k)$	$=$	$\sigma_k \frac{\lvert Y_k \rvert}{\sqrt{Y_k^2 + 2\lambda\sigma_k^2}}$
	$\hat{\alpha}_0^k(Y_k)$	$=$	0
	$\hat{\alpha}_i^k(Y_k)$	$=$	$\alpha_{i-1}^k \frac{f_{i-1}(\frac{Y_k}{\sqrt{\lambda}\sigma_k})}{\sum_{i\geq 0} \alpha_i^k f_i(\frac{Y_k}{\sqrt{\lambda}\sigma_k})}$ where $f_i(z) = \frac{(2i+1)z^{2i}}{(z^2+2)^{i+\frac{3}{2}}}$
Prediction:	σ_{k+1}^2	$=$	$\theta^2 + \rho^2 \hat{\sigma}_k^2$
	α_i^{k+1}	$=$	$\left(\frac{\rho\hat{\sigma}_k}{\sigma_{k+1}}\right)^{2i} \sum_{j\geq i} C_j^i (\frac{\theta}{\sigma_{k+1}})^{2(i-j)} \hat{\alpha}_i^k$ where $C_j^i = \frac{j!}{i!(j-i)!}$

By construction, we see that starting from a special SG distribution, with finite number of mixture parameters leads to a finite dimension filter.

Finally, we note that the conditional pdf of the observation Y_k given the signal X_k, is independent of the past observations and writes as:

$$g_k(x,y) = \frac{\lambda x^2}{|y|^3}\exp(-\frac{\lambda x^2}{y^2}).$$

It satisfies the conditions to construct and establish the quantization filter convergence rates (see Chapter 1).

2.5 Numerical experiments

2.5.1 Stationary suboptimal quantizers

Quantization based filters use precomputed quantizer grids and companion parameters. Although this preprocessing procedure is done off line, it is worth noting that in some cases, we can recycle optimal quantizers of standard distributions to construct stationary suboptimal quantizers that preserve the announced convergence rates. In fact, suppose there exists a sequence of affine invertible functions $T_k : \mathbb{R}^d \to \mathbb{R}^d$, such that $X_k = T_k(Z)$, for all $0 \leq k \leq n$, Z being a knwon distribution. Taking $\Gamma = \{z^1, \ldots, z^N\}$ an L^2-optimal N-quantizer of Z, we can define a stationary N-quantizer of X_k by setting:

$$\hat{X}_k = \sum_{i=1}^{N} T_k(z^i)\mathbf{1}_{\{T_k^{-1}(X_k)\in \mathbf{C_i}(\Gamma)\}} = T_k(\hat{Z}). \tag{2.5.1}$$

This quantizer is no more optimal (in fact it is not even a Voronoi quantizer), but, since the T_k are assumed to be affine, $\hat{X}_k$ still satisfies the stationary property. That is the property needed to establish first order scheme theoretical convergence rate to zero. Namely, we have :

$$\mathbb{E}[X_k|\hat{X}_k] = T_k(\mathbb{E}[Z|\hat{Z}]) = T_k(\hat{Z}) = \hat{X}_k.$$

Furthermore, even if in this case $\hat{X}_k$ is suboptimal, owing to the optimality of $\hat{Z}$, and to the Zador Theorem (see [4, 25]), we have $\|X_k - \hat{X}_k\|_2 = O(N^{\frac{-1}{d}})$. In fact we have:

$$\|X_k - \hat{X}_k\|_2 = \|T_k(Z) - T_k(\hat{Z})\|_2 \leq |||T_k|||\|Z - \hat{Z}\|_2.$$

This will be particularly interesting when we will consider for example linear signal dynamics with Gaussian innovations and Gaussian initial distribution. This is the case for Kalman filter or for the canonical stochastic volatility model (see (2.4.1) and (2.4.4)). It is also the case for particular explicit filter model, when the initial distribution is Gaussian (i.e. $SG(\sigma_0^2, (\alpha_i^0)_{i\geq 0})$ where $\alpha_0 = 1$ and $\alpha_i = 0$ for all $i > 0$). In these cases, X_k is Gaussian at all dates and we can drastically reduce the offline computation runtime needed to implement the quantization based filters since quantizer grids are obtained from the $\mathcal{N}(0, I_d)$

ones by an affine transformation.
Another important case of interesting preprocessing optimization is the case of stationary signal processes : the transition companion parameters do not depend on the time step considered, so only one transition is needed. Namey, $\mathbb{P}(\hat{X}_{k+1} = x^j | \hat{X}_k = x^i)$ is the same on the unique grid $\Gamma = \Gamma_0 = \{x^1, \ldots, x^N\}$ quantizing the stationnary distribution.

This is even more interesting, as it can be associated to the previous point. Assume for example that for any $0 \leq k \leq n$, we have $X_k \sim \mathcal{N}(0, \Sigma'\Sigma)$ where Σ satisfies $\Sigma'\Sigma = (\rho\Sigma)'\rho\Sigma + \theta'\theta$. In this particular case, we could first compute[1] Γ an L^2-optimal quantizer of the centered reduced Gaussian distribution. The quantizers Γ_k are then deduced by the expansion $\Gamma_k = \Sigma \times \Gamma_0$ coupled with (2.5.1), the companion parameters stored off line are those of one single transition as they are time independant.

2.5.2 Convergence tests

We select the following three test functions:

$$f_1(x) = x, \qquad f_2(x) = |x|^2, \qquad f_3(x) = \exp(-|x|).$$

For quantization based methods, as the schemes are deterministic, we simply need to study the behavior of the filter estimators (or dierctly the errors when a reference value is available), as the total size N of the grids goes to infinity.
For particle filtering methods, as the estimators are random variables, the testing approach is slightly different. In fact, we need to represent the empirical mean as well as a variance estimate as functions of the particle number N. For that purpose, we simulate a large number M of realizations of the filter random estimator, and represent the empirical confidence interval containing 90% of the observed values. This is achieved by computing the 5% and the 95% centiles over the population of the M filter random estimator realizations. These values will be represented as functions of the particle number N and the scheme performance is measured with respect to the confidence interval length.

2.5.3 Results and comments

2.5.3.1 Kalman filter: d=1

True values are computed by the Kalman-Bucy filter algorithm. For particle methods, we represent a realization of the filter estimator with 5000 particles. For quantization based filters, the two first order schemes are compared to the zero order one with $N_k = 100$ for all $0 \leq k \leq n$ (see Table 2.2).

[1] Optimal quantizers for the Gaussian distribution can be downloaded on `http://www.proba.jussieu.fr/pageperso/pages/`

$(\rho, \theta, \alpha, \Sigma_0)$	(0.65,1.0,0.1,0.05)			(0.996,0.0316,0.0632,0.7)		
	$\hat{\Pi}_{y,10} f_1$	$\hat{\Pi}_{y,10} f_2$	$\hat{\Pi}_{y,10} f_3$	$\hat{\Pi}_{y,25} f_1$	$\hat{\Pi}_{y,25} f_2$	$\hat{\Pi}_{y,25} f_3$
KF (Ref. Values)	0.0776126	0.9165	0.011546	1.01053	0.3643086	1.02273
SIS (5000 pts)	0.0605323	0.923421	0.0102159	1.08403	0.339045	1.17991
SIR (5000 pts)	0.0781505	0.915884	0.0115783	1.01208	0.363763	1.02594
QF0 (100 pts)	0.077596	0.916493	0.0115486	1.016	0.36235	1.03397
QF1 1-step (100 pts)	0.077601	0.916491	0.0115493	1.0121	0.363741	1.02593
QF1 2-step (100 pts)	0.0776013	0.916491	0.0115493	1.0121	0.363744	1.02591

Table 2.2: One dimensional Kalman filter case.

2.5.3.2 Kalman filter: d=3

We still consider equation (2.4.1) with parameters:

$$\rho = \begin{pmatrix} 0.996 & 0 & 0 \\ 0 & 0.996 & 0 \\ 0 & 0 & 0.996 \end{pmatrix}, \theta = \begin{pmatrix} 0.02 & 0.02 & 0.01 \\ 0.02 & 0.06 & -0.01 \\ 0.01 & -0.01 & 0.04 \end{pmatrix} \quad \text{and} \quad \alpha = I_3.$$

The initial signal distribution is centered and Gaussian with covariance matrix:

$$\Sigma_0' \Sigma_0 = \begin{pmatrix} 0.11 & 0.23 & 4e-4 \\ 0.23 & 0.53 & -8e-4 \\ 4e-4 & -8e-4 & 0.0018 \end{pmatrix}.$$

The chosen prior distribution is the stationary one. If we denote $\Gamma = \{z^1, \ldots, z^N\}$ the L^2-optimal N-quantizer of a centered reduced Gaussian distribution. At $0 \leq k \leq N$, $X_k \sim \mathcal{N}(0, \Sigma_0' \Sigma_0)$ and according to (2.5.1) we define the marginal stationary (N_k)-quantizer of (X_k) as follows:

$$\hat{X}_k = \sum_{i=1}^{N} \Sigma_0 z^i \mathbf{1}_{\{X_k \in \Sigma_0 \mathbf{C}_i(\Gamma)\}}.$$

Although the quantization filter convergence rate and numerical complexity depend on the signal dimension d, it remains interesting to apply them for medium dimension signals. The chosen quantizers are not optimal but we obtain satisfactory convergence results. Convergence errors are depicted in Figure 2.1 for f_2 and in Figure 2.2 for f_3. From log-log scale regression in Figure 2.1, we can evaluate the convergence rate improvement. The Table 2.3 summarizes the computed slopes of the log-log regressions.
We observe nearly the expected theoretical results. A convergence rate of $\frac{1}{d} \approx 0.33$ for zero order schemes, for first order schemes, the slopes are different from the theoretical one $\frac{2}{d} \approx 0.66$, but still better than the zero order slopes.

Or0	Or1 1-step	Or1 2-step
-0.34	-0.52	-0.81

Table 2.3: Regression slopes on the log-log scale representation (d=3)

Figure 2.3 depicts confidence intervals over 5000 particle filter realizations and the reference values computed exactly for f_2 and approximated via a Monte Carlo estimation for f_3 (see numerical integration in the introduction). While **SIS** degenrates, **SIR** method gives satisfying results. Compared with the quantization filters, it is important to see that the range of N is different from one method to another when the same error range is considered. The error reached by quantization filters for 800 points is farly less than the confidence interval length given by particle filters with $7000 >> \sqrt{800}$ particles.

2.5.3.3 Stochastic volatility model

In Figure 2.4 a comparison is made between particle filter methods over $M = 4000$ realizations and quantization based ones. We note the degeneracy of **SIS** method. **SIR** and quantization based methods converge to the same values. Furthermore, even for small quantizer sizes, we see that quantization based estimations always lie in the confidence interval of **SIR** computed for large particle sizes ($N = 10000$).

2.5.3.4 Non linear explicit filter

For numerical application, we considered the case where the a priori signal distribution is Gaussian. This is a particular serial Gaussian distribution $SG(\sigma^2, (\alpha_i)_{i\geq 0})$, where $\alpha_0 = 1$ and $\alpha_i = 0$ for all $i > 0$. As it has been precised in paragraph 2.4.3, the model allows to construct a semi closed solution to the filtering problem: it is of *Serial Gaussian* distribution with recursively determined parameters (see Algorithm 5). Hence, reference values for the considered test functions can be computed via Monte Carlo simulations from the filter distribution.
Results in Figure 2.5 have been obtained for the set of parameters $(\rho, \theta, \lambda) = (0.5, 1, 0.1)$ and $n = 10$. We choose the stationary distribution $\mathcal{N}(0, \frac{\theta^2}{1-\rho^2})$ as the initial signal one. In Figure 2.6, are depicted resluts for $(\rho, \theta, \lambda) = (0.65, 1, 0.1)$ and $n = 10$. Here, the initial distribution has been fixed to $\mathcal{N}(0, \sigma_0^2)$ with $\sigma_0 = 0.05$.
On one hand the first line depicts the quantization based filter behaviour with respect to the total quantizer size. On the other hand, the second and third line represent respectively the **SIS** and the **SIR** confidence intervals with the empirical mean value and centiles as functions of the particle number. Confidence intervals for particle filters are estimated over $M = 1000$ realizations.

As for the previous example, we observe the degeneracy of **SIS** filter, the convergence

value is far from the reference value. Quantization based filter and **SIR** filter converge to the reference value. It is worth noting that the grid size order is quite different whether the method is a quantization based one or a particle filter one. Once again, in this example, particle methods need much more points than quantization ones. This compensates by far the higher complexity order of quantization approach.

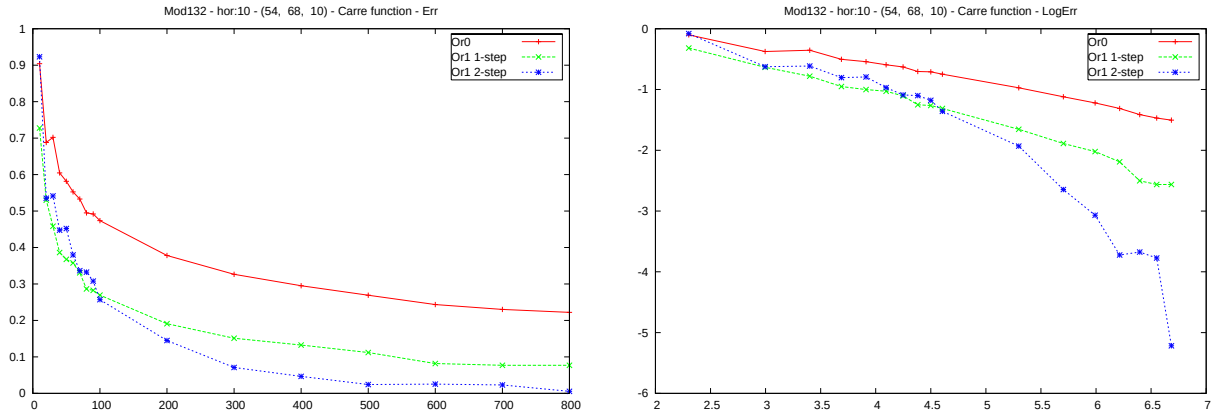

Figure 2.1: Quantization filter estimator errors for 3-dimensional Kalman case as a function of the quantizer size N_k (top: $\|\Pi_{10} f_2 - \hat{\Pi}_{10} f_2\|_2$, bottom: log-log scale representation).

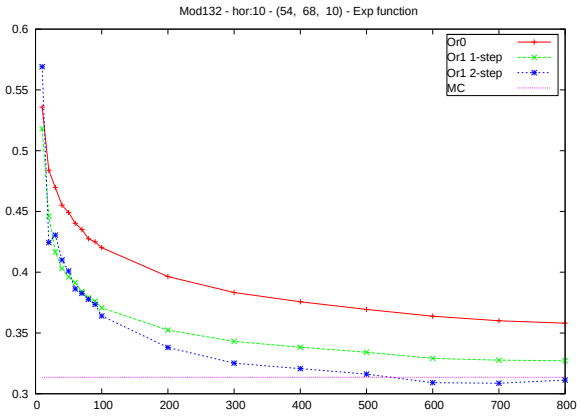

Figure 2.2: Quantization filter estimator errors for 3-dimensional Kalman case as a function of the quantizer size N_k : $\|\Pi_{10} f_3 - \hat{\Pi}_{10} f_3\|_2$.

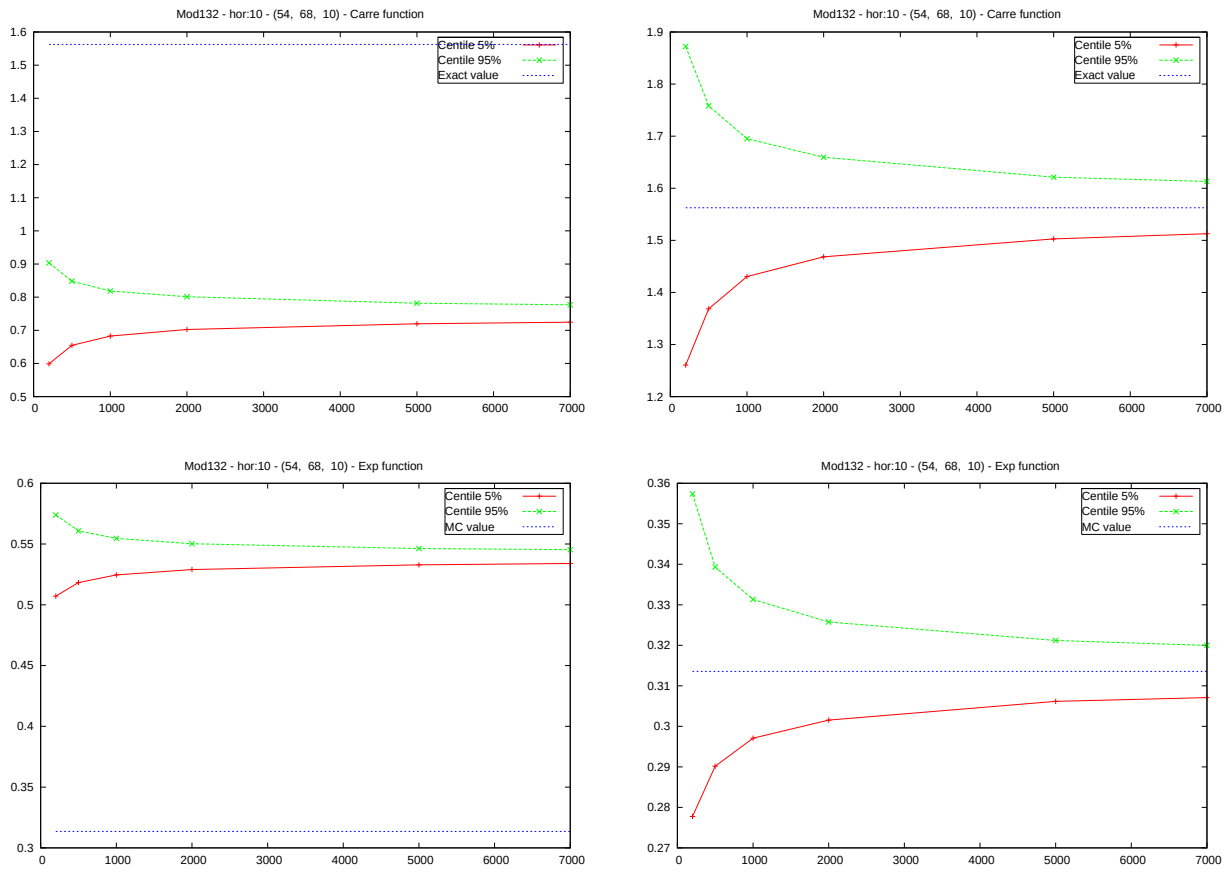

Figure 2.3: Particle filter confidence intervals for 3-dimensional Kalman case as a function of the partcile number N.

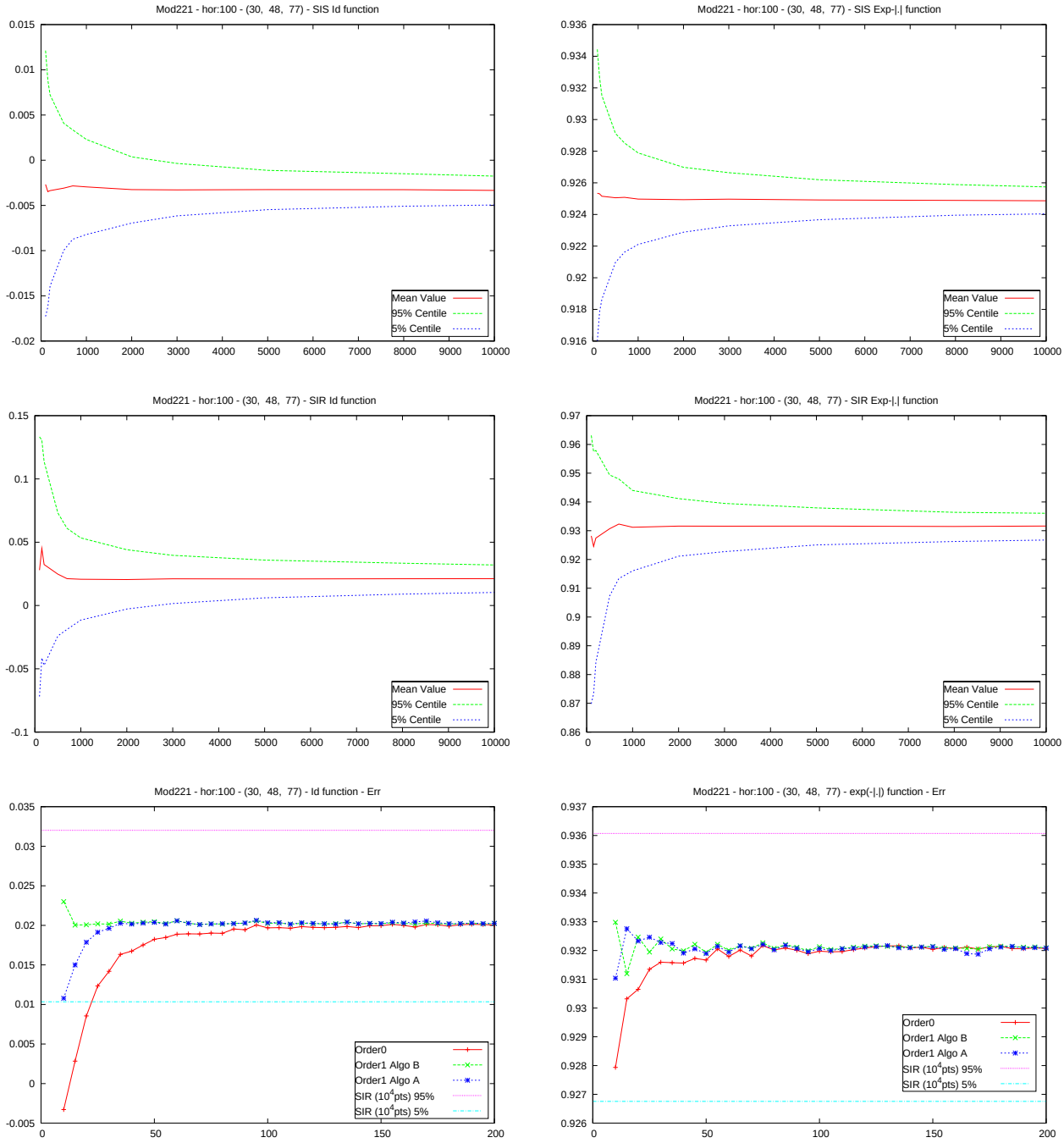

Figure 2.4: Particle filter and quantization filter approximations for SVM (left: $\Pi_{y,100} f_1$, right: $\Pi_{y,100} f_3$) - $(\beta, \sigma) = (0.995, 0.01)$.

