

# Chapitre introductif et présentation du cadre général de la thèse

Afef SELLAMI

Laboratoire de Probabilités et

Modèles Aléatoires

CNRS, UMR 7599

and Université Paris Dauphine

sellami@math.jussieu.fr

September 1, 2005

## 1 Quantification optimale et applications

La quantification est un procédé communément utilisé en traitement du signal et de l'information. Il consiste à approcher un signal à valeurs dans un espace continu par un signal à valeur dans un espace discret. À l'origine, cette technique a été motivée par des raisons pratiques de transmission [7], ensuite elle a vu ses applications s'élargir pour toucher différents domaines pratiques. Dans cette section nous allons rappeler quelques résultats théoriques utiles pour la quantification et s'attarder sur ses applications les plus récentes.

### 1.1 Définitions et résultats préliminaires

#### 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 supposée connue  $\mathbb{P}_X$ . Un entier  $N \in \mathbb{N}^*$  étant fixé, un  $N$ -quantifieur est une application borélienne  $h_N : \mathbb{R}^d \rightarrow \mathbb{R}^d$  appliquant  $\mathbb{R}^d$  dans un ensemble fini  $\Gamma = \{x^1, \dots, x^N\} \subset \mathbb{R}^d$ .

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

- $\Gamma = \{x^1, \dots, x^N\}$  la grille de quantification de taille  $N$  appelée aussi ensemble des centres ou points de quantification, ou encore  $N$ -quantifieur associé à  $h_N$ .

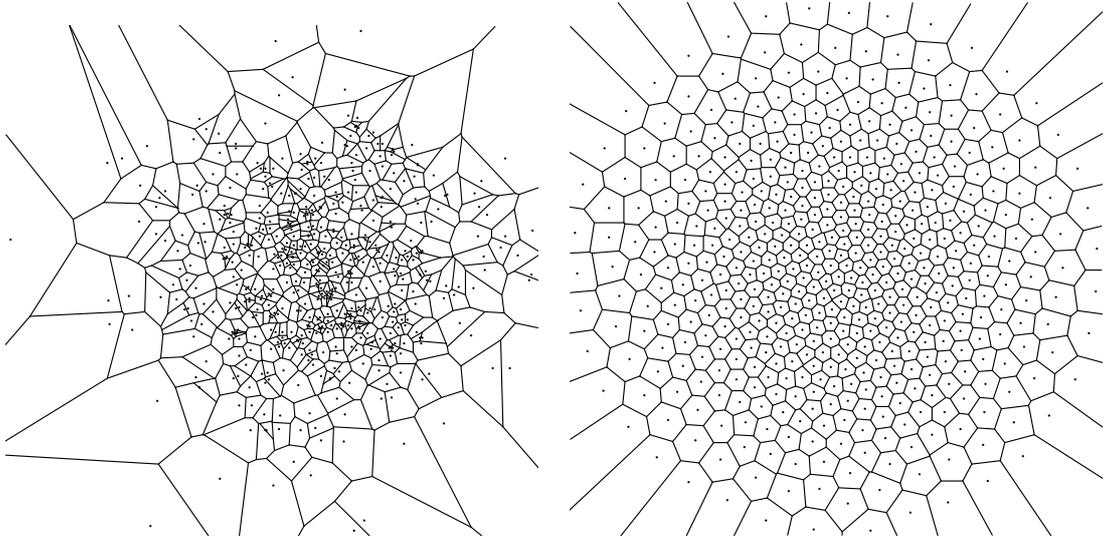


Figure 1: Partitions de l'espace associées à une grille de quantification en deux dimensions

- 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  $x^i \in A^i$ .

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  $N$ :

$$\inf\{\|X - h(X)\|_p^p, h : \mathbb{R}^d \rightarrow \mathbb{R}^d, \text{ application borelienne t.q. } |h(\mathbb{R}^d)| \leq N\}. \quad (1.1)$$

D'après les résultats établis par Graf et Luschgy [10], ce problème admet une solution et sa détermination implique d'une part la détermination d'une grille de quantification optimale qui vérifiera donc:

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

et d'autre part, d'une partition optimale  $(A^i)_{1 \leq i \leq N}$  qui vérifiera:

$$\mathbb{E}|X - h_N^*(X)|^p = \sum_{i=1}^N \inf_{x \in \mathbb{R}^d} \mathbb{E}[|X - x|^p | X \in A^i] \mathbb{P}(X \in A^i).$$