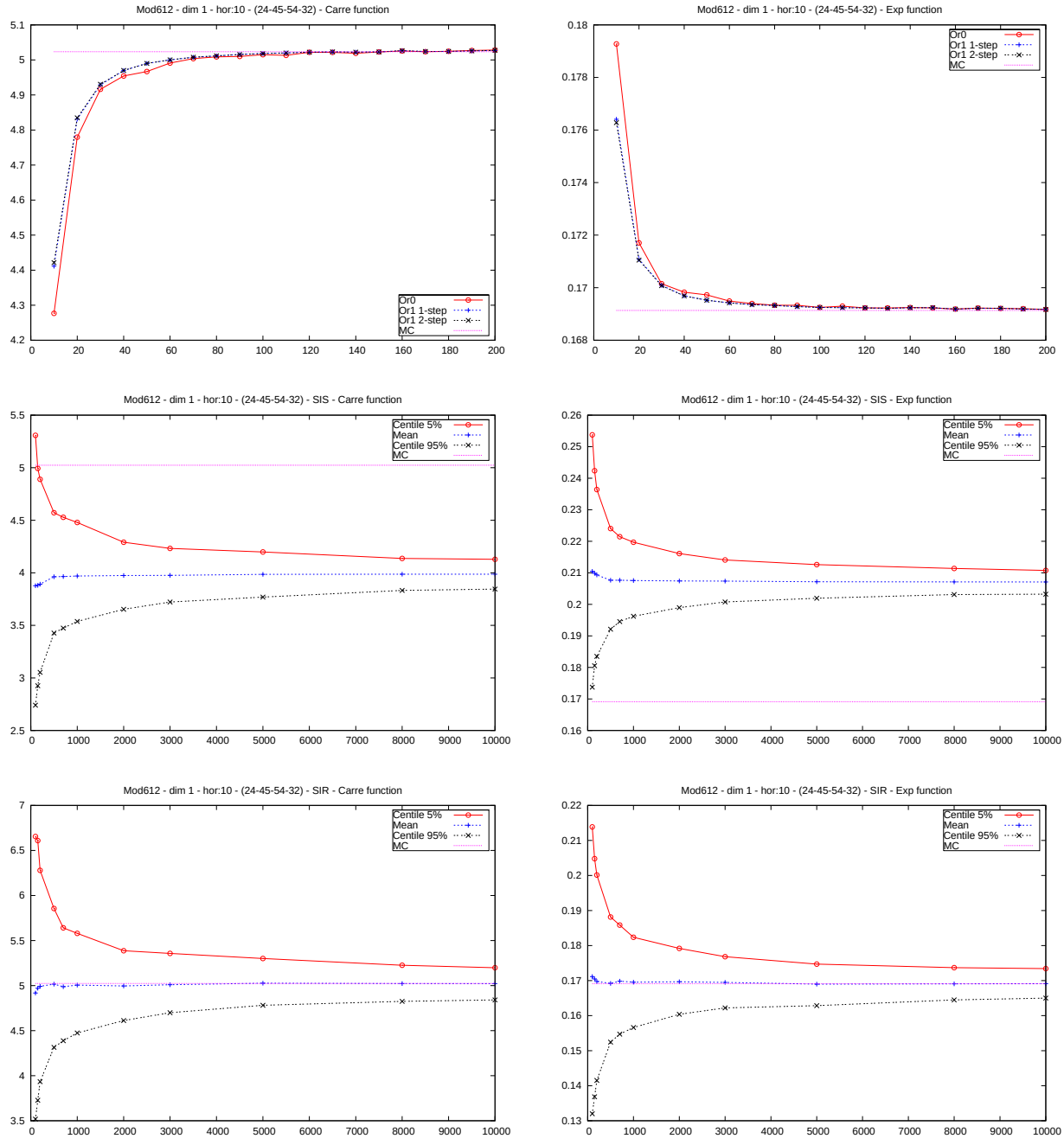

Figure 2.5: Explicit filter estimators as function of grid sizes (left: $\hat{\Pi}_{10} f_2$, right: $\hat{\Pi}_{10} f_3$) - $(\rho, \theta, \lambda, n) = (0.5, 1, 0.1, 10)$.

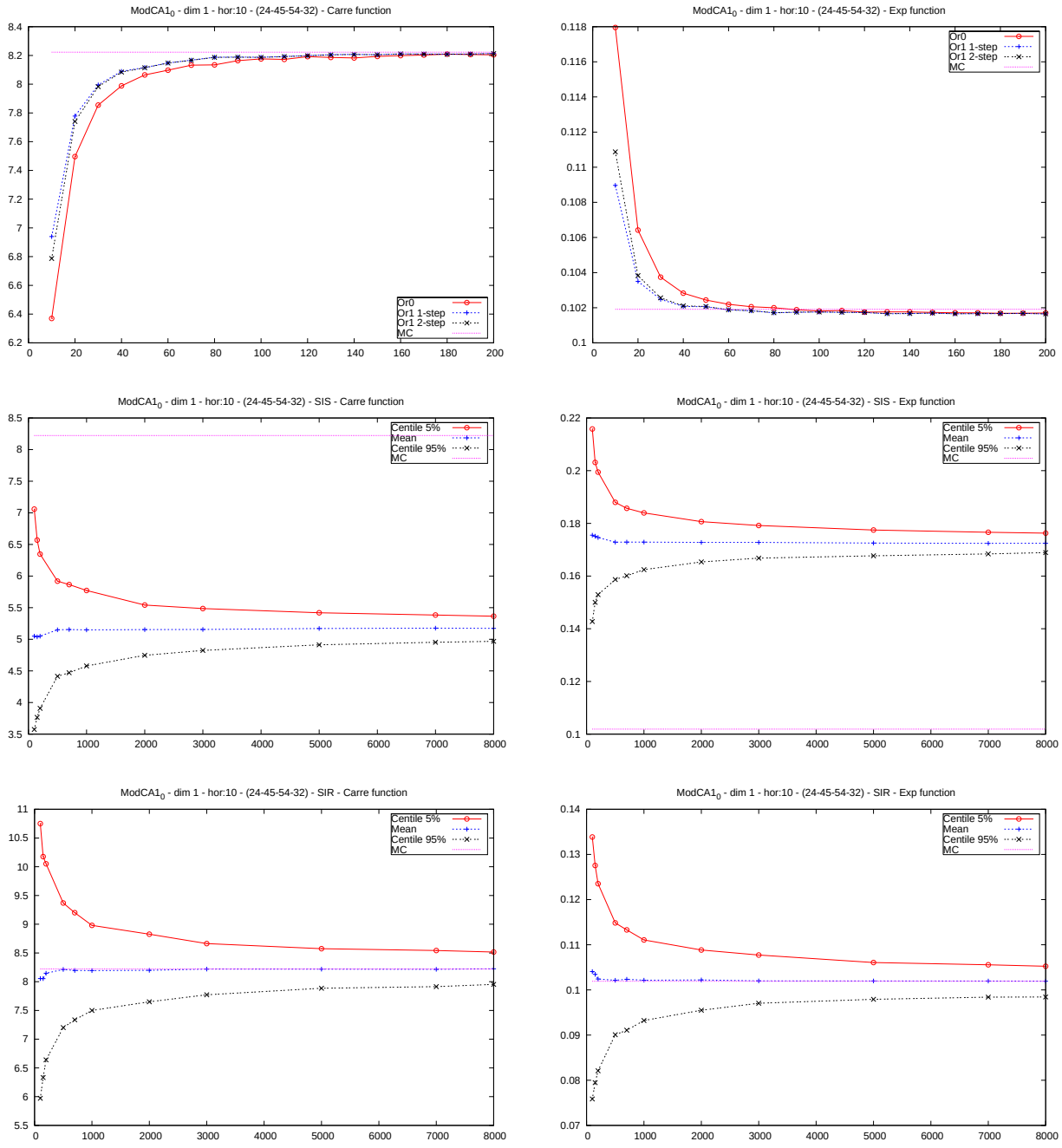

Figure 2.6: Explicit filter estimators as function of grid sizes (left: $\hat{\Pi}_{10} f_2$, right: $\hat{\Pi}_{10} f_3$) - $(\rho, \theta, \lambda, \sigma_0, n) = (0.65, 1, 0.1, 0.05, 10)$.

Deuxième partie

Prétraitements par quantification et application en filtrage

Chapter 3

Observation preprocessing for numerical filtering approximation methods

Preprocessing procedures are presented in order to establish fast and effecient online algorithms to compute discrete time observation filters over a finite horizon. The case of discrete state models is first studied through the use of observation quantization grids, to compute offline the likelihood functions and store the corresponding values in accessible look up tables. Signal quantization based filters are then introduced to reduce the continuous state model to the discrete one and then make possible the use of observation quantization preprocessing procedures. In both cases, the L^1-errors on the random filter estimates are controlled and convergence rates are given when the quantizer size grows to infinity using optimal quantization results. Theoretical convergence rate is confirmed by numerical experiments.

Key words: Numerical filtering, quantization, offline preprocessing, state models, random filter.

3.1 Introduction

Solving a nonlinear filtering problem is still problematic from a practionner point of view. Even if the theory is mature, we are rarely able to give a closed formula to the filter solution. The problem is to determine the measure-valued solution of a discrete filtering problem : the *a posteriori* distribution of a hidden system state x, evolving through some known dynamics and from which noisy observations y are made. In other words, we are facing the problem of determining a Borel function Ξ of the observations, supposed to describe the behaviour of the system according to the information available from these noisy observations.
An explicit solution of this problem can be given in a few cases like discrete signal models, finite dimensional filtering problems like the Kalman-Bucy filters or even some specific infinite dimensional filtering problems. As concerns this last setting, let us mention for example the model introduced in [18] and developed furthermore in [12], where the filter depends upon countably many parameters (most of them being close to zero in usual practical cases). All these examples lead to closed or semi-closed form for Ξ. In a more general context, we are only able to design some numerical approximations, via recursive schemes, which can be rather computationally demanding in some cases (see [34]). A central commun point to the majority of these methods, either for exact solutions or numerical approximations, is the use of sequential prediction-update transitions, where generally we are brought to consider the evaluation of likelihood functions depending on the observations.

For filtering applications, it is important to reduce the online computation runtime, as the filter value needs to be estimated before we get the next observation. One way to reduce the computation complexity and runtime is to compute offline a set of tools that can be used in the online algorithm. In that sense, the idea of precomputing the likelihood function on a predefined grid was introduced in [37] for filtering discrete finite state signal models. For this particular case, this amounts to the following approximation scheme, where the observation y is approximated by its quantization $\hat{y} = q(y)$ and $q : \mathbb{R}^{d'} \to \Gamma$ ($\Gamma \subset \mathbb{R}^{d'}$, finite) :

$$\Xi(y) \overset{\text{Obs. Quantization}}{\longrightarrow} \Xi(\hat{y}). \tag{3.1.1}$$

Quantization is a procedure for approaching a continuous random variable by a discrete one in a somewhat optimal way (see Intorduction chapter and [25]). For (3.1.1), we use as a quantized observation $\hat{y}$ the projection of the observation y on a predefined quantization grid, for which all likelihood values would have been precomputed and stored. As the likelihood function depends on both signal and observation values, the procedure is made possible by the discrete nature of the state space. For general continous state models, the use of quantization based filters introduced in [41, 43] allows to make one further step in the preprocessing and approximation procedure. We first define some signal quantization

based filter estimate and then evaluate it at the observation quantization $\hat{y}$.

$$\Xi(y) \overset{\text{Signal Quantization}}{\longrightarrow} \hat{\Xi}(y) \overset{\text{Obs. Quantization}}{\longrightarrow} \hat{\Xi}(\hat{y}). \tag{3.1.2}$$

For (3.1.2), the optimal quantization of the signal allows to come back to a discrete space case, which according to [41] allows in turn to define both numerically computable and converging filter estimates. The offline computation principle of the quantization based filtering method makes possible the precomputing of likelihood values on a product quantization grid of the signal-observation process. By product quantization,we mean a product grid $\Gamma^X \times \Gamma^Y$ where Γ^X and Γ^Y are some quantization grids of respectively X and Y.

The paper is organised as follows : the next section is devoted to the particular case of optimal discrete signal filtering, in which we examine the effect of adding an observation quantization procedure according to scheme (3.1.1). A numerical application is carried out on a three state stochastic volatility model. In Section 3.3, the same idea is associated to quantization based filtering methods developed in [41] in order to enlarge observation quantization filtering to continuous state space models. Here, we will also perform some numerical tests on the classical case of Kalman filter.

3.2 Optimal filtering : discrete signals

3.2.1 Sequential schemes

In this section we treat the particular case of a discrete signal process with a finite observation horizon n. Assume the signal is an $\mathbb{R}^d$-valued discrete Markov chain X_k of known transition matrix $\mathbf{P}_k$ such that, at each time step k, X_k is a random vector taking N_k states $x_k^1, \ldots, x_k^{N_k}$ in $\mathbb{R}^d$. The initial distribution μ_0 of X_0 is supposed to be known a priori. We denote, for $1 \leq i \leq N_0$, $\mu_0^i := \mathbb{P}[X_0 = x_0^i]$.

At each time step k, an $\mathbb{R}^{d'}$-valued noisy observation Y_k of X_k is made. The dynamics of process $(Y_k)_{0 \leq k \leq n}$ are driven by a Borel function $G_k : \mathbb{R}^d \times \mathbb{R}^{d'} \times \mathbb{R}^d \times \mathbb{R}^{q'} \to \mathbb{R}^{d'}$ such that:

$$Y_0 = 0 \quad \text{and} \quad Y_k = G_k(X_{k-1}, Y_{k-1}, X_k, \eta_k), \quad 1 \leq k \leq n, \tag{3.2.1}$$

where (η_k) is a sequence of iid $\mathbb{R}^{q'}$-valued random variables, independent of $\sigma(X_k, k \geq 0)$. We assume for convenience that for every $1 \leq k \leq n$, the distribution of Y_k given (X_{k-1}, Y_{k-1}, X_k) admits a continuous conditional pdf $y \mapsto g_k(X_{k-1}, Y_{k-1}, X_k, y)$.

The problem we aim to solve is to compute the measure-valued random variable $\Pi_{Y,n}$ which represents a regular version of the conditional distribution of X_n given the whole observation process $Y = (Y_1, \ldots, Y_n)$. Namely, using a fixed set of regular test functions f, we want to compute an approximation of:

$$\Pi_{Y,n} f = \mathbb{E}[f(X_n)|Y_1, \ldots, Y_n].$$

More precisely this means for each test function $f : \mathbb{R}^d \to \mathbb{R}$, we can associate a Borel function:

$$\begin{array}{rcl} \Xi_n : (\mathbb{R}^{d'})^n & \to & \mathbb{R} \\ (y_1, \ldots, y_n) & \mapsto & \Pi_{y,n} f = \mathbb{E}[f(X_n) | Y_1 = y_1, \ldots, Y_n = y_n]. \end{array}$$

As the signal is discrete, this function can be evaluated sequentially as follows. First, the Kallianpur-Striebel formula yields:

$$\Xi_n(y_1, \ldots, y_n) = \Pi_{y,n} f = \frac{\pi_{y,n} f}{\pi_{y,n} \mathbf{1}}, \tag{3.2.2}$$

where $\pi_{y,n} f = \mathbb{E}[f(X_n) \prod_{k=1}^n g_k(X_{k-1}, y_{k-1}, X_k, y_k)]$ is the so-called unnormalized filter. Owing to the Markov property of the signal, the unnormalized filter $\pi_{y,n}$ satisfies the following sequential formula:

$$\begin{array}{rcl} H_{y,k} f(x_{k-1}) & = & \mathbb{E}[f(X_k) g_k(x_{k-1}, y_{k-1}, X_k, y_k) | X_{k-1} = x_{k-1}], \\ \pi_{y,0} f & = & \mathbb{E}[f(X_0)], \\ \pi_{y,k} f & = & \pi_{y,k-1} H_{y,k} f. \end{array} \tag{3.2.3}$$

So, considering that $\pi_{y,k} \in \mathcal{M}_{1,N_k}(\mathbb{R})$ and that $H_{y,k} \in \mathcal{M}_{N_{k-1},N_k}(\mathbb{R})$, (3.2.3) may be written in a matricial form as follows :

$$\begin{array}{rcl} H^{ij}_{y,k} & = & \mathbf{P}^{ij}_{k-1} g_k(x^i_{k-1}, y_{k-1}, x^j_k, y_k), \quad 0 < k \leq n, \\ \pi_{y,0} & = & \mu_0, \\ \pi_{y,k} & = & \pi_{y,k-1} H_{y,k}. \end{array} \tag{3.2.4}$$

Finally, a non linear normalization step gives:

$$\Pi^i_{y,n} = \frac{\pi^i_{y,n}}{\sum_{j=1}^{N_n} \pi^j_{y,n}} \qquad \text{and} \qquad \Xi_n(y) = \Pi_{y,n} f = \sum_{i=1}^{N_n} \Pi^i_{y,n} \times f(x^i_n).$$

The inductive formulation of the filter computation has the advantage of needing only the last two observations and the last intermediate filter value, to compute the transition in (3.2.4). Having in mind the quantization of the observations, we aim to reduce the computation procedure, by evaluating offline the likelihood $g_k(x^i_{k-1}, y_{k-1}, x^j_k, y_k)$. Unfortunately, this is not possible on real observations which of course come out online. The idea of quantizing the observation process is to construct an approximation $\hat{Y}_k$ for Y_k taking a finite number N^Y_k of states in $\Gamma^Y_k = \{y^1_k, \ldots, y^{N^Y_k}_k\}$ and such that for $1 \leq k \leq n$, there exists a Borel function $q_k : \mathbb{R}^{d'} \to \mathbb{R}^{d'}$ such that $\hat{Y}_k = q_k(Y_k)$.

Once this approximation is processed, offline, we can compute and then keep in some accessible "look up tables" the likelihood values $g_k(x^i_{k-1}, y^l_{k-1}, x^j_k, y^m_k)$. Then we approximate $\Xi_n(y)$ by $\Xi_n(\hat{y})$, where $\hat{y} = (q_1(y_1), \ldots, q_n(y_n))$. This consists in plugging the quantized likelihood $g_k(x^i_{k-1}, q_{k-1}(y_{k-1}), x^j_k, q_k(y_k))$ instead of $g_k(x^i_{k-1}, y_{k-1}, x^j_k, y_k)$ in (3.2.4)

Algorithm 6 Discrete state filtering algorithm with preprocessed observation

$\pi_{\hat{y},0} = \mu_0$

Transition from k to $k+1$: $\pi_{\hat{y},k}$ is defined and $\hat{y}_k = y_k^{l_0}$.

- New observation y_{k+1},
- Projection : $\hat{y}_{k+1} = \sum_{m=1}^{N_{k+1}^Y} y_{k+1}^m \mathbf{1}_{A_{k+1}^m}(y_{k+1})$,
- Define m_0 the indice of y_{k+1} nearest neighbour on the grid Γ_{k+1}^Y.
- $H_{\hat{y},k+1}^{ij} = \mathbf{P}_k^{ij} g_{k+1}(x_k^i, y_k^{l_0}, x_{k+1}^j, y_{k+1}^{m_0})$.
- $\pi_{\hat{y},k+1} = \pi_{\hat{y},k} H_{\hat{y},k+1}$.

to spare the online computations of the true one. The quantized likelihood are "read" in a table and need not to be computed. This leads to Algorithm 6.

Finally, we define the estimator $\Pi_{\hat{Y},n}$ of the random filter $\Pi_{Y,n}$ by the recursive scheme (3.2.4) involving the quantized observations. We have consequently:

$$\Pi_{\hat{Y},n} f \stackrel{Def}{=} (\Pi_{y,n})_{y=\hat{Y}} f = \Xi_n(\hat{Y}).$$

In the forthcoming subsection, we establish some error bounds for the L^1-error induced by such a preprocessing procedure. This will emphasize the interest of using optimal (quadratic) N_k^Y-quantizations [1] of the observations (Y_k). We have in fact a convergence rate result, induced by the so-called Zador theorem, namely:

$$\|Y_k - \hat{Y}_k\|_2 = O((N_k^Y)^{-\frac{1}{d'}}).$$

3.2.2 Stability with respect to observation imprecision

In view of studying the estimation L^1-error $\|\Pi_{Y,n} f - \Pi_{\hat{Y},n}\|_1$, we examine some stability results when some error is introduced in the observations used in the filtering scheme. We consider two different observation n-tuples: $y = y_1, \ldots, y_n$ and $y' = y'_1, \ldots, y'_n$. Let ϕ_n denote the joint pdf of $Y_{1:n}$ defined by:

$$\phi_n(y) = \mathbb{E}[\prod_{k=1}^{n} g_k(X_{k-1}, y_{k-1}, X_k, y_k)].$$

Lemma 3.2.1 *For every $k \in \{0, \ldots, n\}$, assume g_k satisfies the following assumption:*

H 1 *The function g_k is bounded and we denote $L_g := \max_{1 \le k \le n} \|g_k\|_\infty$. Moreover, there exists a constant $[g_k]_{Lip}^1 > 0$ such that for any u, u', $\bar{u}$ and $\bar{u}'$ in $\mathbb{R}^{d'}$:*

$$|\mathbb{E}[g_k(X_{k-1}, u, X_k, u')] - \mathbb{E}[g_k(X_{k-1}, \bar{u}, X_k, \bar{u}')]| \le [g_k]_{Lip}^1 \left(|u - u'| + |\bar{u} - \bar{u}'|\right).$$

[1] meaning a quantizer of size N_k^Y minimizing the L^2-error $\|Y_k - \hat{Y}_k\|_2$ among all N_k^Y-quantizers.

Hence, for any y, $y' \in (\mathbb{R}^{d'})^n$

$$|\phi_n(y) - \phi_n(y')| \leq L_g^{n-1} \sum_{k=1}^{n} \left([g_k]^1_{Lip} + [g_{k+1}]^1_{Lip} \right) |y_k - y'_k|,$$

with the notation $[g_{n+1}]^1_{Lip} = 0$.

Proof. We first recall $y_0 = 0$.

$$\begin{aligned}
|\phi_n(y) \quad - \quad \phi_n(y')| &\leq |\mathbb{E}[\prod_{k=1}^{n} g_k(X_{k-1}, y_{k-1}, X_k, y_k) - \prod_{k=1}^{n} g_k(X_{k-1}, y'_{k-1}, X_k, y'_k)]| \\
&\leq |\mathbb{E}[\prod_{k=1}^{n} g_k(X_{k-1}, y_{k-1}, X_k, y_k) - \prod_{k=1}^{n-1} g_k(X_{k-1}, y'_{k-1}, X_k, y'_k) g_n(X_{n-1}, y_{n-1}, X_n, y_n)]| \\
&\quad + |\mathbb{E}[\prod_{k=1}^{n-1} g_k(X_{k-1}, y'_{k-1}, X_k, y'_k) g_n(X_{n-1}, y_{n-1}, X_n, y_n) - \prod_{k=1}^{n} g_k(X_{k-1}, y'_{k-1}, X_k, y'_k)]|, \\
&\leq L_g |\phi_{n-1}(y) - \phi_{n-1}(y')| + L_g^{n-1} [g_n]^1_{Lip} \left(|y_{n-1} - y'_{n-1}| + |y_n - y'_n| \right), \\
&\leq L_g^{n-1} \sum_{k=1}^{n-1} \left([g_{k+1}]^1_{Lip} + [g_k]^1_{Lip} \right) |y_k - y'_k| + L_g^{n-1} [g_n]^1_{Lip} |y_n - y'_n|.
\end{aligned}$$

□

The error involved by the quantization of the observation is controlled by the following theorem:

Theorem 3.2.1 *Assume* ***H1*** *and in addition:*

H 2 $\| \frac{1}{\phi_n(Y)} \|_2 < +\infty$.

Then, for every $f \in \mathcal{C}_b$*:*

$$\|\Pi_{Y,n} f - \Pi_{\hat{Y},n} f\|_1 \leq 2 \|f\|_\infty L_g^{n-1} \| \frac{1}{\phi_n(Y)} \|_2 \sum_{k=1}^{n} \left([g_{k+1}]^1_{Lip} + [g_k]^1_{Lip} \right) \|Y_k - \hat{Y}_k\|_2.$$

Remark 3.2.1 **H**2 is generally satisfied by models with finite discrete states. This kind of hypothesis is closely connected to some mixing assumptions needed to establish uniform convergence of particle filters [23, 14, 51]. For the present, we only study the convergence problem of random filter estimates at a fixed horizon. It would be important in a further step to examine the statbility of the filter as the horizon n grows.

Proof.
Let y, y' in $\mathbb{R}^{d'n}$. Using Lemma 3.2.1 we have:

$$\begin{aligned}
|\pi_{y,n} f - \pi_{y',n} f| &\leq \|f\|_\infty |\phi_n(y) - \phi_n(y')|, \\
&\leq \|f\|_\infty L_g^{n-1} \left(\sum_{k=1}^{n-1} \left([g_{k+1}]^1_{Lip} + [g_k]^1_{Lip} \right) |y_k - y'_k| + [g_n]^1_{Lip} |y_n - y'_n| \right).
\end{aligned}$$

According to Lemma 3.1. in [41], we have consequently:

$$|\Pi_{y,n}f - \Pi_{y',n}f| \leq \frac{2L_g^{n-1}\|f\|_\infty}{\phi_n(y) \vee \phi_n(y')} \times \left(\sum_{k=1}^{n-1} \left([g_{k+1}]_{Lip}^1 + [g_k]_{Lip}^1 \right) |y_k - y'_k| + [g_n]_{Lip}^1 |y_n - y'_n| \right). \quad (3.2.5)$$

Now taking the L^1-norm of the previous inequality and using **H**2 yields:

$$\|\Pi_{Y,n}f - \Pi_{\hat{Y},n}f\|_1 \leq 2\|\frac{1}{\phi_n(Y)}\|_2 L_g^{n-1}\|f\|_\infty \left(\sum_{k=1}^{n} \left([g_{k+1}]_{Lip}^1 + [g_k]_{Lip}^1 \right) \|Y_k - \hat{Y}_k\|_2 \right).$$

with the convention$[g_{n+1}]_{Lip}^1 = 0$. □

Theorem 3.2.1, when $\hat{Y}_k$ is an L^2-optimal N_k^Y-quantization of Y_k, ensures the convergence to zero of the L^1-estimation error, when $N_k^Y \to +\infty$ with the rate $(N_k^Y)^{-\frac{1}{d'}}$.

Remark 3.2.2 Note that the result of Theorem 3.2.1 is still available for a general state space model, in particular a continuous state signal.

3.2.3 Example

One application of filtering in finance is the calibration of volatility in stochastic volatility models. We consider here an observable risky asset price (S_k) with dynamics given by :

$$S_{k+1} = S_k \exp\left[\left(r - \frac{1}{2}X_k^2 \right) \delta + X_k \sqrt{\delta} \eta_{k+1} \right], \quad S_0 = s_0 > 0,$$

where (η_k) is a sequence of Gaussian white noise, (X_k) is the unobservable volatility process and $\delta > 0$ represents a discretization time step. Equivalently, we observe the process $(Y_k) = (\log S_k)$, and we notice that the conditional distribution of Y_{k+1} given (X_k, Y_k) has the following density :

$$g_{k+1}(X_k, Y_k, y') = \frac{1}{\sqrt{2\pi X_k^2 \delta}} \exp\left[-\frac{\left(y' - Y_k - (r - \frac{1}{2}X_k^2)\delta \right)^2}{2X_k^2 \delta} \right], \quad y' \in \mathbb{R}.$$

We model here the dynamics of (X, S) under a risk neutral martingale measure $\mathbb{P}$, r representing in this case the riskless interest rate.

We assume that (X_k) is an homogeneous Markov chain taking three possible values $x^b < x^m < x^h$ in $\mathbb{R}_+ \backslash \{0\}$. Its probability transition matrix is given by :

$$\mathbf{P}_k = \begin{pmatrix} 1 - (p_{bm} + p_{bh})\delta & p_{bm}\delta & p_{bh}\delta \\ p_{mb}\delta & 1 - (p_{mb} + p_{mh})\delta & p_{mh}\delta \\ p_{hb}\delta & p_{hm}\delta & 1 - (p_{hb} + p_{hm})\delta \end{pmatrix}. \quad (3.2.6)$$

The considered volatility model is a Markov-chain approximation "à la Kushner" (see [31]) of an Ornstein-Uhlenbeck process :

$$dX_t \;=\; \lambda(x_0 - X_t)dt + \sigma dW_t.$$

Denoting by $\Delta > 0$ the spatial step, this corresponds to a probability transition matrix of the form (3.2.6) with :

$$x^b \;=\; x_0 - \Delta, \quad x^m \;=\; x_0, \quad x^h \;=\; x_0 + \Delta,$$

and

$$\begin{array}{rclcrcl} p_{bm} & = & \lambda + \frac{\sigma^2}{2\Delta^2} & , & p_{bh} & = & 0 \\ p_{mb} & = & \frac{\sigma^2}{2\Delta^2} & , & p_{mh} & = & \frac{\sigma^2}{2\Delta^2} \\ p_{hb} & = & 0 & , & p_{hm} & = & \lambda + \frac{\sigma^2}{2\Delta^2}. \end{array} \tag{3.2.7}$$

Remark 3.2.3 Here the factor $\|\frac{1}{\phi_n(Y)}\|_2$ depends on x^h and x^b, and on the observation horizon n. The case where n grows deserves to be examined numerically, as it would be expected that the error control deteriorates.

In order to ensure that $\mathbf{P}_k$ is indeed a probability transition matrix, we have the consistency conditions:

$$1 - \left(\lambda + \frac{\sigma^2}{2\Delta^2}\right)\delta \geq 0 \quad \text{and} \quad 1 - \frac{\sigma^2}{\Delta^2}\,\delta \geq 0.$$

Numerical tests are performed with:

- Price parameters : $r = 0.05$, $S_0 = 110$,
- Volatility parameters : $\lambda = 1$, $\sigma = 0,1$, $X_0 = 0.15$,
- Spatial step : $\Delta = 0,05$.
- Quantization : Observation quantization grids all have the same size $\bar{N}^Y$ for each time step $t_k = k\delta$ with $\delta = \frac{1}{n}$. We perform optimal quantization using a stochastic gradient descent algorithm (see [4, 44]).

The L^1-error is represented in Figure 3.1 and Figure 3.2 for two different observation horizons $n = 5$ and $n = 20$. The convergence rate to zero is depicted through the regression on the log-scale representations. The slope is close of $\frac{1}{d'} = 1$ which is coherent with theoretical results.

3.3 Continuous state signals

We suppose now that the signal process (X_k) is an $\mathbb{R}^d$-valued discrete time hidden Markov chain $(X_k)_{0\leq k\leq n}$, which evolves according to a signal equation of the form:

$$X_0 \sim \mu_0 \quad \text{and} \quad X_{k+1} = F_{k+1}(X_k, \varepsilon_{k+1}), \quad 0 \leq k \leq n-1, \tag{3.3.1}$$

where $(\varepsilon_k)_{1<k\leq n}$ is a sequence of iid $\mathbb{R}^q$-valued random variables, independent of X_0, and $F_k : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}^d$ are some Borel function. The probability transition from X_k to X_{k+1} will be denoted $\mathbf{P}_k(x, dx')$.
The distribution μ_0 of X_0 is supposed to be known. We will denote:

$$\mu_0 f = \int f(x)\mu_0(dx) \quad \textit{and} \quad \mathbf{P}_k f(x) = \int f(x')\mathbf{P}_k(x, dx').$$

In this new framework, the Kallianpur-Striebel formula (3.2.2) and the recursive construction of the filter (3.2.3) are still true. The difference is that the $H_{y,k}$ operators involve now some "continuous state" conditional expectations whose evaluation in most general cases can only be achieved by numerical approximations. The method we use here is that developed by Pagès and Pham in [41] : each X_k is approximated by an L^2-optimal N_k^X-quantization $\hat{X}_k$ in order to reduce the problem to the discrete state case (note however that $(\hat{X}_k)$ is no longer a Markov chain). Namely, we define

$$\hat{X}_k = \sum_{i=1}^{N_k^X} x_k^i \mathbf{1}_{\mathbf{C_i}(\Gamma_k^X)}(X_k),$$

where $\mathbf{C_i}(\Gamma_k^X)$ is the so called Voronoi tessellation of the grid Γ_k^X. Then one sets recursively:

$$\begin{array}{rcll} \hat{H}_{y,k}^{ij} & = & \hat{\mathbf{P}}_{k-1}^{ij} g_k(x_{k-1}^i, y_{k-1}, x_k^j, y_k), & 0 < k \leq n, \\ \hat{\pi}_{y,0}^i & = & \hat{\mu}_0^i = \mathbb{P}[X_0 \in \mathbf{C_i}(\Gamma_0^X)], & \\ \hat{\pi}_{y,k} & = & \hat{\pi}_{y,k-1}\hat{H}_{y,k}. & \end{array} \tag{3.3.2}$$

Then, we define $\hat{\Xi}_n(y)$ associated to some test function $f \in \mathcal{C}_{b,Lip}$, and a fixed observation realization y as the normalized filter defined by:

$$\hat{\Pi}_{y,n}^i = \frac{\hat{\pi}_{y,k}^i}{\sum_{j=1}^{N_n} \hat{\pi}_{y,k}^j} \qquad \text{and} \qquad \hat{\Xi}_n(y) = \hat{\Pi}_{y,n} f = \sum_{i=1}^{N_n} \hat{\Pi}_{y,n}^i \times f(x_n^i).$$

Using results from [41], we derive the following theorem which establishes a convergence result for the filter estimation, depending on the quantization error estimates:

Theorem 3.3.1 *For $k = 1, \ldots, n$, assume:*

H 3 *The signal Markov operator $\mathbf{P}_k$ is K-Lipschitz, namely for any Lipschitz function f, $\mathbf{P}_k f$ is $(K[f]_{Lip})$-Lipschitz.*

H 4 *(i) The function g_k is bounded on $\mathbb{R}^d \times \mathbb{R}^{d'} \times \mathbb{R}^d \times \mathbb{R}^{d'}$, we set $L_g = \max_{1\leq k\leq n} \|g_k\|_\infty$.*

(ii) There exists a real constant $[g_k]_{Lip}^2$ such that for any u, u' in $\mathbb{R}^{d'}$, $g_k(., u, ., u')$ is Lipschitz on $\mathbb{R}^d \times \mathbb{R}^d$. For x, x' and $\bar{x}, \bar{x}'$ in $\mathbb{R}^d$:

$$|g_k(x, u, x', u') - g_k(\bar{x}, u, \bar{x}', u')| \leq [g_k]_{Lip}^2 \left(|x - \bar{x}| + |x' - \bar{x}'|\right).$$

Let $y = (y_1, \ldots, y_n)$ be a fixed observation n-tuple. Then, for every test function $f \in \mathcal{C}_{b,Lip}$ there exists a sequence of nonnegative real constants $(B_k^n(f))_{0\leq k\leq n}$ not depending on the observation sequence such that:

$$|\Pi_{y,n} f - \hat{\Pi}_{y,n} f| \leq \frac{L_g^n}{\phi_n(y) \vee \hat{\phi}_n(y)} \sum_{k=0}^{n} B_k^n(f) \|X_k - \hat{X}_k\|_2,$$

where $\hat{\phi}_n(y) = \hat{\pi}_{y,n}\mathbf{1}$
and $B_k^n(f) \leq K^{n-k}[f]_{Lip} + 2\frac{\|f\|_\infty}{L_g}\left(([g_{k+1}]^2 + [g_k]^2) + (1+K)\sum_{j=k+1}^{n} K^{j-k-1}[g_j]^2\right)$.

The signal quantization allows to come back to the recursive computations of the discrete state case. For a given observation y, we have here also to evaluate g_k on the signal quantization grid points Γ_k^X and the collected observations. This can be avoided (at least in the online procedure) by combining this with an observation preprocessing procedure.
Using a quantized version of the observation process $\hat{Y}_k$, the likelihood function g_k can be pre-evaluated offline on a product quantization grid $\Gamma_k^X \times \Gamma_k^Y$ of (X_k, Y_k) and kept in some accessible "look up tables". The online computation cost is then reduced to some more elementary operations.

Comparatively with the discrete state case, the preprocessing procedure involves here to quantize both the signal and observation processes. This means predefining optimal quantization grids in $\mathbb{R}^d \times \mathbb{R}^{d'}$ of the form $\Gamma_k^X \times \Gamma_k^Y$. As Y_k depends of X_k, it could have been intuitive to consider the quantization of the simulatable Markov process $Z_k = (X_k, Y_k)$, by choosing a quantization size (N_k^Z) and performing a gradient descent algorithm. According to known results on optimal vector quantization, we have:

$$\begin{aligned} \|X_k - \hat{X}_k\|_2 &\leq \|Z_k - \hat{Z}_k\|_2 = O((N_k^Z)^{\frac{-1}{d+d'}}), \\ \|Y_k - \hat{Y}_k\|_2 &\leq \|Z_k - \hat{Z}_k\|_2 = O((N_k^Z)^{\frac{-1}{d+d'}}). \end{aligned}$$

The fact is that, in our filtering problem, we have no need to quantize this joint distribution (X_k, Y_k). For the signal, we only need to store the quantization grids and then companion parameters like probability transitions (see [41]). As concerns the observation, it will be the quantization grids and likelihood evaluation on the product quantization grid of (X_k, Y_k). The idea is then to perform separate optimization procedures for the quantization of both marginal processes (X_k) and (Y_k). In fact, as X_k and Y_k are separately simulatable, we can e.g. perform a stochastic gradient descent algorithm for each fixed quantization size N_k^X and N_k^Y. The quantizer obtained for (X_k, Y_k) is still Voronoi (although usually not optimal for the couple (X_k, Y_k)). The point is that :

$$\begin{aligned} \|X_k - \hat{X}_k\|_2 &= O((N_k^X)^{\frac{-1}{d}}), \\ \|Y_k - \hat{Y}_k\|_2 &= O((N_k^Y)^{\frac{-1}{d'}}). \end{aligned} \tag{3.3.3}$$

This is moreover interesting as it allows to recycle signal quantizers to construct product grids, and accelerates the preprocessing procedure. It is worthnoting that the storage capacity for the precomputed grids (with companion parameters and associated likelihood values) can be drastically reduced in case of stationnary signal and observation processes (see Chapter 2).

Once the quantizers are preprocessed, the Algorithm 7 described below can be implementd.

Algorithm 7 Quantization based filtering algorithm with observation preprocessed

$\hat{\pi}_{\hat{y},0} = \hat{\mu}_0$

Transition from k to $k+1$: $\hat{\pi}_{\hat{y},k}$ is defined and $\hat{y}_k = y_k^{l_0}$.

- New observation y_{k+1},
- Projection : $\hat{y}_{k+1} = \sum_{m=1}^{N_{k+1}^Y} y_{k+1}^m \mathbf{1}_{A_{k+1}^m}(y_{k+1})$,
- Define m_0 the indice of y_{k+1} nearest neighbour on the grid Γ_{k+1}^Y.
- $\hat{H}_{\hat{y},k+1}^{ij} = \hat{\mathbf{P}}_k^{ij} g_{k+1}(x_k^i, y_k^{l_0}, x_{k+1}^j, y_{k+1}^{m_0})$.
- $\hat{\pi}_{\hat{y},k+1} = \hat{\pi}_{\hat{y},k} \hat{H}_{\hat{y},k+1}$.

The L^1-error for the filter estimate $\hat{\Pi}_{\hat{Y},n}$ resulting from Algorithm 7, after normalization, is controlled by Theorem 3.3.2.

Theorem 3.3.2 *Assume that for* $1 \leq k \leq n$:
H 1' For all $u, u', \bar{u}, \bar{u}'$ in $\mathbb{R}^{d'}$, x, x' in $\mathbb{R}^d$ we have:

$$|g_k(x,u,x',u') - g_k(x,\bar{u},x',\bar{u}')| \leq [g_k]_{Lip}^1 \left(|u-u'| + |\bar{u}-\bar{u}'|\right).$$

Hence, for each test function $f \in \mathcal{C}_{b,Lip}$ *and under assumptions* ***H2****,* ***H3*** *and* ***H4****, we have:*

$$\|\Pi_{Y,n} f - \hat{\Pi}_{\hat{Y},n} f\|_1 \leq 2\|f\|_\infty L_g^{n-1} c_n^N \sum_{k=1}^n ([g_k]_{Lip}^1 + [g_{k+1}]_{Lip}^1)\|Y_k - \hat{Y}_k\|_2 + L_g^n c_n^N \sum_{k=0}^n B_k^n(f)\|X_k - \hat{X}_k\|_2, \quad (3.3.4)$$

where $c_n^N = \min\{\|\frac{1}{\phi_n(Y)}\|_2, \|\frac{1}{\phi_n(\hat{Y})}\|_2, \|\frac{1}{\hat{\phi}_n(Y)}\|_2, \|\frac{1}{\hat{\phi}_n(\hat{Y})}\|_2\} \leq \|\frac{1}{\phi_n(Y)}\|_2 < +\infty$ *and* $N = (N_k)_{0\leq k\leq n}$.

When (3.3.3) is satisfied by the preprocessed quantizers, Theorem 3.3.2 gives a convergence rate to zero of the random filter estimation L^1-error.

Proof.

$$\|\Pi_{Y,n} f - \hat{\Pi}_{\hat{Y},n} f\|_1 \leq \|\Pi_{Y,n} f - \hat{\Pi}_{Y,n} f\|_1 + \|\hat{\Pi}_{Y,n} f - \hat{\Pi}_{\hat{Y},n} f\|_1.$$

The first term on the right hand side sum is controlled by Theorem 3.3.1 and **H** 2. For the second one we need further developments for the proof.
We consider a discrete state Markov chain with transition $\hat{\mathbf{P}}_k$ and such that $Z_k(\Omega) = X_k(\Omega)$ for each $0 \leq k \leq n$.
Then, we have:

$$\begin{aligned}\hat{H}_{y,k}f(x_{k-1}) &= \mathbb{E}[f(\hat{X}_k)g_k(x_{k-1}, y_{k-1}, \hat{X}_k, y_k)|\hat{X}_{k-1} = x_{k-1}],\\ &= \mathbb{E}[f(Z_k)g_k(x_{k-1}, y_{k-1}, Z_k, y_k)|Z_{k-1} = x_{k-1}].\end{aligned}$$

We denote by $\hat{\phi}_n(y) = \hat{\pi}_{y,n}\mathbf{1}$.
By the Markov property of Z_k, we have $\hat{\phi}_n(y) = \mathbb{E}[\prod_{k=1}^n g_k(Z_{k-1}, y_{k-1}, Z_k, y_k)]$. Now, by analogy with the discrete state case and particularly with Lemma 3.2.1, for two n-tuples of $\mathbb{R}^{d'}$, y and y':

$$\begin{aligned}|\hat{\phi}_n(y) - \hat{\phi}_n(y')| &\leq |\mathbb{E}[\prod_{k=1}^n g_k(Z_{k-1}, y_{k-1}, Z_k, y_k) - \prod_{k=1}^n g_k(Z_{k-1}, y'_{k-1}, Z_k, y'_k)]|,\\ &\leq |\mathbb{E}[\prod_{k=1}^n g_k(Z_{k-1}, y_{k-1}, Z_k, y_k) - \prod_{k=1}^{n-1} g_k(Z_{k-1}, y'_{k-1}, Z_k, y'_k)g_n(Z_{n-1}, y_{n-1}, Z_n, y_n)]|\\ &\quad +|\mathbb{E}[\prod_{k=1}^{n-1} g_k(Z_{k-1}, y'_{k-1}, Z_k, y'_k)g_n(Z_{n-1}, y_{n-1}, Z_n, y_n) - \prod_{k=1}^n g_k(Z_{k-1}, y'_{k-1}, Z_k, y'_k)]|,\\ &\leq L_g|\hat{\phi}_{n-1}(y) - \hat{\phi}_{n-1}(y')| + L_g^{n-1}[g_n]^1_{Lip}\left(|y_{n-1} - y'_{n-1}| + |y_n - y'_n|\right),\\ &\leq L_g^{n-1}\sum_{k=1}^{n-1}\left([g_{k+1}]^1_{Lip} + [g_k]^1_{Lip}\right)|y_k - y'_k| + L_g^{n-1}[g_n]^1_{Lip}|y_n - y'_n| + L_g^{n-1}[g_1]^1_{Lip}|y_0 - y'_0|.\end{aligned}$$

Then, by Lemma 3.1 in [41], we have :

$$\|\hat{\Pi}_{Y,n}f - \hat{\Pi}_{\hat{Y},n}f\|_1 \leq 2\|f\|_\infty L_g^n\|\frac{1}{\hat{\phi}_n(Y)\vee\hat{\phi}_n(\hat{Y})}\|_2\sum_{k=1}^n([g_k]^1_{Lip} + [g_{k+1}]^1_{Lip})\|Y_k - \hat{Y}_k\|_2$$

Hence,

$$\begin{aligned}\|\Pi_{Y,n}f - \hat{\Pi}_{\hat{Y},n}f\|_1 \leq 2\|f\|_\infty L_g^n\|\frac{1}{\hat{\phi}_n(Y)\vee\hat{\phi}_n(\hat{Y})}\|_2\sum_{k=1}^n([g_k]^1_{Lip} + [g_{k+1}]^1_{Lip})\|Y_k - \hat{Y}_k\|_2\\ + 2\|f\|_\infty L_g^n\|\frac{1}{\phi_n(Y)\vee\phi_n(\hat{Y})}\|_2\sum_{k=0}^n B_k^n(f)\|X_k - \hat{X}_k\|_2. \quad (3.3.5)\end{aligned}$$

By symmetry, we can construct an inequality equivalent to (3.3.5), starting from :

$$\|\Pi_{Y,n}f - \hat{\Pi}_{\hat{Y},n}f\|_1 \leq \|\Pi_{Y,n}f - \Pi_{\hat{Y},n}f\|_1 + \|\Pi_{\hat{Y},n}f - \hat{\Pi}_{\hat{Y},n}f\|_1.$$

So, using Theorem 3.2.1 and Theorem 3.3.1, we have:

$$\|\Pi_{Y,n}f - \hat{\Pi}_{\hat{Y},n}f\|_1 \leq 2\|f\|_\infty L_g^n \|\frac{1}{\phi_n(Y)\vee\phi_n(\hat{Y})}\|_2 \sum_{k=1}^{n}([g_k]^1_{Lip} + [g_{k+1}]^1_{Lip})\|Y_k - \hat{Y}_k\|_2$$
$$+ 2\|f\|_\infty L_g^n \|\frac{1}{\phi_n(\hat{Y})\vee\hat{\phi}_n(\hat{Y})}\|_2 \sum_{k=0}^{n} B_k^n(f)\|X_k - \hat{X}_k\|_2. \quad (3.3.6)$$

Setting in (3.3.5) or in (3.3.6):

$$c_n^N = \min\{\|\frac{1}{\phi_n(Y)}\|_2, \|\frac{1}{\phi_n(\hat{Y})}\|_2, \|\frac{1}{\hat{\phi}_n(Y)}\|_2, \|\frac{1}{\hat{\phi}_n(\hat{Y})}\|_2\},$$

achieves the proof. □

3.3.1 Example

In this section we choose to treat numerically a linear Gaussian state model. This leads to the well known Kalman filter. In this case, the filter distribution is Gaussian and an explicit solution is given by sequential computations of its first two moments. We set:

$$\begin{cases} X_k = \rho X_{k-1} + \theta\varepsilon_{k+1}, \quad X_0 \sim \mathcal{N}(0, trans\Sigma_0\Sigma_0) \\ Y_k = X_k + \alpha\eta_k, \\ \varepsilon_k \ \ and \ \ \eta_k \quad iid \quad \sim \mathcal{N}(0,1), \\ |\rho| < 1, \quad \rho, \theta \in \mathbb{R}. \end{cases} \quad (3.3.7)$$

$\Pi_{y,n}$ is a Gaussian distribution with mean and variance depending on the observation process and which can be computed sequentially by the means of a deterministic algorithm [21]. In our special Kalman filter case, we note that both signal and observation marginal distributions are Gaussian. Hence, a fast preprocessing procedure will consist in using pre-computed Gaussian quantizers, kept online on `http://www.proba.jussieu.fr/pageperso/pages.html`. In addition, we take the signal stationary distribution, as the a initial one which allows to consider one single transition step for transition parameters $\hat{\mathbf{P}}_k^{ij}$ (Cf. Chapter 2).

For application, we fix $n = 25$ and choose three test functions $f_1(x) = x$, $f_2(x) = x^2$ and $f_3(x) = e^{-|x|}$. Note that assumptions needed in Theorem 3.3.2 are satisfied : $K \leq \rho$, $L_g \leq \frac{1}{\sqrt{2\pi}\alpha}$ and $[g_k]^1_{Lip} = [g_k]^2_{Lip} \leq \sqrt{\frac{e}{2\pi\alpha^3}}$.

The L^1-error is computed via a Monte Carlo estimation over $MC = 5000$ observation realizations. In Figure 3.3 we represent it as a function of $N = N_k^X = N_k^Y$ the quantizer size of both the signal and the observation processes. We see that the use of observation quantization has a very little impact on the error rates. On the log scale figure (Figure 3.4), the convergence rate is deduced from the linear regression slope. As in the discrete state case the theoretical rate $\frac{1}{d} = 1$ is obtained.

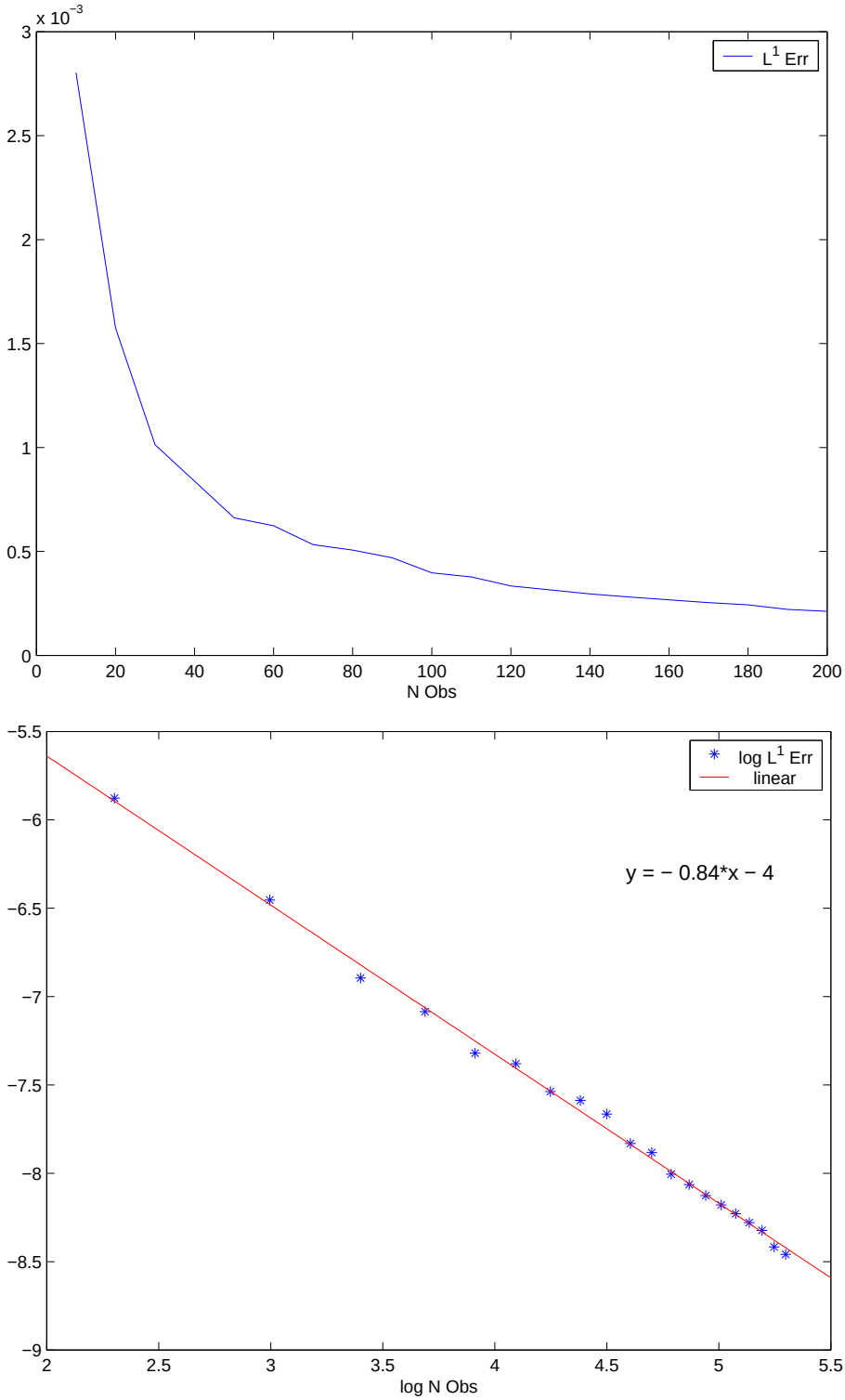

Figure 3.1: L^1-error on discrete state filters computed with observation quantization - $f = Id$ - $(n = 5)$

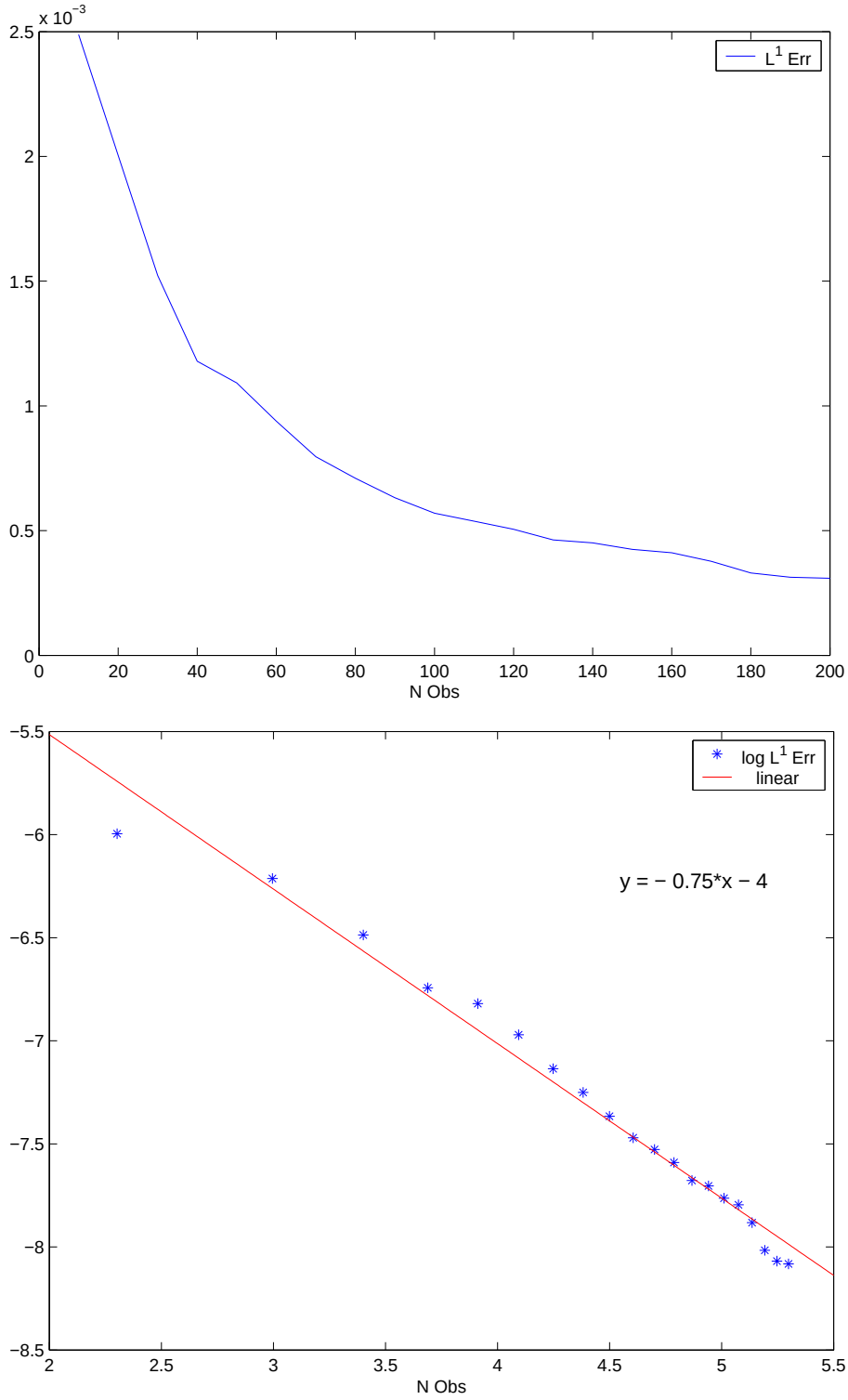

Figure 3.2: L^1-error on discrete state filters computed with observation quantization - $f = Id$ - $(n = 20)$

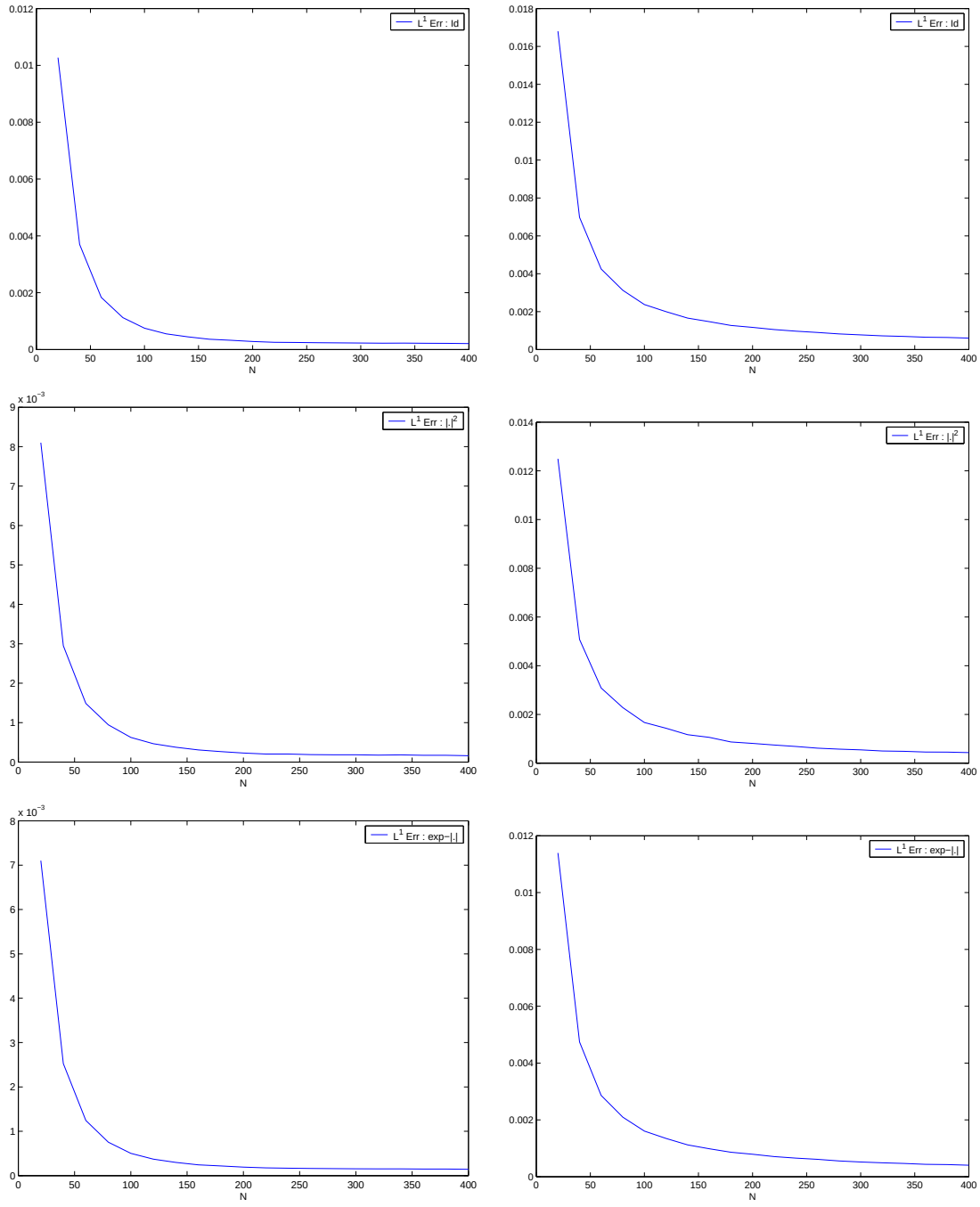

Figure 3.3: L^1-error on continuous state filter estimates evaluated for three test functions - (n=25) - (left : with signal quantization, right : with both signal and observation quantization)

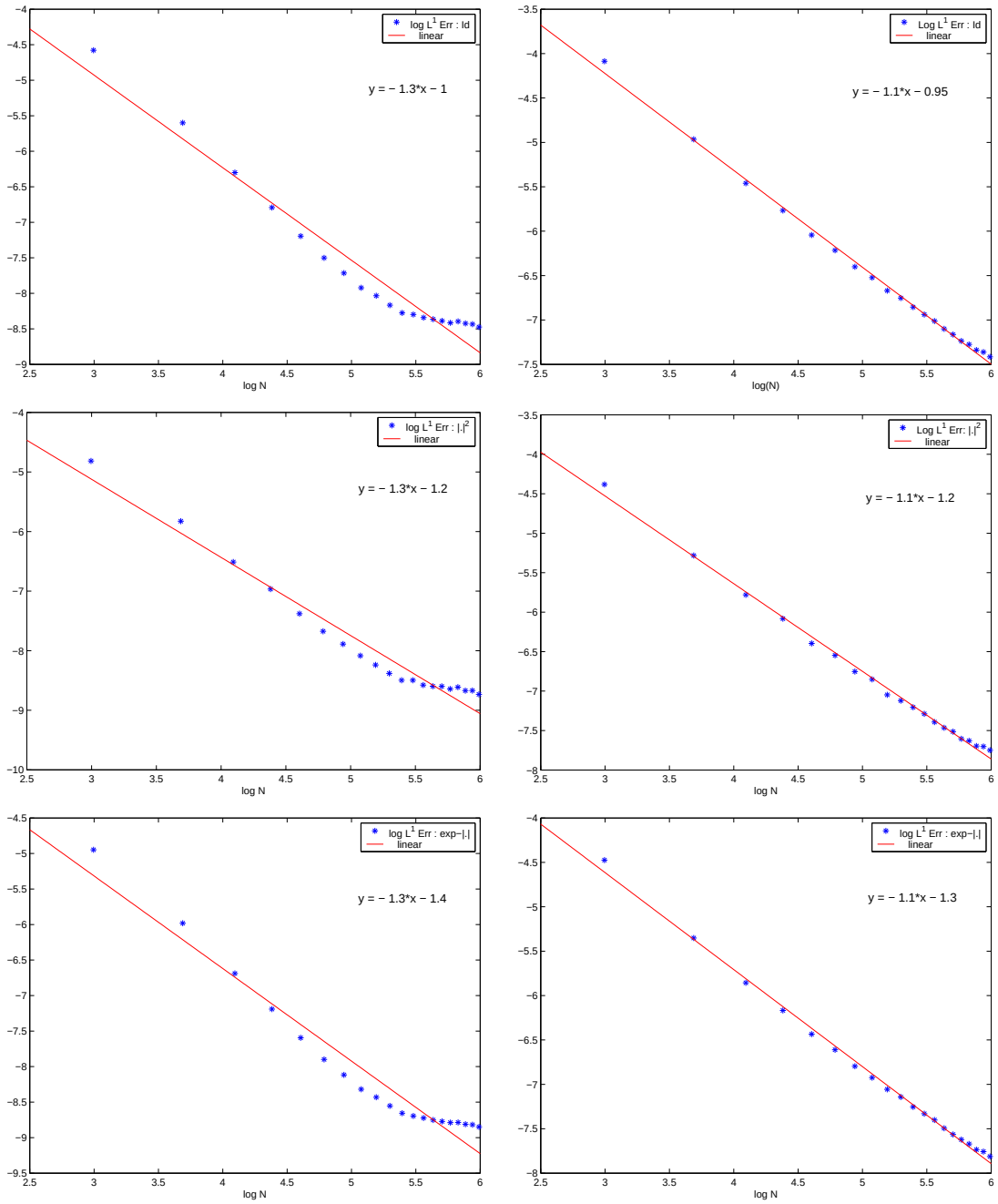

Figure 3.4: L^1-error on continuous state filter estimates evaluated for three test functions in $\log - \log$ scale - (n=25) - (left : with signal quantization, right : with both signal and observation quantization)

Chapter 4

Approximation by quantization of the filter process and applications to optimal stopping problems under partial observation

Joint paper with H. Pham and W. Runggaldier [47].

We present an approximation method for discrete time nonlinear filtering in view of solving dynamic optimization problems under partial information. The method is based on quantization of the Markov pair process filter-observation (Π, Y) and is such that, at each time step k and for a given size N_k of the quantization grid in period k, this grid is chosen to minimize a suitable quantization error. The algorithm is based on a stochastic gradient descent combined with Monte-Carlo simulations of (Π, Y). Convergence results are given and applications to optimal stopping under partial observation are discussed. Numerical results are presented for a particular stopping problem : American option pricing with unobservable volatility.

Key words: Nonlinear filtering, Markov chain, quantization, stochastic gradient descent, partial observation, optimal stopping.

4.1 Introduction

We consider a discrete time, partially observable process (X, Y) where X represents the state or signal process that may not be observable, while Y is the observation. The signal process $\{X_k, k \in \mathbb{N}\}$ is valued in a measurable space $(E, \mathcal{E})$ and is a Markov chain with probability transition (P_k) (i.e. the transition from time $k-1$ to time k), and initial law μ. The observation sequence (Y_k) is valued in $\mathbb{R}^d$ and such that the pair (X_k, Y_k) is a Markov chain and

(H) The law of Y_k conditional on (X_{k-1}, Y_{k-1}, X_k), $k \geq 1$, denoted $q_k(X_{k-1}, Y_{k-1}, X_k, dy')$, admits a bounded density

$$y' \longmapsto g_k(X_{k-1}, Y_{k-1}, X_k, y').$$

For simplicity, we assume that Y_0 is a known deterministic constant equal to y_0. Notice that the probability transition of the Markov chain $(X_k, Y_k)_{k\in\mathbb{N}}$ is then given by $P_k(x, dx')g_k(x, y, x', y')dy'$ with initial law $\mu(dx)\delta_{y_0}(dy)$.

We denote by $(\mathcal{F}_k^Y)$ the filtration generated by the observation process (Y_k) and by Π_k the filter conditional law of X_k given $\mathcal{F}_k^Y$:

$$\Pi_k(dx) \quad = \quad \mathbb{P}\left[X_k \in dx | \, \mathcal{F}_k^Y\right], \quad k \in \mathbb{N}.$$

The filter process $(\Pi_k)_k$ allows one to transform problems related to the partially observed process (X, Y) into equivalent problems under complete observation related to the pair (Π, Y) and this latter pair turns out to be Markov with respect to the observation filtration $(\mathcal{F}_k^Y)$. An important class of problems related to partially observable processes are stochastic control and stopping problems that are dynamic stochastic optimization problems where the partial observation concerns the state/signal process. We shall, in particular, consider optimal stopping under partial information and, in this context, a financial application concerning the pricing of Asian options in a partially observed stochastic volatility model.

We shall assume that the state space E of the state/signal process (X_k) consists of a finite number m of points. Problems with a more general state space can be approximated by problems with a finite state space (see e.g. [31]). With X_k taking the m values $x^i\,(i = 1, \cdots, m)$, the filter process is characterized by an $m-$vector with components $\Pi_k^i = \mathbb{P}[X_k = x^i | \mathcal{F}_k^Y]$ and takes thus values in the $m-$simplex K_m in $\mathbb{R}^m$. While the filter process allows thus to transform a problem under partial observation into one under full observation, it has the drawback that a finite-valued state space becomes infinite. This leads to difficulties when trying to solve dynamic optimization problems also in discrete time. For actual computation one has thus to discretize the process Π_k, approximating it with a process that takes a finite number of values in the simplex K_m. Approaches to this effect have appeared in the literature; they are based on discretizing the observations process (Y_k) and then approximating Π_k by a filter of X_k, given the discretized

observations path (see e.g. [35], [9], [55], [38] and references therein). Such an approach has the drawback that the number of possible values for the approximating filter grows exponentially fast with the time step.

In this paper, we propose a new approach by exploiting the Markov property of $(\Pi, Y)_k$ with respect to the observation filtration $(\mathcal{F}_k^Y)$. This means that the conditional law of X_{k+1} given $\mathcal{F}_k^Y$ is summarized by the sufficient statistics (Π_k, Y_k) valued in $K^m \times \mathbb{R}^d$. This suggests to approximate the couple process (Π, Y) with $(\hat{\Pi}, \hat{Y})$ that in the generic time step k takes a number of values N_k that can be assigned arbitrarily. Following standard usage, we shall call this *quantization* and the problem then arises to find an *optimal* quantization, namely such that it minimizes for each time period k the L^2-approximation error induced by the quantization. This error is related to what e.g. in information theory is called *distorsion*. The implementation of the minimizing algorithm itself is based on a stochastic gradient descent method combined with Monte-Carlo simulations of (Π, Y). As a byproduct, we obtain an approximation of the probability transition matrices of the Markov chain (Π, Y). These companion parameters provide in turn an approximation algorithm , via the dynamic programming principle, for computing optimal values associated to dynamic optimization problems under partial observation.

Optimal quantization methods have been developed recently in numerical probability and various problems of optimal stopping, control or nonlinear filtering, see [40], [4], [5], [42], [43], [41]. In our context, we obtain error bounds and rate of convergence for the approximation of the pair filter-observation process (Π, Y). We then give an application to the approximation of optimal stopping problem under partial observation and study associated convergence results. Finally, we present a numerical illustration for the American option pricing problem with unobservable volatility.

The rest of the paper is organized as follows. In Section 2, we recall some preliminaries on nonlinear filtering. Section 3 is the heart of the paper. We first present the quantization approximation method for the filter-observation Markov chain, and then analyze the induced error. The rate of convergence and the practical implementation of the algorithm for the optimal quantization are discussed. In Section 4, we apply our quantization method to the approximation method of an optimal stopping problem under partial observation. Convergence results for the associated optimal value functions are provided. The last Section 5 presents numerical tests for the American option pricing with unobservable volatility.

4.2 Preliminaries

We denote by $\mathcal{M}(E)$ the set of finite nonnegative measures on $(E, \mathcal{E})$ and by $\mathcal{P}(E)$ the subset of probability measures on $(E, \mathcal{E})$. It is known that $\mathcal{M}(E)$ is a Polish space equipped with the weak topology, hence a measurable space endowed with the Borel σ-field. We recall

that from Bayes formula, Π_k is given in inductive form by :

$$\begin{aligned} \Pi_0 &= \mu \\ \Pi_k &= \bar{G}_k(\Pi_{k-1}, Y_{k-1}, Y_k), \quad k \geq 1, \end{aligned} \tag{4.2.1}$$

where G_k is the continuous function from $\mathcal{M}(E) \times \mathbb{R}^d \times \mathbb{R}^d$ into $\mathcal{M}(E)$ defined by :

$$G_k(\pi, y, y')(dx') = \int_E g_k(x, y, x', y') P_k(x, dx') \pi(dx),$$

and $\bar{G}_k$ is the normalized continuous function valued in $\mathcal{P}(E)$:

$$\bar{G}_k(\pi, y, y') = \frac{G_k(\pi, y, y')}{\int_E G_k(\pi, y, y')(dx')}.$$

Denoting by $(\mathcal{F}_k)$ the filtration generated by (X_k, Y_k), and using the law of iterated conditional expectations, we have for any k and bounded Borelian function φ on $\mathcal{P}(E) \times \mathbb{R}^d$:

$$\begin{aligned} & \mathbb{E}\left[\varphi(\Pi_{k+1}, Y_{k+1}) \middle| \mathcal{F}_k^Y\right] \\ = \; & \mathbb{E}\left[\mathbb{E}[\varphi(\bar{G}_{k+1}(\Pi_k, Y_k, Y_{k+1}), Y_{k+1}) | \mathcal{F}_k] \middle| \mathcal{F}_k^Y\right] \\ = \; & \mathbb{E}\left[\int \varphi(\bar{G}_{k+1}(\Pi_k, Y_k, y'), y') P_{k+1}(X_k, dx') q_{k+1}(X_k, Y_k, x', dy') \middle| \mathcal{F}_k^Y\right] \\ = \; & \int \varphi(\bar{G}_{k+1}(\Pi_k, Y_k, y'), y') P_{k+1}(x, dx') q_{k+1}(x, Y_k, x', dy') \Pi_k(dx). \end{aligned} \tag{4.2.2}$$

This proves that the pair $(\Pi_k, Y_k)_k$ is a $(\mathbb{P}, (\mathcal{F}_k^Y)_k)$ Markov chain in $\mathcal{P}(E) \times \mathbb{R}^d$, with initial value (μ, y_0). Moreover, under **(H)**, this also shows that the (unnormalized) law of Y_k conditional on (Π_{k-1}, Y_{k-1}), denoted $Q_k(\Pi_{k-1}, Y_{k-1}, dy')$ admits a density given by :

$$y' \longmapsto \int G_k(\Pi_{k-1}, Y_{k-1}, y')(dx') = \int g_k(x, Y_{k-1}, x', y') P_k(x, dx') \Pi_{k-1}(dx) \tag{4.2.3}$$

Notice that the probability transition R_k (from time $k-1$ to k) of the Markov chain $(\Pi_k, Y_k)_k$ is not explicit in general. Actually, from (4.2.2), we can write :

$$R_k \varphi(\pi, y) = \int \varphi(\bar{G}_k(\pi, y, y'), y') Q_k(\pi, y, dy'), \quad \forall (\pi, y) \in \mathcal{P}(E) \times \mathbb{R}^d. \tag{4.2.4}$$

In the sequel, we denote by $|.|_2$ the Euclidian norm and by $|.|_1$ the l^1 norm on $\mathbb{R}^l$. For any $\mathbb{R}^l$-valued random variable U, we denote :

$$\|U\|_2 = \left(\mathbb{E}|U|_2^2\right)^{\frac{1}{2}} \quad \text{and} \quad \|U\|_1 = \mathbb{E}|U|_1.$$

4.3 An optimal quantization approach for the approximation of the filter process

4.3.1 The approximation method

We assume here that the state space E of the signal process (X_k) is finite consisting of m points : $E = \{x^1, \ldots, x^m\}$. The initial discrete law $\mu = (\mu^i)$ and the probability transition matrix P_k are defined by :

$$\begin{aligned} \mu^i &= \mathbb{P}[X_0 = x^i], \quad i = 1, \ldots, m, \\ P_k^{ij} &= \mathbb{P}[X_k = x^j | X_{k-1} = x^i], \quad i, j = 1, \ldots, m. \end{aligned}$$

The random filter Π_k is characterized by its random weights :

$$\Pi_k^i = \mathbb{P}[X_k = x^i | \mathcal{F}_k^Y], \quad i = 1, \ldots, m,$$

and may then be identified with a random vector valued in the m-simplex K_m in $\mathbb{R}^m$ of dimension $m-1$:

$$K_m = \left\{ \pi = (\pi^i) \in \mathbb{R}^m \ : \pi^i \geq 0, \ 1 \leq i \leq m, \ |\pi|_1 = \sum_{i=1}^m \pi^i = 1 \right\}.$$

By (4.2.1), it is expressed in the recursive form :

$$\begin{aligned} \Pi_0 &= \mu \\ \Pi_k &= \bar{G}_k(\Pi_{k-1}, Y_{k-1}, Y_k) = \frac{GP_k(Y_{k-1}, Y_k)^\top \Pi_{k-1}}{|GP_k(Y_{k-1}, Y_k)^\top \Pi_{k-1}|_1}, \end{aligned} \tag{4.3.1}$$

where $GP_k(Y_{k-1}, Y_k)$ is a $m \times m$ random matrix given by :

$$GP_k(Y_{k-1}, Y_k)_{ij} = g_k(x_{k-1}^i, Y_{k-1}, x_k^j, Y_k) P_k^{ij}, \quad 1 \leq i, j \leq m. \tag{4.3.2}$$

Here $M^\top$ denotes the transpose of a matrix M.

A first approach for approximating the filter process (Π_k) consists in discretizing the observation process (Y_k) by replacing it by a process $(\hat{Y}_k)$ taking a finite number N of values, and then approximate for each k the random filter Π_k by a random filter of X_k given $\hat{Y}_1, \ldots, \hat{Y}_k$. So at each time step k, given $\hat{Y}_k$ and $\hat{\Pi}_k$, the random variable $\hat{\Pi}_{k+1}$ may take N values. Thus, at time n, the random filter $\hat{\Pi}_n$ is identified with a random vector taking N^n possible values. All these values are precomputed and stored in a "look-up table" but this could be very heavy, typically when n is large. Such an approach was introduced in [35] and [9], and investigated further in [55], [38] and Chapter 3.

We propose here a new approach. This starts from the key remark that the pair process (Π_k, Y_k) is Markov with respect to the observation filtration $(\mathcal{F}_k^Y)$. In other words, the

conditional law of X_{k+1} given $\mathcal{F}_k^Y$ may be summarized by the sufficient statistic (Π_k, Y_k). Therefore, since the Markov chain (Π_k, Y_k) is completely characterized by its probability transitions, the idea is to approximate these probability transitions by suitable probability transition matrices.

In a first step, we discretize for each k the couple (Π_k, Y_k) by approximating it by $(\hat{\Pi}_k, \hat{Y}_k)$ taking a finite number of values. The space discretization (or quantization) of the random vector $Z_k = (\Pi_k, Y_k)$ in $K_m \times \mathbb{R}^d$ is constructed as follows. At initial time $k = 0$, recall that Z_0 is a known deterministic vector equal to $z_0 = (\mu, y_0)$, so we start from the grid with one point in $K_m \times \mathbb{R}^d$:

$$\Gamma_0 \;=\; \{z_0 = (\mu, y_0)\}\,.$$

At time $k \geq 1$, we are given a grid Γ_k of N_k points in $K_m \times \mathbb{R}^d$:

$$\Gamma_k \;=\; \left\{z_k^1 = (\pi_k(1), y_k^1), \ldots, z_k^{N_k} = (\pi_k(N_k), y_k^{N_k})\right\},$$

and we denote by $C_i(\Gamma_k)$, $i = 1, \ldots, N_k$, the associated Voronoi tesselations :

$$C_i(\Gamma_k) \;=\; \left\{z \in K_m \times \mathbb{R}^d \;: \mathrm{Proj}_{\Gamma_k}(z) = z_k^i\right\}, \quad i = 1, \ldots, N_k.$$

Here Proj_{Γ_k} is a closest neighbor projection for the Euclidian norm :

$$\left|z - \mathrm{Proj}_{\Gamma_k}(z)\right|_2 \;=\; \min_{i=1,\ldots,N_k} |z - z_k^i|_2, \quad \forall\, z \in K_m \times \mathbb{R}^d.$$

Notice that by definition of the Euclidian norm, we clearly have :

$$\mathrm{Proj}_{\Gamma_k} \;=\; \left(\mathrm{Proj}_{\Gamma_k^\Pi}, \mathrm{Proj}_{\Gamma_k^Y}\right), \qquad C_i(\Gamma_k) \;=\; C_i(\Gamma_k^\Pi) \times C_i(\Gamma_k^Y),$$

where :

$$\begin{aligned}
\Gamma_k^\Pi \;&=\; \mathrm{Proj}_{K_m}(\Gamma_k) \;=\; \{\pi_k(1), \ldots, \pi_k(N_k)\} \\
\Gamma_k^Y \;&=\; \mathrm{Proj}_{\mathbb{R}^d}(\Gamma_k) \;=\; \left\{y_k^1, \ldots, y_k^{N_k}\right\}.
\end{aligned}$$

We then approximate the pair $Z_k = (\Pi_k, Y_k)$ by $\hat{Z}_k = (\hat{\Pi}_k, \hat{Y}_k)$ valued in Γ_k and defined by :

$$\hat{Z}_k \;=\; \mathrm{Proj}_{\Gamma_k}(Z_k) \;=\; \left(\mathrm{Proj}_{\Gamma_k^\Pi}(\Pi_k), \mathrm{Proj}_{\Gamma_k^Y}(Y_k)\right).$$

In a second step, we approximate the probability transitions of the Markov chain (Z_k) :

$$R_k(z, dz') \;=\; \mathbb{P}\left[Z_k \in dz' | Z_{k-1} = z\right], \quad k \geq 1,\; z \in K_m \times \mathbb{R}^d,$$

by the following probability transition matrix :

$$\begin{aligned}
\hat{r}_k^{ij} \;&=\; \mathbb{P}\left[\hat{Z}_k \;=\; z_k^j \middle|\, \hat{Z}_{k-1} \;=\; z_{k-1}^i\right] \\
&=\; \frac{\mathbb{P}\left[Z_k \in C_j(\Gamma_k),\; Z_{k-1} \in C_i(\Gamma_{k-1})\right]}{\mathbb{P}\left[Z_{k-1} \in C_i(\Gamma_{k-1})\right]} \;=:\; \frac{\hat{\beta}_k^{ij}}{\hat{p}_{k-1}^i},
\end{aligned}$$

for all $k \geq 1$, $i = 1, \ldots, N_{k-1}$, $j = 1, \ldots, N_k$. We shall see later how the grids Γ_k and the number of points N_k are optimally chosen and implemented, and how the associated probability transition matrix $\hat{r}_k$ can be estimated.

Example : Computation of predictor conditional expectation.

Suppose we are interested in the computation for any $k = 0, \ldots, n$, and for arbitrary measurable function φ_{k+1} on $K_m \times \mathbb{R}^d$, of the filter predictor :

$$U_k \;=\; \mathbb{E}\left[\varphi_{k+1}(X_{k+1}, Y_{k+1}) \middle| \mathcal{F}_k^Y\right].$$

A precise application where such $\mathcal{F}_k^Y$-measurable random variables appear is presented in Section 4. Then, by introducing the function :

$$\hat{\varphi}_{k+1}(\pi, y) \;=\; \sum_{i=1}^{m} \varphi_{k+1}(x^i, y)\pi^i, \quad \forall\, \pi = (\pi^i)_i \in K_m,\ \forall y \in \mathbb{R}^d,$$

and using the law of iterated conditional expectation, we can rewrite U_k as :

$$U_k \;=\; \mathbb{E}\left[\hat{\varphi}_{k+1}(\Pi_{k+1}, Y_{k+1}) \middle| \mathcal{F}_k^Y\right] \;=\; \mathbb{E}\left[\hat{\varphi}_{k+1}(Z_{k+1}) \middle| \mathcal{F}_k^Y\right].$$

We thus approximate the sequence of $\mathcal{F}_k^Y$-measurable random variables U_k by $\hat{U}_k = \hat{v}_k(\hat{Z}_k)$ where the functions $\hat{v}_k$, $k = 0, \ldots, n$, are defined on Γ_k by :

$$\begin{aligned} \hat{v}_k(z_k^i) \;&=\; \mathbb{E}\left[\hat{\varphi}_{k+1}(\hat{Z}_{k+1}) \middle| \hat{Z}_k = z_k^i\right] \\ &=\; \sum_{j=1}^{N_{k+1}} \hat{r}_{k+1}^{ij} \hat{\varphi}_{k+1}(z_{k+1}^j), \quad \forall\, z_k^i \in \Gamma_k. \end{aligned}$$

4.3.2 The error analysis

The quality of the approximation described in the previous paragraph is measured as follows. We denote for any subset D in $\mathbb{R}^l$:

$$BL_1(D) \;=\; \left\{\varphi \;\text{ Borelian from } D \;\text{ into } \mathbb{R} \;:\; \|\varphi\|_{sup} := \sup_{x \in D} |\varphi(x)| \leq 1, \quad [\varphi]_{lip} := \sup_{x,y \in D, x \neq y} \frac{|\varphi(x) - \varphi(y)|}{|x - y|_1} \leq 1 \right\}.$$

We make the following assumption.

(H1) There exists a constant L_g such that for all $k \geq 1$:

$$\sum_{i,j=1}^{m} P_k^{ij} \int \left|g_k(x^i, y, x^j, y') - g_k(x^i, \hat{y}, x^j, y')\right| dy' \;\leq\; L_g |y - \hat{y}|_1, \quad \forall y, \hat{y} \in \mathbb{R}^d.$$

Proposition 4.3.1 *Under* **(H1)**, *we have for any* n *and* $\varphi_1, \ldots, \varphi_n \in BL_1(K_m \times \mathbb{R}^d)$:

$$\begin{aligned} &\left| \int \varphi_1(z_1) \ldots \varphi_n(z_n) \left(R_1(z_0, dz_1) \ldots R_n(z_{n-1}, dz_n) \; - \; \hat{r}_1(z_0, dz_1) \ldots \hat{r}_n(z_{n-1}, dz_n) \right) \right| \\ \leq \; & \sum_{k=1}^{n} \frac{3\sqrt{m+d}}{2\bar{L}_g - 1} \left[(2\bar{L}_g)^{n-k+1} - 1 \right] \; \left\| Z_k - \hat{Z}_k \right\|_2 , \end{aligned} \tag{4.3.3}$$

where $\bar{L}_g = \max(L_g, 1)$.

We first state a preliminary result on the Lipschitz property of the transition of the Markov chain $(Z_k)_k = (\Pi_k, Y_k)_k$.

Lemma 4.3.1 *Under* **(H1)**, *we have for all* $k \geq 1$ *and Borelian function* φ *on* $K_m \times \mathbb{R}^d$:

$$|R_k\varphi(z) - R_k\varphi(\hat{z})| \;\; \leq \;\; \left(2[\varphi]_{lip} + \|\varphi\|_{sup} \right) \bar{L}_g \, |z - \hat{z}|_1, \quad \forall z, \hat{z} \in K_m \times \mathbb{R}^d.$$

Proof. Recall from (4.2.3) that the conditional law $Q_k(\pi, y, dy')$ of Y_k given (Π_{k-1}, Y_{k-1}) $= (\pi, y)$ admits a density given by :

$$y' \;\; \longmapsto \;\; f_k(\pi, y, y') \; = \; \sum_{i,j=1}^{m} g_k(x^i, y, x^j, y') P_k^{ij} \pi^i.$$

This conditional density satisfies the Lipschitz property : for all (π, y) and $(\hat{\pi}, \hat{y}) \in K_m \times \mathbb{R}^d$,

$$\begin{aligned} \int \left| f_k(\pi, y, y') - f_k(\hat{\pi}, \hat{y}, y') \right| dy' \;\; \leq \;\; & \sum_{i,j=1}^{m} P_k^{ij} \int \left| g_k(x^i, y, x^j, y') - g_k(x^i, \hat{y}, x^j, y') \right| dy' \\ & + \sum_{i}^{m} |\pi^i - \hat{\pi}^i|, \end{aligned} \tag{4.3.4}$$

where we used the fact that $\sum_j P_k^{ij} \int g_k(x^i, y, x^j, y') dy' = 1$. From (4.2.4), we then have for any $z = (\pi, y)$ and $\hat{z} = (\hat{\pi}, \hat{y}) \in K_m \times \mathbb{R}^d$:

$$\begin{aligned} |R_k\varphi(z) - R_k\varphi(\hat{z})| \;\; \leq \;\; & \int \left| \varphi(\bar{G}_k(\pi, y, y'), y') - \varphi(\bar{G}_k(\hat{\pi}, \hat{y}, y'), y') \right| Q_k(\pi, y, dy') \\ & + \left| \int \varphi(\bar{G}_k(\hat{\pi}, \hat{y}, y'), y') \left(Q_k(\pi, y, dy') - Q_k(\hat{\pi}, \hat{y}, dy') \right) \right| \\ \leq \;\; & [\varphi]_{lip} \int \left| \bar{G}_k(\pi, y, y') - \bar{G}_k(\hat{\pi}, \hat{y}, y') \right|_1 f_k(\pi, y, y') dy' \\ & + \|\varphi\|_{sup} \int \left| f_k(\pi, y, , y') - f_k(\hat{\pi}, \hat{y}, y') \right| dy' \\ \leq \;\; & [\varphi]_{lip} \int \left| \bar{G}_k(\pi, y, y') - \bar{G}_k(\hat{\pi}, \hat{y}, y') \right|_1 f_k(\pi, y, y') dy' \\ & + \|\varphi\|_{sup} \sum_{i,j=1}^{m} P_k^{ij} \int \left| g_k(x^i, y, x^j, y') - g_k(x^i, \hat{y}, x^j, y') \right| dy' \\ & + \|\varphi\|_{sup} \sum_{i}^{m} |\pi^i - \hat{\pi}^i|. \end{aligned} \tag{4.3.5}$$

Now, from (4.3.2), we have :

$$\begin{aligned}
& \int \left|\bar{G}_k(\pi,y,y') - \bar{G}_k(\hat{\pi},\hat{y},y')\right|_1 f_k(\pi,y,y')dy' \\
\leq & \sum_{j=1}^{m} \int \left|\bar{G}_k^j(\pi,y,y') - \bar{G}_k^j(\hat{\pi},\hat{y},y')\right| f_k(\pi,y,y')dy' \\
= & \sum_{i,j=1}^{m} \int \left| \frac{g_k(x^i,y,x^j,y')P_k^{ij}\pi^i}{f_k(\pi,y,y')} - \frac{g_k(x^i,\hat{y},x^j,y')P_k^{ij}\hat{\pi}^i}{f_k(\hat{\pi},\hat{y},y')} \right| f_k(\pi,y,y')dy' \\
\leq & \sum_{i,j=1}^{m} P_k^{ij}\hat{\pi}^i \int \frac{\left|g_k(x^i,y,x^j,y')f_k(\hat{\pi},\hat{y},y') - g_k(x^i,\hat{y},x^j,y')f_k(\pi,y,y')\right|}{f_k(\hat{\pi},\hat{y},y')}dy' \\
& + \sum_{i=1}^{m} |\pi^i - \hat{\pi}^i| \\
\leq & \sum_{i,j=1}^{m} P_k^{ij} \int \left|g_k(x^i,y,x^j,y') - g_k(x^i,\hat{y},x^j,y')\right| dy' \\
& + \int \left|f_k(\pi,y,y') - f_k(\hat{\pi},\hat{y},y')\right| dy' \; + \; \sum_{i=1}^{m} |\pi^i - \hat{\pi}^i| \\
\leq & \; 2\sum_{i,j=1}^{m} P_k^{ij} \int \left|g_k(x^i,y,x^j,y') - g_k(x^i,\hat{y},x^j,y')\right| dy' \; + \; 2\sum_{i=1}^{m} |\pi^i - \hat{\pi}^i|,
\end{aligned}$$

where we used again (4.3.4). Plugging into (4.3.5) and using **(H1)** yield :

$$\begin{aligned}
& |R_k\varphi(z) - R_k\varphi(\hat{z})| \\
\leq & \left(2[\varphi]_{lip} + \|\varphi\|_{sup}\right) \left(\sum_{i,j=1}^{m} P_k^{ij} \int \left|g_k(x^i,y,x^j,y') - g_k(x^i,\hat{y},x^j,y')\right| dy' \; + \; \sum_{i=1}^{m} |\pi^i - \hat{\pi}^i| \right) \\
\leq & \left(2[\varphi]_{lip} + \|\varphi\|_{sup}\right) \left(L_g|y-\hat{y}|_1 + |\pi - \hat{\pi}|_1\right), \qquad (4.3.6)
\end{aligned}$$

and then the required result. □

Remark 4.3.1 In the case where the law of Y_k conditional on (X_{k-1}, Y_{k-1}, X_k) does not depend on Y_{k-1}, i.e. its density g_k depends only X_{k-1}, X_k, the condition **(H1)** is empty, or in other words is trivially satisfied wth $L_g = 0$. Then, inequality (4.3.6) shows that for all $z = (\pi, y)$, $\hat{z} = (\hat{\pi}, \hat{y}) \in K_m \times \mathbb{R}^d$,

$$|R_k\varphi(z) - R_k\varphi(\hat{z})| \quad \leq \quad \left(2[\varphi]_{lip} + \|\varphi\|_{sup}\right) |\pi - \hat{\pi}|_1. \qquad (4.3.7)$$

Proof of Proposition 4.3.1. For $k = 1, \ldots, n$, we define the measurable functions on

$K_m \times \mathbb{R}^d$, resp. on Γ_k :

$$\begin{aligned} v_k(z) &= \varphi_k(z) \int \varphi_{k+1}(z_{k+1}) \dots \varphi_n(z_n) R_{k+1}(z, dz_{k+1}) \dots R_n(z_{n-1}, dz_n), \\ \hat{v}_k(z) &= \varphi_k(z) \int \varphi_{k+1}(z_{k+1}) \dots \varphi_n(z_n) \hat{r}_{k+1}(z, dz_{k+1}) \dots \hat{r}_n(z_{n-1}, dz_n), \end{aligned}$$

with the convention that for $k = n$, $v_n = \hat{v}_n = \varphi_n$, we then have the backward induction formulas :

$$v_k(z) = \varphi_k(z) R_{k+1} v_{k+1}(z) := \varphi_k(z) \mathbb{E}[v_{k+1}(Z_{k+1}) | Z_k = z] \tag{4.3.8}$$

$$\hat{v}_k(z) = \varphi_k(z) \hat{r}_{k+1} \hat{v}_{k+1}(z) := \varphi_k(z) \mathbb{E}[\hat{v}_{k+1}(\hat{Z}_{k+1}) | \hat{Z}_k = z], \tag{4.3.9}$$

for all $k = 1, \dots, n-1$.

Step 1. We clearly have $\|v_k\|_{sup} \leq 1$. Moreover, from (4.3.8) and using Lemma 4.3.1, we have :

$$\begin{aligned} [v_k]_{lip} &\leq [\varphi_k]_{lip} + [R_{k+1} v_{k+1}]_{lip} \\ &\leq 1 + \bar{L}_g + 2\bar{L}_g [v_{k+1}]_{lip}. \end{aligned}$$

Since $[v_n]_{lip} \leq 1$, a standard backward induction yields :

$$[v_k]_{lip} \leq \frac{\frac{3}{2}(2\bar{L}_g)^{n-k+1} - 1 - \bar{L}_g}{2\bar{L}_g - 1}, \tag{4.3.10}$$

for all $k = 0, \dots, n$.

Step 2. From (4.3.8)-(4.3.9), we may write :

$$\begin{aligned} \left\| v_k(Z_k) - \hat{v}_k(\hat{Z}_k) \right\|_1 &\leq \left\| v_k(Z_k) - \mathbb{E}[v_k(Z_k) | \hat{Z}_k] \right\|_1 \\ &\quad + \left\| \mathbb{E}\left[\left(\varphi_k(Z_k) - \varphi_k(\hat{Z}_k) \right) R_{k+1} v_{k+1}(Z_k) \middle| \hat{Z}_k \right] \right\|_1 \\ &\quad + \left\| \mathbb{E}\left[\varphi(\hat{Z}_k) \left(R_{k+1} v_{k+1}(Z_k) - \hat{r}_{k+1} \hat{v}_{k+1}(\hat{Z}_k) \right) \middle| \hat{Z}_k \right] \right\|_1 \\ &= I_1 + I_2 + I_3. \end{aligned} \tag{4.3.11}$$

By the law of iterated conditional expectation, we have :

$$\begin{aligned} I_1 &\leq \left\| v_k(Z_k) - v_k(\hat{Z}_k) \right\|_1 + \left\| \mathbb{E}[v_k(\hat{Z}_k) | \hat{Z}_k] - \mathbb{E}[v_k(Z_k) | \hat{Z}_k] \right\|_1 \\ &\leq 2 \left\| v_k(Z_k) - v_k(\hat{Z}_k) \right\|_1 \leq 2[v_k]_{lip} \left\| Z_k - \hat{Z}_k \right\|_1. \end{aligned}$$

Since conditional expectation (here with respect to $\hat{Z}_k$) is a L^1-contraction, and v_{k+1} is bounded by 1, we have :

$$I_2 \leq \left\| \varphi_k(Z_k) - \varphi_k(\hat{Z}_k) \right\|_1 \leq \left\| Z_k - \hat{Z}_k \right\|_1,$$

recalling that φ_k is in $BL_1(K_m \times \mathbb{R}^d)$. Since $\hat{Z}_k$ is $\sigma(Z_k)$-measurable, and recalling also that φ_k is bounded by 1, we have :

$$I_3 \leq \left\| v_{k+1}(Z_{k+1}) - \hat{v}_{k+1}(\hat{Z}_{k+1}) \right\|_1 .$$

Plugging these estimates of I_1, I_2 and I_3 into (4.3.11), we get

$$\left\| v_k(Z_k) - \hat{v}_k(\hat{Z}_k) \right\|_1 \leq \left(1 + 2[v_k]_{lip}\right) \left\| Z_k - \hat{Z}_k \right\|_1 + \left\| v_{k+1}(Z_{k+1}) - \hat{v}_{k+1}(\hat{Z}_{k+1}) \right\|_1 .$$

Since $\left\| v_n(Z_n) - \hat{v}_n(\hat{Z}_n) \right\|_1 \leq \left\| Z_n - \hat{Z}_n \right\|_1$, a direct backward induction yields :

$$\left\| v_k(Z_k) - \hat{v}_k(\hat{Z}_k) \right\|_1 \leq \sum_{j=k}^{n} \left(1 + 2[v_j]_{lip}\right) \left\| Z_j - \hat{Z}_j \right\|_1 .$$

The required result is proved by taking $k = 0$, substituting the estimate (4.3.10) and using Cauchy-Schwarz inequality : $\left\| Z_j - \hat{Z}_j \right\|_1 \leq \sqrt{m+d} \left\| Z_j - \hat{Z}_j \right\|_2$. □

4.3.3 Optimal quantization and rate of convergence

The estimation error (4.3.3) in Proposition 4.3.1 shows that to obtain the best approximation of the sequence of probability transitions (R_k) of the Markov chain $(Z_k)_k = (\Pi_k, Y_k)_k$ by this quantization approach, one has to minimize at each time $k \geq 1$ the L^2 quantization error $\|Z_k - \hat{Z}_k\|_2$. By identifying a grid $\Gamma_k = \{z^1, \ldots, z^{N_k}\}$ of size $|\Gamma_k| = N_k$ points in $K_m \times \mathbb{R}^d$, with the N_k-tuple $(z^1, \ldots, z^{N_k}) \in (K_m \times \mathbb{R}^d)^{N_k}$, the objective is then to minimize the symmetric function :

$$\begin{aligned} D_{N_k}^{Z_k}(z^1, \ldots, z^{N_k}) &= \left\| Z_k - \mathrm{Proj}_{\Gamma_k}(Z_k) \right\|_2^2 \\ &= \mathbb{E}\left[\min_{1 \leq i \leq N_k} \left| Z_k - z^i \right|_2^2 \right], \ \forall \Gamma_k = (z^1, \ldots, z^{N_k}) \in (K_m \times \mathbb{R}^d)^{N_k} \end{aligned} \quad (4.3.12)$$

which is the square of the L^2 quantization error and is usually called distorsion. The optimal quantization consists, for each $k \geq 1$ and given a number of points N_k, to find a grid $\hat{\Gamma}_k$ of size N_k that reaches the minimum of the distorsion function $D_{N_k}^{Z_k}$. This question has been tackled for a long time as part of quantization for information theory and signal processing, and more recently in probability for both numerical and theoretical purpose (see [25] or [40]). We recall these results and apply them in our context in the next paragraph. Now, we focus on the behaviour of the minimum of this function, i.e. the minimal L^2 quantization error, when the number of points N_k goes to infinity. For this, we first recall the so-called Zador theorem, see e.g. [25].

Theorem 4.3.1 *Let X be a $\mathbb{R}^l$-valued random variable with distribution $\mathbb{P}_X$, s.t. $\mathbb{E}|X|_2^{2+\delta} < \infty$ for some $\delta > 0$. Then*

$$\lim_N \left(N^{\frac{2}{l}} \min_{|\Gamma| \leq N} \|X - \mathrm{Proj}_\Gamma(X)\|_2^2 \right) \quad = \quad J_{2,l} \left(\int_{\mathbb{R}^l} |f|^{\frac{l}{l+2}}(x)\, dx \right)^{\frac{l+2}{l}} \tag{4.3.13}$$

where $\mathbb{P}_X(dx) = f(x)\,\lambda_l(dx) + \nu(dx)$ is the Lebesgue decomposition of $\mathbb{P}_X$ with respect to the Lebesgue measure λ_l on $\mathbb{R}^l$. The constant $J_{2,l}$ corresponds to the case where X is the uniform distribution on $[0,1]^l$.

Remark 4.3.2 In dimension $l = 1$ and 2, the values of $J_{2,l}$ are known : $J_{2,1} = 1/6$ and $J_{2,2} = \frac{5}{18\sqrt{3}}$. In higher dimension, the true value of $J_{2,l}$ is unknown but we have an equivalent $J_{2,l} \sim \frac{l}{2\pi e}$ as l goes to infinity.

Here, we cannot apply directly this theorem to Z_k since the distribution $\mathbb{P}_{Z_k}$ of (Π_k, Y_k) is not known in general, and in particular its decomposition with respect to the Lebesgue measure. However, one can prove the following error bound for the minimal distorsion error of Z_k.

Proposition 4.3.2 *For $k \geq 1$, assume that there exists some $\varepsilon > 0$ s.t. :*

$$\int |y_k|_2^{2+\varepsilon} \prod_{l=1}^{k} g_l(x_{l-1}, y_{l-1}, x_l, y_l) dy_l \quad < \quad \infty, \quad \forall\, x_0, \ldots, x_k \;\in\; E. \tag{4.3.14}$$

Then, we have :

$$\limsup_{N_k \to \infty} N_k^{\frac{2}{m-1+d}} \min_{|\Gamma_k| \leq N_k} \|Z_k - \mathrm{Proj}_{\Gamma_k}(Z_k)\|_2^2 \quad \leq \quad C(m, d, f_k),$$

where

$$C(m,d,f_k) \quad = \quad \frac{m(d+m-1)}{(md)^{\frac{d}{d+m-1}}} (J_{2,d})^{\frac{d}{d+m-1}} \left(\int_{\mathbb{R}^d} |f_k|^{\frac{d}{d+2}}(y)\, dy \right)^{\frac{d+2}{d+m-1}},$$

and f_k is the marginal density of Y_k given by :

$$f_k(y) \quad = \quad \int g_k(x_{k-1}, y_{k-1}, x_k, y) P_k(x_{k-1}, dx_k) \prod_{l=1}^{k-1} g_l(x_{l-1}, y_{l-1}, x_l, y_l) P_l(x_{l-1}, dx_l) dy_l \mu(dx_0).$$

Remark 4.3.3 In the case where the law of Y_k conditional on (X_{k-1}, Y_{k-1}, X_k) does not depend on Y_{k-1}, i.e. the function g_k does not depend on Y_{k-1}, the expression f_k of the marginal density of Y_k simplifies into :

$$f_k(y) \quad = \quad \mathbb{E}[g_k(X_{k-1}, X_k, y)], \quad y \in \mathbb{R}^d, \tag{4.3.15}$$

and condition (4.3.14) is written as :

$$\int |y_k|_2^{2+\delta} g_k(x_{k-1}, x_k, y_k) dy_k \;\; < \;\; \infty, \quad \forall\, x_{k-1}, x_k \;\in\; E,$$

for some $\delta > 0$.

Proof of Proposition 4.3.2. For any grids

$$\begin{aligned} \Gamma^{\Pi} &= \{\pi(1), \ldots, \pi(M)\} \quad \text{of size } |\Gamma^{\Pi}| = M \text{ points in } K_m, \\ \Gamma^{Y} &= \left\{y^1, \ldots, y^L\right\} \quad \text{of size } |\Gamma^{Y}| = L \text{ points in } \mathbb{R}^d, \end{aligned}$$

we denote

$$\Gamma^{\Pi} \otimes \Gamma^{Y} \;\; = \;\; \left\{(\pi(i), y^j) \;: 1 \le i \le M,\; 1 \le j \le L\right\} \quad \text{of size } \; ML \; \text{ points in } K_m \times \mathbb{R}^d.$$

We then have by definition of the norm $\|.\|_2$ and of the projection :

$$\left\|Z_k - \mathrm{Proj}_{\Gamma^{\Pi} \otimes \Gamma^{Y}}(Z_k)\right\|_2^2 \;\; = \;\; \|\Pi_k - \mathrm{Proj}_{\Gamma^{\Pi}}(\Pi_k)\|_2^2 + \|Y_k - \mathrm{Proj}_{\Gamma^{Y}}(Y_k)\|_2^2,$$

and so

$$\begin{aligned} & \min_{|\Gamma_k| \le N_k} \|Z_k - \mathrm{Proj}_{\Gamma_k}(Z_k)\|_2^2 \\ \le \;\; & \min_{M,L \,:ML \le N_k} \left(\min_{|\Gamma^{\Pi}| \le M} \|\Pi_k - \mathrm{Proj}_{\Gamma^{\Pi}}(\Pi_k)\|_2^2 + \min_{|\Gamma^{Y}| \le L} \|Y_k - \mathrm{Proj}_{\Gamma^{Y}}(Y_k)\|_2^2 \right) \end{aligned} \tag{4.3.16}$$

For M_0 in $\mathbb{N}\backslash\{0\}$, consider the grid Γ^{Π} of size $M = M_0^{m-1}$ points in K_m :

$$\Gamma^{\Pi} \;\; = \;\; \left\{ \left(\frac{i_1}{M_0}, \ldots, \frac{i_{m-1}}{M_0}, 1 - \sum_{l=1}^{m-1} \frac{i_l}{M_0} \right) \;: i_1, \ldots, i_{m-1} \;=\; 1, \ldots, M_0, \; \sum_{l=1}^{m-1} i_l \le M_0 \right\},$$

Denoting for all $a \in \mathbb{R}$, $[a]$ the smallest integer smaller than a, we have for all $\pi = (\pi^i)_{1 \le i \le m}$ $\in K_m$:

$$\begin{aligned} |\pi - \mathrm{Proj}_{\Gamma^{\Pi}}(\pi)|_2^2 \;\; &= \;\; \min_{\substack{i_1, \ldots, i_{m-1} = 1, \ldots, M_0 \\ i_1 + \ldots + i_{m-1} \le M_0}} \; \sum_{l=1}^{m-1} \left| \pi^l - \frac{i_l}{M_0} \right|^2 + \left| \sum_{l=1}^{m-1} \left(\pi^l - \frac{i_l}{M_0} \right) \right|^2 \\ &\le \;\; m \min_{\substack{i_1, \ldots, i_{m-1} = 1, \ldots, M_0 \\ i_1 + \ldots + i_{m-1} \le M_0}} \; \sum_{l=1}^{m-1} \left| \pi^l - \frac{i_l}{M_0} \right|^2 \\ &\le \;\; m \sum_{l=1}^{m-1} \left| \pi^l - \frac{[\pi^l M_0]}{M_0} \right|^2 \\ &\le \;\; \frac{m(m-1)}{M_0^2}. \end{aligned}$$

This shows that :

$$\min_{|\Gamma^{\Pi}|\leq M} \|\Pi_k - \mathrm{Proj}_{\Gamma^{\Pi}}(\Pi_k)\|_2^2 \quad \leq \quad \frac{m(m-1)}{M^{\frac{2}{m-1}}}. \tag{4.3.17}$$

On the other hand, notice that condition (4.3.14) ensures that $\mathbb{E}|Y_k|_2^{2+\varepsilon} < \infty$. Thus, from Theorem 4.3.1 applied to Y_k, we have for all $\delta > 0$ and L large enough,

$$\min_{|\Gamma^{Y}|\leq L} \|Y_k - \mathrm{Proj}_{\Gamma^{Y}}(Y_k)\|_2^2 \quad \leq \quad \frac{\left(J_{2,d}\|f_k\|_{\frac{d}{d+2}} + \delta\right)}{L^{\frac{2}{d}}}, \tag{4.3.18}$$

where we set :

$$\|f_k\|_{\frac{d}{d+2}} \quad = \quad \left(\int_{\mathbb{R}^d} |f_k|^{\frac{d}{d+2}}(y)\, dy\right)^{\frac{d+2}{d}}.$$

Substituting (4.3.17) and (4.3.18) into (4.3.16) yields :

$$\begin{aligned} &\min_{|\Gamma_k|\leq N_k} \|Z_k - \mathrm{Proj}_{\Gamma_k}(Z_k)\|_2^2 \\ \leq \quad &\min_{M,L\,:ML\leq N_k} \left[\frac{m(m-1)}{M^{\frac{2}{m-1}}} + \frac{\left(J_{2,d}\|f_k\|_{\frac{d}{d+2}} + \delta\right)}{L^{\frac{2}{d}}}\right]. \end{aligned}$$

We conclude with the elementary result that for all $a, b > 0$:

$$\min_{M,L\,:ML\leq N} \left[\frac{a}{M^{\frac{2}{l}}} + \frac{b}{L^{\frac{2}{d}}}\right] \quad = \quad (d+l)\left(\frac{a}{l}\right)^{\frac{l}{d+l}}\left(\frac{b}{d}\right)^{\frac{d}{d+l}} \frac{1}{N^{\frac{2}{d+l}}}.$$

□

4.3.4 Practical implementation of the optimal approximating filter process

We now come back to the numerical implementation of an algorithm that computes for each k :

- an optimal grid $\hat{\Gamma}_k$ which minimizes the distorsion :

$$D_{N_k}^{Z_k}(\Gamma_k) \quad = \quad \|Z_k - \mathrm{Proj}_{\Gamma_k}(Z_k)\|_2^2$$

as well as an estimation of this error,

- the weights of the Voronoi tesselations :

$$\begin{aligned} \hat{p}_k^i \quad &= \quad \mathbb{P}\left[Z_k \in C_i(\hat{\Gamma}_k)\right], \quad i = 1, \ldots, N_k, \\ \hat{\beta}_{k+1}^{ij} \quad &= \quad \mathbb{P}\left[Z_{k+1} \in C_j(\hat{\Gamma}_{k+1}),\ Z_k \in C_i(\hat{\Gamma}_k)\right], \quad i = 1, \ldots, N_k,\ j = 1, \ldots, N_{k+1}, \end{aligned}$$

and so the probability transition matrix $\hat{r}_{k+1}^{ij} = \hat{\beta}_{k+1}^{ij}/\hat{p}_k^i$.

This program is based on the following key property of the distorsion : The function $D_{N_k}^{Z_k}$ is continuously differentiable at any N_k-tuple $\Gamma_k = (z^1, \ldots, z^{N_k}) \in (K_m \times \mathbb{R}^d)^{N_k}$ having pairwise distinct components and its gradient is obtained by formal differentiation inside the expectation operator in (4.3.12) (see [40]) :

$$\nabla D_{N_k}^{Z_k}(\Gamma_k) \;=\; 2\,\mathbb{E}[H(\Gamma_k, Z_k)], \tag{4.3.19}$$

where the $(K_m \times \mathbb{R}^d)^{N_k}$-vector valued function H is given by :

$$H(\Gamma_k, z) \;=\; \big((z^i - z)1_{z \in C_i(\Gamma_k)}\big)_{1 \le i \le N_k}, \quad \Gamma_k = (z^1, \ldots, z^{N_k}) \in (K_m \times \mathbb{R}^d)^{N_k},\ z \in K_m \times \mathbb{R}^d.$$

This above integral representation for $\nabla D_{N_k}^{Z_k}$ suggests to implement a stochastic gradient descent, whenever one is able to simulate easily independent copies of Z_k. We will come back below on the simulation of Z_k. The stochastic gradient procedure is recursively defined by

$$\Gamma_k^{s+1} \;=\; \Gamma_k^s - \delta_{s+1}\, H(\Gamma_k^s, \xi_k^{s+1}) \tag{4.3.20}$$

where the initial grid Γ_k^0 has N_k pairwise distinct components, $(\xi_k^s)_{s \ge 1}$ is an i.i.d. sequence of $\mathbb{P}_{Z_k}$-distributed random vectors, and $(\delta_s)_{s \ge 1}$ is a sequence of step parameters satisfying the usual conditions :

$$\sum_s \delta_s \;=\; \infty \quad \text{and} \quad \sum_s \delta_s^2 \;<\; \infty.$$

In an abstract framework (see *e.g.* [16] or [32]), under some appropriate assumptions, a stochastic gradient descent associated to the integral representation of a so-called potential function ($D_{N_k}^{Z_k}$ in our problem) converges a.s., when s goes to infinity, toward a local minimum $\hat{\Gamma}_k$ of this potential function :

$$\nabla D_{N_k}^{Z_k}(\hat{\Gamma}_k) \;=\; 0.$$

Although these assumptions are not fulfilled by $D_{N_k}^{Z_k}$, the encountered theoretical problems can be partially overcome (see [40]) Practical implementation does provide satisfactory results (a commonly encountered situation with gradient descents). Moreover, computation of the weights of the tesselations and of the distorsion can be implemented as by-product of the procedure. We now describe this algorithm, known as the *Competitive Learning Vector Quantization* algorithm.

Simulation of the Markov chain $(Z_k)_k$:

We notice that from (4.2.4), we are able to simulate the probability transition R_k of the $(\mathbb{P}, \mathcal{F}_k^Y)$ Markov chain $(Z_k)_k = (\Pi_k, Y_k)_k$. For $k = 0$, recall that Z_0 is a known deterministic vector equal to $z_0 = (\mu, y_0)$. For $k \ge 1$, starting from (Π_{k-1}, Y_{k-1}),

- we simulate X_{k-1} with probability law Π_{k-1}, and then X_k according to the probability transition P_k.
- we simulate Y_k according to the probability transition $q_k(X_{k-1}, Y_{k-1}, X_k, dy')$.
- we compute Π_k by the formula (4.3.1) :

$$\Pi_k \;=\; \frac{GP(Y_{k-1}, Y_k)^\top \Pi_{k-1}}{|GP(Y_{k-1}, Y_k)^\top \Pi_{k-1}|_1}.$$

Subsequently, we stock S independent copies of the Markov chain $(Z_0, \ldots, Z_n)$, that we denote $\xi^s = (\xi_0^s, \ldots, \xi_n^s)$, $s = 1, \ldots, S$. The algorithm reads as follows :

Initialisation phase :

- Initialize the n grids $\Gamma_k^0 = (z_k^{0,1}, \ldots, z_k^{0,N_k}) \in (K_m \times \mathbb{R}^d)^{N_k}$ for $k = 0, \ldots, n$, with $\Gamma_0^0 = z_0$ reduced to $N_0 = 1$ point for $k = 0$.
- Initialize the weights vectors : $p_k^{0,i} = 1/N_k$, $\beta_{k+1}^{0,ij} = 0$, $i = 1, \ldots, N_k$, $j = 1, \ldots, N_{k+1}$, and the distorsion $D_{N_k}^0 = 0$, for $k = 0, \ldots, n$.

Updating $s \to s+1$: At step s, the n grids $\Gamma_k^s = (z_k^{s,1}, \ldots, z_k^{s,N_k})$, the weights vectors $p_k^{s,i}$, $\beta_{k+1}^{s,ij}$, $i = 1, \ldots, N_k$, $j = 1, \ldots, N_{k+1}$, the distorsion $D_{N_k}^s$ have been obtained and we use the sample ξ^{s+1} of $(Z_0, \ldots, Z_n)$ to update them as follows : for all $k = 0, \ldots, n$,

- *Competitive phase :* select $i_k(s+1) \in \{1, \ldots, N_k\}$ such that

$$\xi^{s+1} \;\in\; C_{i_k(s+1)}(\Gamma_k^s), \quad \text{i.e.} \quad i_k(s+1) \;\in\; \operatorname{argmin}_{1 \le i \le N_k} |z_k^{s,i} - \xi^{s+1}|_2.$$

- *Learning phase :*

  ⋆ Updating of the grid :

$$z_k^{s+1,i} \;=\; z_k^{s,i} - \delta_{s+1}\, 1_{i=i_k(s+1)} \left(z_k^{s,i} - \xi^{s+1}\right), \quad i = 1, \ldots, N_k$$

  ⋆ Updating of the weights vectors and of the probability transition

$$\begin{aligned} p_k^{s+1,i} &= p_k^{s,i} - \delta_{s+1} \left(p_k^{s,i} - 1_{i=i_k(s+1)}\right), \\ \beta_{k+1}^{s+1,ij} &= \beta_{k+1}^{s,ij} - \delta_{s+1} \left(\beta_{k+1}^{s,ij} - 1_{i=i_k(s+1), j=i_{k+1}(s+1)}\right), \\ r_{k+1}^{s+1,ij} &= \frac{\beta_{k+1}^{s+1,ij}}{p_k^{s+1,i}}, \end{aligned}$$

  for all $i = 1, \ldots, N_k$, $j = 1, \ldots, N_{k+1}$.

  ⋆ Updating of the distorsion

$$D_{N_k}^{s+1} \;=\; D_{N_k}^s - \delta_{s+1} \left(D_{N_k}^s - \left|z_k^{s,i_k(s+1)} - \xi^{s+1}\right|_2^2\right),$$

It is shown in [40] that on the event $\{\Gamma_k^s \to \hat{\Gamma}_k\}$, set of trajectories of $(\Gamma_k^s)_s$ that converge to $\hat{\Gamma}_k$ local minimum of the distorsion, we have :

$$\begin{aligned} p_k^{s,i} &\longrightarrow \hat{p}_k^i = \mathbb{P}\left[Z_k \in C_i(\hat{\Gamma}_k)\right], \quad a.s. \\ \beta_{k+1}^{s,ij} &\longrightarrow \hat{\beta}_{k+1}^{ij} = \mathbb{P}\left[Z_{k+1} \in C_j(\hat{\Gamma}_{k+1}),\ Z_k \in C_i(\hat{\Gamma}_k\right], \quad a.s. \\ D_{N_k}^s &\longrightarrow D_{N_k}^{Z_k}(\hat{\Gamma}_k), \quad a.s. \end{aligned}$$

for all $k = 0, \ldots, n$, $i = 1, \ldots, N_k$, $j = 1, \ldots, N_{k+1}$, as s goes to infinity.

4.4 Application : optimal stopping under partial observation

We consider the framework of Section 3. We denote by $\mathcal{T}_n^Y$ the set of stopping times adapted with respect to the observation filtration $(\mathcal{F}_k^Y)$ and valued in $\{0, \ldots, n\}$. Given two measurable functions f and h on $E \times \mathbb{R}^d$, we consider the following optimal stopping problem under partial observation :

$$u_0 = \sup_{\tau \in \mathcal{T}_n^Y} \mathbb{E}\left[\sum_{k=0}^{\tau} f(X_k, Y_k) + h(X_\tau, Y_\tau)\right]. \tag{4.4.1}$$

We shall transform this problem into an optimal stopping problem under complete observation. We denote

$$J(\tau) = \mathbb{E}\left[\sum_{k=0}^{\tau} f(X_k, Y_k) + h(X_\tau, Y_\tau)\right], \quad \tau \in \mathcal{T}_n^Y,$$

the expected gain function associated to (4.4.1). We shall also introduce the functions :

$$\begin{aligned} \hat{f}(\pi, y) &= \sum_{i=1}^{m} f(x^i, y)\pi^i, \quad \forall\, \pi = (\pi^i)_i \in K_m,\ \forall y \in \mathbb{R}^d, \\ \hat{h}(\pi, y) &= \sum_{i=1}^{m} h(x^i, y)\pi^i, \quad \forall\, \pi = (\pi^i)_i \in K_m,\ \forall y \in \mathbb{R}^d. \end{aligned}$$

Then, by using the law of iterated conditional expectations and the definition of the filter (Π_k), we have for all $\tau \in \mathcal{T}_n^Y$:

$$\begin{aligned} J(\tau) &= \mathbb{E}\left[\sum_{j=0}^{n} 1_{\tau=j}\left(\sum_{k=0}^{j} f(X_k,Y_k)+h(X_j,Y_j)\right)\right] \\ &= \mathbb{E}\left[\sum_{j=0}^{n} 1_{\tau=j}\left(\sum_{k=0}^{j} \mathbb{E}\left[f(X_k,Y_k)|\,\mathcal{F}_k^Y\right]+\mathbb{E}\left[h(X_j,Y_j)|\,\mathcal{F}_j^Y\right]\right)\right] \\ &= \mathbb{E}\left[\sum_{j=0}^{n} 1_{\tau=j}\left(\sum_{k=0}^{j} \hat{f}(\Pi_k,Y_k)+\hat{h}(\Pi_j,Y_j)\right)\right] \\ &= \mathbb{E}\left[\sum_{k=0}^{\tau} \hat{f}(Z_k)+\hat{h}(Z_\tau)\right]. \end{aligned}$$

Therefore, problem (4.4.1) may be rewritten as :

$$u_0 = \sup_{\tau\in\mathcal{T}_n^Y} \mathbb{E}\left[\sum_{k=0}^{\tau} \hat{f}(Z_k)+\hat{h}(Z_\tau)\right]. \tag{4.4.2}$$

Since the process $(Z_k)_k = (\Pi_k, Y_k)_k$ is a $(\mathbb{P}, (\mathcal{F}_k^Y))$ Markov chain, problem (4.4.2) is then an optimal stopping problem under complete observation.

By the dynamic programming principle, we have $u_0 = v_0(z_0)$, where the sequence of measurable functions $v_k : K_m \times \mathbb{R}^d \mapsto \mathbb{R}$, $k = 0,\ldots,n$, is given in recursive form by the backward formula :

$$v_n(z) = \hat{h}(z), \quad \forall z \in K_m \times \mathbb{R}^d \tag{4.4.3}$$

$$\begin{aligned} v_k(z) &= \max\left\{\hat{h}(z); \mathbb{E}\left[\hat{f}(Z_{k+1})+v_{k+1}(Z_{k+1})\Big|\, Z_k = z\right]\right\}, \\ &\qquad \forall z \in K_m \times \mathbb{R}^d, \;\; k = 0,\ldots,n-1. \end{aligned} \tag{4.4.4}$$

Applying the quantization approach described in Section 3 and setting :

$$\hat{Z}_k = \mathrm{Proj}_{\Gamma_k}(Z_k), \quad \Gamma_k = (z_k^1,\ldots,z_k^{N_k}) \in (K_m \times \mathbb{R}^d)^{N_k},\ k = 0,\ldots,n,$$

we then approximate, following [4], the sequence of functions (v_k) by the sequence of functions $\hat{v}_k : \Gamma_k \to \mathbb{R}$, $k = 0,\ldots,n$ defined by :

$$\hat{v}_n(z) = \hat{h}(z), \quad \forall z \in \Gamma_n$$

$$\begin{aligned} \hat{v}_k(z) &= \max\left\{\hat{h}(z); \mathbb{E}\left[\hat{f}(\hat{Z}_{k+1})+\hat{v}_{k+1}(\hat{Z}_{k+1})\Big|\, \hat{Z}_k = z\right]\right\}, \\ &\qquad \forall z \in \Gamma_k, \;\; k = 0,\ldots,n-1. \end{aligned}$$

From an algorithmic viewpoint, this reads as follows :

$$\begin{aligned}\hat{v}_n(z_n^i) &= \hat{h}(z_n^i), \quad i=1,\dots,N_n \\ \hat{v}_k(z_k^i) &= \max\left\{\hat{h}(z_k^i); \sum_{j=1}^{N_{k+1}} \hat{r}_{k+1}^{ij}\left(\hat{f}(z_{k+1}^j)+\hat{v}_{k+1}(z_{k+1}^j)\right)\right\}, \\ & \qquad i=1,\dots,N_k, \quad k=0,\dots,n-1.\end{aligned}$$

The optimal grids Γ_k and the associated probability transition matrix $(\hat{r}_{k+1}^{ij})$ are estimated follwing the procedure described in Section 3.

The induced approximation error is provided in the following theorem.

Theorem 4.4.2 *Assume that f and h are bounded and Lipschitz, uniformly in $x \in E$, i.e.*

$$\|f\|_{sup} := \sup_{x\in E, y\in\mathbb{R}^d} |f(x,y)| < \infty \quad \textit{and} \quad [f]_{lip} := \sup_{x\in E, y,\hat{y}\in\mathbb{R}^d, y\neq\hat{y}} \frac{|f(x,y)-f(x,\hat{y})|}{|y-\hat{y}|_1} < \infty.$$

Then under **(H1)**, *we have for all $k=0,\dots,n$:*

$$\begin{aligned}&\left\|v_k(Z_k)-\hat{v}_k(\hat{Z}_k)\right\|_1 \\ \leq\ & \sqrt{m+d}(\bar{f}+\bar{h})\sum_{j=k}^{n}\left[(7+2(n-j))\bar{L}_g+(n-j+2)\frac{(2\bar{L}_g)^{n-j+1}}{2\bar{L}_g-1}\right]\|Z_j-\hat{Z}_j\|_2, \qquad (4.4.5)\end{aligned}$$

where $\bar{f}=\max(\|f\|_{sup},[f]_{lip})$, $\bar{h}=\max(\|h\|_{sup},[h]_{lip})$.

Remark 4.4.4 In view of Proposition 4.3.2, the estimation (4.4.5) provides a rate of convergence for the approximation of $v_0(Z_0)$ of order

$$\frac{n(2\bar{L}_g)^n}{N^{\frac{1}{m-1+d}}},$$

when $N_k=N$ is the number of points at each grid Γ_k used for the optimal quantization of Z_k, $k=0,\dots,n$. The term $(2\bar{L}_g)^n$ is important when n is large, but this is consistent with the rate of convergence obtained in approximation of nonlinear filtering by quantization, see [41] or by Monte-Carlo particle methods, see [36].

Lemma 4.4.2 *Under the assumptions of Theorem 4.4.2, the functions v_k, $k=0,\dots,n$, defined in* (4.4.4), *are bounded and Lipschitz with :*

$$\| v_k \|_{sup} \leq (n-k+1)(\bar{f}+\bar{h}), \qquad (4.4.6)$$

$$[v_k]_{lip} \leq \frac{n-k+3}{2}\frac{(\bar{f}+\bar{h})}{2\bar{L}_g-1}(2\bar{L}_g)^{n-k+1}. \qquad (4.4.7)$$

Proof. From (4.4.4), we have $\| v_k \|_{sup} \leq \|\hat{h}\|_{sup} + \|\hat{f}\|_{sup} + \| v_{k+1} \|_{sup}$. Since $\| v_n \|_{sup} = \|\hat{h}\|_{sup}$, $\|\hat{f}\|_{sup} \leq \|f\|_{sup}$, $\|\hat{h}\|_{sup} \leq \|h\|_{sup}$, we obtain immediately the inequality (4.4.6) by induction.

On the other hand, relation (4.4.4) and Lemma 4.3.1 also show

$$\begin{aligned} [v_k]_{lip} &\leq [\hat{h}]_{lip} + [R_{k+1}(\hat{f} + v_{k+1})]_{lip} \\ &\leq [\hat{h}]_{lip} + \bar{L}_g \left(\|\hat{f}\|_{sup} + 2[\hat{f}]_{lip} + \|v_{k+1}\|_{sup} + 2[v_{k+1}]_{lip} \right). \end{aligned} \tag{4.4.8}$$

By definition of $\hat{f}$, we clearly have for all $z = (\pi, y)$ and $z' = (\pi', y')$ in $K^m \times \mathbb{R}^d$,

$$\begin{aligned} |\hat{f}(z) - \hat{f}(z')| &\leq \|f\|_{sup} |\pi - \pi'|_1 + [f]_{lip} |y - y'|_1 \\ &\leq \bar{f} |z - z'|_1, \end{aligned}$$

A similar inequality holds for $[h]_{lip}$ i.e. $[\hat{h}]_{lip} \leq \bar{h}$. Plugging into (4.4.8) and using (4.4.6) yields

$$\begin{aligned} [v_k]_{lip} &\leq \bar{h} + \bar{L}_g \left(3\bar{f} + (n-k)(\bar{f} + \bar{h}) \right) \\ &\quad + 2\bar{L}_g [v_{k+1}]_{lip}. \end{aligned} \tag{4.4.9}$$

Since $[v_n]_{lip} = [\hat{h}]_{lip}$, a straightforward induction gives (4.4.7). □

Proof of Theorem 4.4.2.
We set $\Phi_k(z) = \mathbb{E}[\hat{f}(Z_{k+1}) + v_{k+1}(Z_{k+1}) | Z_k = z]$ and $\hat{\Phi}_k(z) = \mathbb{E}[\hat{f}(\hat{Z}_{k+1}) + \hat{v}_{k+1}(\hat{Z}_{k+1}) | \hat{Z}_k = z]$. Then, for $k = 0, \ldots, n-1$,

$$\begin{aligned} \left\| v_k(Z_k) - \hat{v}_k(\hat{Z}_k) \right\|_1 &\leq \left\| \hat{h}(Z_k) - \hat{h}(\hat{Z}_k) \right\|_1 + \left\| \Phi_k(Z_k) - \hat{\Phi}_k(\hat{Z}_k) \right\|_1 \\ &\leq [\hat{h}]_{lip} \|Z_k - \hat{Z}_k\|_1 + \left\| \Phi_k(Z_k) - \Phi_k(\hat{Z}_k) \right\|_1 \qquad (4.4.10) \\ &\quad + \left\| \mathbb{E}[\Phi_k(\hat{Z}_k) | \hat{Z}_k] - \mathbb{E}[\Phi_k(Z_k) | \hat{Z}_k] \right\|_1 + \left\| \mathbb{E}[\Phi_k(Z_k) | \hat{Z}_k] - \hat{\Phi}_k(\hat{Z}_k) \right\|_1 \\ &\leq [\hat{h}]_{lip} \|Z_k - \hat{Z}_k\|_1 + 2 \left\| \Phi_k(Z_k) - \Phi_k(\hat{Z}_k) \right\|_1 \\ &\quad + \left\| \mathbb{E}[\Phi_k(Z_k) | \hat{Z}_k] - \hat{\Phi}_k(\hat{Z}_k) \right\|_1, \qquad (4.4.11) \end{aligned}$$

by the law of iterated conditional expectation. Since $\hat{Z}_k$ is $\sigma(Z_k)$-measurable, we have

$$\begin{aligned} & \left\| \mathbb{E}[\Phi_k(Z_k) | \hat{Z}_k] - \hat{\Phi}_k(\hat{Z}_k) \right\|_1 \\ = & \left\| \mathbb{E}[\hat{f}(Z_{k+1}) + v_{k+1}(Z_{k+1}) | \hat{Z}_k] - \mathbb{E}[\hat{f}(\hat{Z}_{k+1}) + \hat{v}_{k+1}(\hat{Z}_{k+1}) | \hat{Z}_k] \right\|_1 \\ \leq & [\hat{f}]_{lip} \left\| Z_{k+1} - \hat{Z}_{k+1} \right\|_1 + \left\| v_{k+1}(Z_{k+1}) - \hat{v}_{k+1}(\hat{Z}_{k+1}) \right\|_1. \end{aligned}$$

Plugging into (4.4.11) yields :

$$\begin{aligned}\left\|v_k(Z_k)-\hat{v}_k(\hat{Z}_k)\right\|_1 &\leq \left([\hat{h}]_{lip}+2[\Phi_k]_{lip}\right)\|Z_k-\hat{Z}_k\|_1+[\hat{f}]_{lip}\left\|Z_{k+1}-\hat{Z}_{k+1}\right\|_1\\ &\quad+\left\|v_{k+1}(Z_{k+1})-\hat{v}_{k+1}(\hat{Z}_{k+1})\right\|_1.\end{aligned}$$

Since $\left\|v_n(Z_n)-\hat{v}_n(\hat{Z}_n)\right\|_1 \leq [\hat{h}]_{lip}\|Z_n-\hat{Z}_n\|_1$, a direct induction gives :

$$\left\|v_k(Z_k)-\hat{v}_k(\hat{Z}_k)\right\|_1 \leq \sum_{j=k}^{n} a_j\|Z_j-\hat{Z}_j\|_1 \tag{4.4.12}$$

where

$$a_j = \begin{cases} [\hat{h}]_{lip}+2[\Phi_k]_{lip}, & j=k \\ [\hat{h}]_{lip}+2[\Phi_j]_{lip}+[\hat{f}]_{lip}\ , & j=k+1,\ldots n-1 \\ [\hat{h}]_{lip}\ , & j=n \end{cases} \tag{4.4.13}$$

Now, by Lemmata 4.3.1 and 4.4.2, we have

$$\begin{aligned}[\Phi_k]_{lip} &= [R_{k+1}(\hat{f}+v_{k+1})]_{lip}\\ &\leq \bar{L}_g\left(\|\hat{f}\|_{sup}+2[\hat{f}]_{lip}+\|v_{k+1}\|_{sup}+2[v_{k+1}]_{lip}\right)\\ &\leq \bar{L}_g[3\bar{f}+(n-k)(\bar{f}+\bar{h})]\\ &\quad+(\bar{f}+\bar{h})(\frac{n-k}{2}+1)\frac{(2\bar{L}_g)^{n-k+1}}{2\bar{L}_g-1}.\end{aligned}$$

Substituting into (4.4.13) and (4.4.12) provides the required result by using also Cauchy-Schwarz inequality.

4.5 Numerical illustration : Bermudean options in a partially observed stochastic volatility model

We consider an observable risky asset price (S_k) with dynamics given by :

$$S_{k+1} = S_k\exp\left[\left(r-\frac{1}{2}X_k^2\right)\delta+X_k\sqrt{\delta}\varepsilon_{k+1}\right],\quad S_0 = s_0>0,$$

where (ε_k) is a sequence of Gaussian white noise, and (X_k) is the unobservable volatility process. $\delta>0$ may represent some discretization time step. Equivalently, we observe the process $(Y_k)=(\ln S_k)$, and we notice that the conditional law of Y_{k+1} given (X_k,Y_k) has a density given by :

$$g(X_k,Y_k,y') = \frac{1}{\sqrt{2\pi X_k^2\delta}}\exp\left[-\frac{\left(y'-Y_k-(r-\frac{1}{2}X_k^2)\delta\right)^2}{2X_k^2\delta}\right],\quad y'\in\mathbb{R}.$$

We model here the dynamics of (X, S) under some risk neutral martingale measure $\mathbb{P}$, r representing in this case the riskless interest rate.

We assume that (X_k) is an homogeneous Markov chain taking three possible values $x^b < x^m < x^h$ in $\mathbb{R}_+\backslash\{0\}$. Its probability transition matrix is given by :

$$P_k = \begin{pmatrix} 1-(p_{bm}+p_{bh})\delta & p_{bm}\delta & p_{bh}\delta \\ p_{mb}\delta & 1-(p_{mb}+p_{mh})\delta & p_{mh}\delta \\ p_{hb}\delta & p_{hm}\delta & 1-(p_{hb}+p_{hm})\delta \end{pmatrix}. \quad (4.5.1)$$

In this context of a partially observed stochastic volatility model, we consider a Bermudean put option with payoff :

$$h(y) = (\kappa - e^y)_+, \quad y \in \mathbb{R},$$

and we want to compute its price given by :

$$u_0 = \sup_{\tau \in \mathcal{T}_n^Y} \mathbb{E}\left[e^{-r\tau\delta} h(Y_\tau)\right]. \quad (4.5.2)$$

We consider a model where the volatility (X_k) is a Markov-chain approximation à la Kushner (see [31]) of a mean-reverting process :

$$dX_t = \lambda(x_0 - X_t)dt + \eta dW_t.$$

Denoting by $\Delta > 0$ the spatial step, this corresponds to a probability transition matrix of the form (4.5.1) with :

$$x^b = x_0 - \Delta, \quad x^m = x_0, \quad x^h = x_0 + \Delta,$$

and

$$\begin{aligned} p_{bm} &= \lambda + \frac{\eta^2}{2\Delta^2}, & p_{bh} &= 0 \\ p_{mb} &= \frac{\eta^2}{2\Delta^2}, & p_{mh} &= \frac{\eta^2}{2\Delta^2} \\ p_{hb} &= 0, & p_{hm} &= \lambda + \frac{\eta^2}{2\Delta^2}. \end{aligned}$$

In order to ensure that P_k is indeed a probability transition matrix, we have the consistency conditions :

$$1 - \left(\lambda + \frac{\eta^2}{2\Delta^2}\right)\delta \geq 0 \quad \text{and} \quad 1 - \frac{\eta^2}{\Delta^2}\,\delta \geq 0$$

We perform numerical tests with

- Price and put option parameters : $r = 0.05$, $S_0 = 110$, $\kappa = 100$,
- Volatility parameters : $\lambda = 1$, $\eta = 0,1$, $X_0 = 0.15$,

- Spatial step : $\Delta = 0,05$.
- Quantization : Grids are of same size N fixed for each time period with step $\delta = \frac{1}{n}$.

We first compare in Table 4.1 the filter expectation at the final date $n = ...$ computed by the optimal quantization method with increasing grid size N , and by 10^6 Monte Carlo iterations.

	$E[\Pi_n^1]$	$E[\Pi_n^2]$	$E[\Pi_n^3]$	Relative error (%)
Monte Carlo	0.287608	0.422833	0.289558	
Quant. with $N = 300$	0.301651	0.421725	0.276624	0.898
Quant. with $N = 600$	0.301604	0.421458	0.276938	0.886
Quant. with $N = 900$	0.301598	0.421316	0.277086	0.881
Quant. with $N = 1200$	0.301618	0.42122	0.277162	0.879
Quant. with $N = 1500$	0.301605	0.421205	0.27719	0.878

Table 4.1: Comparison of quantized filter value to its Monte Carlo estimation

We see that besides the very low error level, we can observe error and relative error decreasing as the grid sizes grow.

Secondly, in order to illustrate the effect of the time step, we compute the American option price under partial observation when the time step δ decreases to zero (i.e. n increases) and compare it with the American option price with complete observation of (X_k, Y_k). Indeed, in the limit for $\delta \to 0$ we fully observe the volatility, and so the partial observation price should converge to the complete observation price.

Moreover, when we have more and more observations, the difference between the two prices should decrease and converge to zero. This is shown in figure 4.2, where we performed option pricing over grids of size $N_{\Pi,Y} = 1500$ in case of partial observation. The total observation price is given by the same pricing algorithm carried out on $N_{X,Y} = 45$ point product grids of (X_k, Y_k). We have seen in Remark 4.4.4 that for fixed n, the rate of convergence for the approximation of the value function under partial observation is of order $N_{\Pi,Y}^{1/(m-1+d)}$ where $N_{\Pi,Y}$ is the number of points used at each time k, for the grid of (Π_k, Y_k) valued in $K^m \times \mathbb{R}^d$. From results of [4], we also know that the rate of convergence for the approximation of the value function under full observation is of order $m \times N_Y$ where $N_{X,Y} = m \times N_Y$ is the number of points at each time k, used for the grid of (X_k, Y_k) valued in $E \times \mathbb{R}^d$. This explains why, in order to have comparable results, and with $m = 3$ and $d = 1$, we have chosen $N_Y \sim N_{\Pi,Y}^{1/3}$.

In addition, it is possible to observe the effect of information enrichment as the time step decreases. In fact, if we consider multiples of n as the time step parameter, we notice that the American option price increases for both total and partial observation models (see tables 4.2 and 4.3).

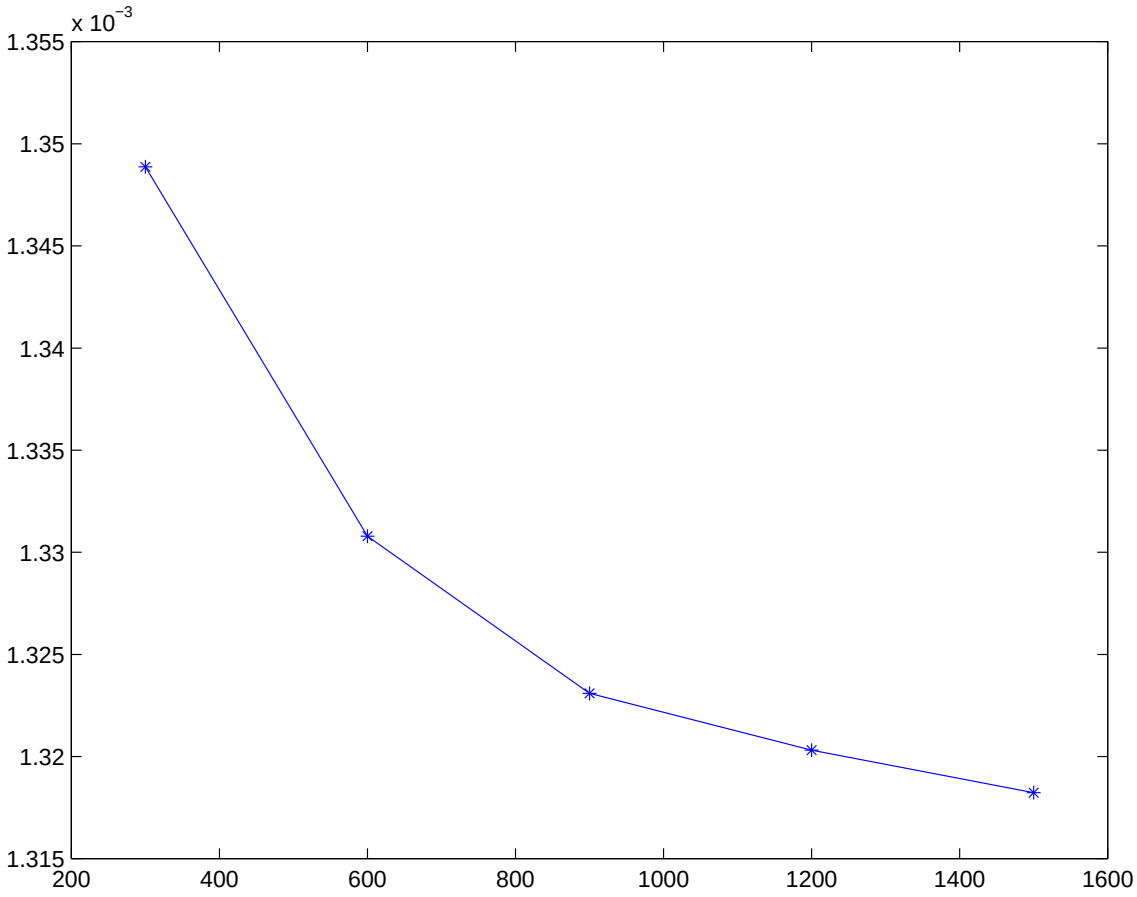

Figure 4.1: Filter error convergence as N grows

n	4	8	16
Tot. Obs. ($N_{X,Y} = 30$)	1.45863	1.75689	1.77642
Part. Obs. ($N_{\Pi,Y} = 1000$)	0.921729	1.13898	1.47089
Variation	0.53	0.61	0.30

Table 4.2: American option price for embedded filtrations - First Example

n	5	10	20
Tot. Obs. ($N_{X,Y} = 45$)	1.57506	1.72595	1.91208
Part. Obs. ($N_{\Pi,Y} = 1500$)	0.988531	1.30616	1.59632
Variation	0.58	0.42	0.31

Table 4.3: American option price for embedded filtrations - Second Example

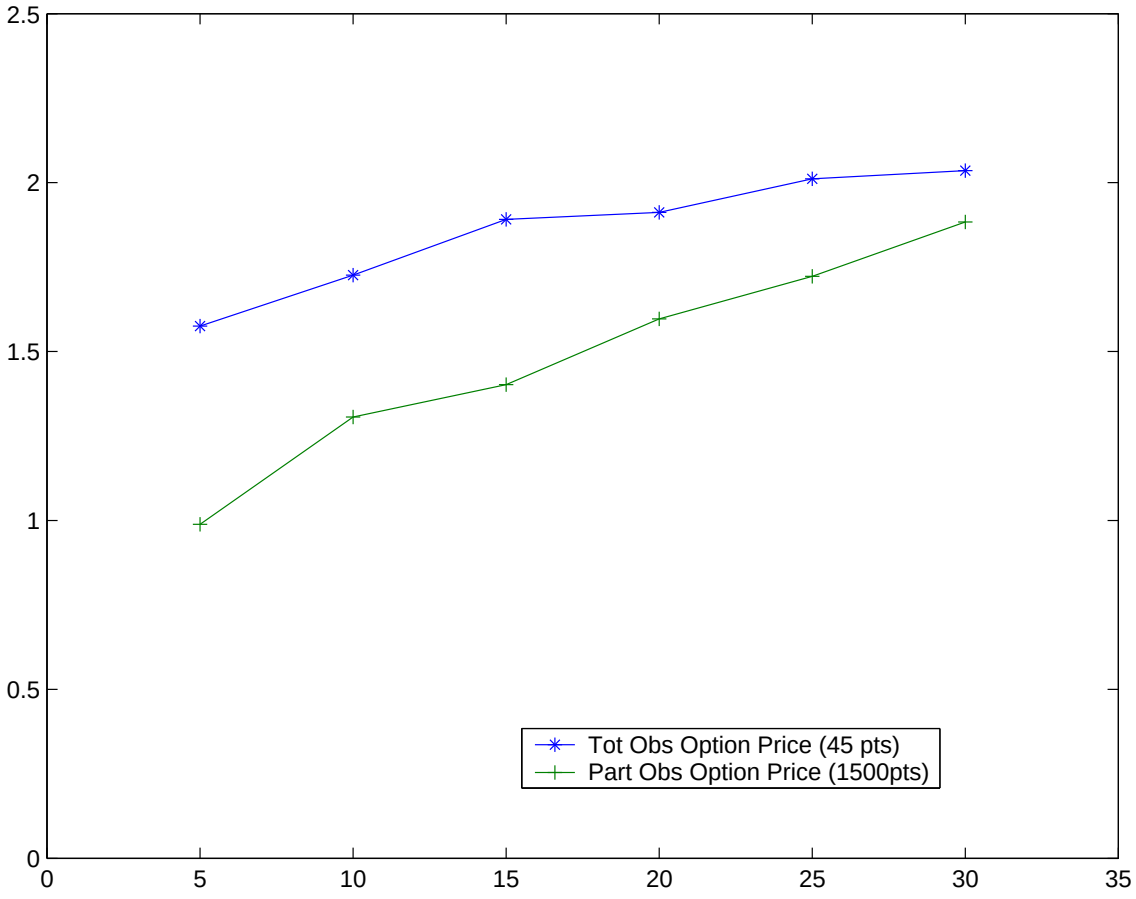

Figure 4.2: Partial and total observation option prices as $\delta \to 0$

Table des figures

Liste des tableaux

List of Algorithms

Bibliographie

[1] S. Arulampalam, T. Clapp, N. Gordon, and S. Maskall. A tutorial on particle filters for On-line Non-linear/Non Gaussian Bayesian tracking. QinetiQ Ltd, DSTO, IEEE, 2001.

[2] J. G. Attali. *1. Méthodes de stabilité pour des chaînes de Markov non félleriennes 2. Sur quelques autres problèmes issus des réseaux de neurones.* PhD thesis, Université de Paris I Panthéon Sorbonne, 1999.

[3] V. Bally, M.P. Bavouzet, and M. Messaoud. Integration by parts formula for locally smooth laws and applications to sensitivity computations. Rapport de recherche de l'INRIA : RR-5567, 2005.

[4] V. Bally and G. Pagès. A quantization algorithm for solving discrete time multi-dimensional optimal stopping problems. *Bernoulli*, 9 :1003–1049, 2003.

[5] V. Bally, G. Pagès, and J. Printems. A stochastic quantization method for non linear problems. *Monte Carlo methods and applications*, 7(1-2) :21–34, 2001.

[6] V. Bally, G. Pagès, and J. Printems. First order schemes in the numerical quantization method. *Mathematical finance*, 13(1) :1–16, 2003.

[7] V. Bally, G. Pagès, and J. Printems. A quantization tree method for pricing and hedging multidimensional american options. *Mathematical Finance*, 15(1) :119–168, 2005.

[8] V.E Beneš. Exact finite dimensional filters for certain diffusions with non linear drift. *Stochastics*, 5 :65–92, 1981.

[9] A. Bensoussan and W. Runggaldier. An approximation method for stochastic control problems with partial observation of the state : a method for constructing ε-optimal controls. *Acta Appli. Math.*, 10 :145–170, 1987.

[10] D. Brigo, B. Hanzon, and F. Legland. A Differential Geometric Approach to Nonlinear Filtering : the Projection Filter. *IEEE Transactions on Automatic Control*, 43(2) :247–252, 1998.

[11] D. Brigo and F. Legland. A finite dimensional filter with exponential conditional density. In *Proceedings of the 36th IEEE Conference on Decision and Control*, pages 1643–1644, 1997.

[12] M. Chaleyat-Maurel and V. Genon-Catalot. Computable inifite dimensional filters with applications to discretized diffusion processes. Preprint of Laboratoire de Probabilités et Modèles Aléatoires - PMA-989, 2005.

[13] M. Chaleyat-Maurel and D. Michel. Des résultats de non existence de filtre de dimension finie. *Stochastics*, 13(1-2) :83–102, 1984.

[14] D. Crisan and A. Doucet. A Survey of Convergence Results on Particle Filtering Methods for Practionners. *IEEE Transactions on signal processing*, 50 :736–746, 2002.

[15] A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo methods in Practice*. Springer, 1st edition, 2001.

[16] M. Duflo. *Random Iterative Models*. Springer-Verlag, 1997.

[17] J. C. Fort and G. Pagès. Asymptotics of optimal quantizers for some scalar distributions. *Journal of Computational and Applied Mathematics*, 146 :253–275, 2002.

[18] V. Genon-Catalot. A non linear explicit filter. *Statist. and Prob. letters*, 61 :145–154, 2003.

[19] V. Genon-Catalot and M. Kessler. Random scale perturbation of an AR(1) process and its properties as a nonlinear explicit filter. *Bernoulli*, 10(4) :701–720, 2004.

[20] A. Gersho and R.M. Gray. *Vector Quantization and Signal Compression*. The Kluwer International Series in Engineering and Computer Science. Springer, 2000.

[21] F. Le Gland. *Introduction au filtrage en temps discret. Filtre de Kalman, Modèles de Markov cachés*. IRISA/INRIA, 2002-2003.

[22] F. Le Gland. Filtrage particulaire. In *Actes du 19ème Colloque GRETSI sur le Traitement du Signal et des Images*, volume 2, pages 1–8, 2003.

[23] F. Le Gland and N. Oudjane. Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters. *The Annals of Applied Probability*, 4(1) :144–187, 2004.

[24] N. Gordon, D.J. Salmond, and A.F.M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Proceedings*, 140(2) :107–113, April 1993.

[25] S. Graf and H. Luschgy. *Foundations of quantization for probability distributions*. Lecture Notes in Mathematics. Springer, 2000.

[26] S. Ben Hamida and M. Mrad. Optimal Quantization using Evolutionary Algorithms. CMAP Preprint 564, Ecole Polytechnique-France, 2005.

[27] G. Kallianpur and C. Striebel. Estimation of stochastic systems : Arbitrary system process with additive white noise observation errors. *Ann. Math. Statist.*, 39(2) :785–801, 1968.

[28] J. Kieffer. Exponential rate of convergence for the Lloyd's method I. *IEEE on Information Theory, Special issue on Quantization*, 28(2) :205–210, 1982.

[29] G. Kitagawa. Non-Gaussian state-space modeling of nonstationary time series. *Journal of the American statistical association*, 82(400) :1032–1063, 1987.

[30] G. Kitagawa. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5(1) :1–25, 1996.

[31] H.J. Kushner and P.G. Dupuis. *Numerical Methods for Stochastic Control Problems in Continuous Time*. Springer, 1992.

[32] H.J. Kushner and G.G. Yin. *Stochastic Approximation Algorithms and Applications*. Springer, 1997.

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

[34] T.J. Lyons and T.S. Salisbury, editors. *Numerical Methods and Stochastics*, USA, 1999. American Mathematical Society.

[35] G.B. Di Masi. and W. Runggaldier. An approach to discrete-time stochastic control problems under partial observation. *SIAM J. Control & Optimiz.*, 25 :38–48, 1987.

[36] P. Del Moral, J. Jacod, and P. Protter. The Monte Carlo method for filtering with discrete-time observations. *Probab. Theory and Relat. Fields*, 120(2) :346–368, June 2001.

[37] N. Newton. Observations preprocessing and quantization for nonlinear filters. *SIAM Journal of Control and Optimization*, 38(2) :482–502, 2000.

[38] N. Newton. Approximations for nonlinear filters based on quantization. *Monte Carlo methods and applications*, 7 :311–320, 2001.

[39] N. Oudjane. *Stabilité et approximations particulaires en filtrage non linéaire : Application au pistage*. PhD thesis, Université de Rennes, 2000.

[40] G. Pagès. A space vector quantization method for numerical integration. *SIAM J. Control & Optimiz.*, 89 :1–38, 1997.

[41] G. Pagès and H. Pham. Optimal quantization methods for nonlinear filtering with discrete-time observations. *Bernoulli*, 11(5) :893–932, 2005.

[42] G. Pagès, H. Pham, and J. Printems. An optimal Markovian quantization algorithm for multidimensional stochastic problems. *Stochastics and Dynamics*, 4 :501–545, 2004.

[43] G. Pagès, H. Pham, and J. Printems. Optimal quantization methods and applications to numerical problems in finance. In *Handbook of Computational and Numerical Methods in Finance*. S.T. Rachev, Birkhauser, Boston, 2004.

[44] G. Pagès and J. Printems. Optimal quadratic quantization for numerics : the Gaussian case. *Monte Carlo methods and applications*, 9(2) :135–168, 2003.

[45] G. Pagès and J. Printems. Functional quantization for pricing derivatives. Preprint of Laboratoire de Probabilités et Modèles Aléatoires PMA-930, 2004.

[46] E. Pardoux. *Filtrage non linéaire et équations aux dérivées partielles stochastiques associées*. LNM 1468 - Ecole d'été de probabilités de Saint Flour, 1989.

[47] H. Pham, W. Runggaldier, and A. Sellami. Approximation by quantization of the filter process and applications to optimal stopping problems under partial observation. *Monte Carlo methods and applications*, 11 :57–81, 2005.

[48] J. Picard. Efficiency of the Extended Kalman Filter. *SIAM Journal on Applied Mathematics*, 51(3) :843–885, 1991.

[49] C. P. Robert and G. Casella. *Monte Carlo statistical methods*. Springer texts in statistics. Springer, 2nd edition, 1999.

[50] W. Runggaldier and F. Spizzichino. Sufficient conditions for finite dimensionality of filters in discrete time : a Laplace transform-based approach. *Bernoulli*, 7(2) :211–221, 2001.

[51] S. Rubenthaler. *Méthodes de Monte Carlo en filtrage non linéaire et pour certaines équations différentielles stochastiques*. PhD thesis, Université Paris VI, 2002.

[52] G. Sawitzki. Finite dimensional filter systems in discrete time. *Stochastics*, 5 :107–114, 1981.

[53] A. Sellami. Comparative survey on non linear filtering methods : the quantization and the particle filtering approaches. Preprint of Laboratoire de Probabilités et Modèles Aléatoires. PMA-1021, 2005.

[54] A. Sellami. Quantization based filtering method using first order approximation. Preprint of Laboratoire de Probabilités et Modèles Aléatoires. PMA-1009, 2005.

[55] L. Stettner W. Runggaldier. On the construction of nearly optimal strategies for a general problem of control of partially observed diffusions. *Stochastics and Stochastics Reports*, 37 :15–47, 1991.

Résumé

Nous développons une approche de résolution numérique du filtrage par méthode de grille, en utilisant des résultats de quantification optimale de variables aléatoires. Nous mettons en oeuvre deux algorithmes de calcul de filtres utilisant les techniques d'approximation du type ordre 0 et ordre 1. Nous proposons les versions implémentables de ces algorithmes et étudions le comportement de l'erreur des approximations en fonction de la taille des quantifieurs en s'appuyant sur la propriété de stationnarité des quantifieurs optimaux. Nous positionnons cette approche par grille par rapport à l'approche particulaire du type Monte Carlo à travers la comparaison des deux méthodes et leur expérimentation sur différents modèles d'états. Dans une seconde partie, nous nous intéressons à l'avantage qu'offre la quantification pour le prétraitement des données offline pour développer un algorithme de filtrage par quantification des observations (et du signal). L'erreur est là aussi étudiée et un taux de convergence est établi en fonction de la taille des quantifieurs. Enfin, la quantification du filtre en tant que variable aléatoire est étudiée dans le but de la résolution d'un problème d'évaluation d'option américaine dans un marché à volatilité stochastique non observée. Tous les résultats sont illustrés à travers des exemples numériques.
Mots clés : Filtrage, quantification optimale, quantifieur stationnaire, filtres particulaires, Monte Carlo, prétraitement offline, option américaine, volatilité stochastique.

Abstract

We develop a grid based numerical approach to solve a filtering problem, using results on optimal quantization of random variables. We construct two filtering algorithms using zero order and first order approximation techniques. We suggest implementable versions of these algorithms and study the approximation error behavior by considering the stationnary property of optimal quantizers. The grid approach is then compared to the particle one based on Monte Carlo methods. The study is done over a set of different state models. In a second part, we have been interested in the advantadge given by quantization methods to preprocess offline the information. This permitted to develop a filtering algorithm based on observation (and signal) quantization. Here also the error convergence rate to zero as the quantizer size goes to infinity is studied. Finally, the quantization of the filter as a random variable is studied in order to solve a problem of pricing an American option in an unobserved stochastic volatility market. All results are illustrated by numerical experiments.
Keywords : Filtering, optimal quantization, stationary quantizer, particle method, Monte Carlo, offline preprocessing, American option, stochastic volatility.

Vu : le Président
M. .....

Vu : les suffrageants
MM. .....

Vu et permis d'imprimer :
Le Vice-Président du Conseil Scientifique chargé de la Recherche de l'Université PARIS IX DAUPHINE.