Il est aussi établi que l'optimalité du quantifieur passe par le choix de la partition connue sous le nom de la partition de Voronoï  $(C_i(\Gamma^*))_{1 \leq i \leq N}$  définissant  $h_N^*$  comme une projection au plus proche voisin sur l'ensemble des centres  $(x^i)_{1 \leq i \leq N}$  (voir Figure 1.1.1). Soit:

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

L'erreur de quantification ou distorsion s'écrit alors:

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

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

**Theorem 1.1** (Cf. [10, 1]) *On suppose que  $\int_{\mathbb{R}^d} |\xi|^{p+\eta} \mathbb{P}_X(d\xi) < +\infty$  pour  $\eta > 0$ . Alors,*

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

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

Nous pouvons ainsi écrire que  $\|X - h_N^*(X)\|_p = O(N^{-\frac{1}{d}})$  au voisinage de  $+\infty$ .

Par ailleurs, il sera utile de noter que les quantifieurs  $L^2$ -optimaux vérifient une propriété dite de stationnarité, à savoir que:

$$\mathbb{E}[X|h_N^*(X)] = h_N^*(X). \quad (1.2)$$

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 [2] ou de filtrage.

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. Des méthodes de résolution numérique peuvent cependant être utilisées pour trouver des solutions approchées au problème, on citera par exemple la méthode du point fixe ou celle du gradient, qui deviennent difficiles à mettre en oeuvre en dimensions supérieures à 1. Quelques autres méthodes numériques peuvent être adoptées dans des cas particuliers de lois à quantifier. Enfin, les algorithmes stochastiques d'apprentissage (*Competitive Learning Vector Quantization* CLVQ algorithm) offrent une bonne alternative aux problèmes posés par les méthodes précédentes. Nous ne détaillerons pas plus cet aspect de la quantification car il n'est pas au centre de notre travail. Nous nous intéresserons plutôt à l'exploitation des quantifieurs une fois définis pour différentes applications pratiques. A cet effet, il est important d'introduire tout d'abord une définition de processus quantifié. Ce sera l'objet du prochain paragraphe.

### 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 connaît la dynamique d'évolution. Une approche possible dans ce cas c'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  $X_k \in L^p$  qu'on notera  $\Gamma_k = \{x_k^1, \dots, x_k^{N_k}\}$ .

Par conséquent, on définit:

$$\hat{X}_k = \sum_{i=1}^{N_k} x_k^i \mathbf{1}_{\mathbf{C}_i(\Gamma_k)}(X_k). \quad (1.3)$$

Le processus 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  $i \in \{1, \dots, N_k\}$  et  $j \in \{1, \dots, N_{k+1}\}$ :

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

D'une manière générale, pour  $0 \leq k < n$  et  $i \in \{1, \dots, N_k\}$ , on notera:

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

Pour des horizons  $n$  pas trop élevés, il est donc possible de calculer et de stocker dans des tables facilement accessibles les grilles de quantifications et les paramètres compagnons. Ce pre-traitement des données, dit off line, permet de minimiser les calculs d'éventuels estimateurs utilisant la quantification.

## 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 d'évaluation de l'intégrale  $\mathbb{E}[f(X)]$ , pour  $X$  de loi connue absolument continue par rapport à la mesure de Lebesgue, de densité  $\mathbf{p}$ . 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 l'estimateur se résumera à une somme pondérée finie. En reprenant les notations du paragraphe précédent, on pose:

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

Les pondérations  $\hat{\mathbf{p}}^i$  sont aussi des paramètres compagnon qui peuvent être calculés en même temps que la grille de quantification  $\Gamma$  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  $f \in \mathcal{C}_b^1$ , il existe  $\xi \in (X, \hat{X})$  tel que:

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

Ce qui donne la majoration d'erreur qu'on appellera *d'ordre zero*:

$$|\mathbb{E}[f(X)] - \mathbb{E}[f(\hat{X})]| \leq C \|\Delta\|_1 \leq C \|\Delta\|_2 \quad (1.4)$$

Quand  $f \in \mathcal{C}_b^2$ , 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:

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

Ainsi, comme  $\hat{X}$  vérifie la propriété de stationnarité (1.2), il est possible d'établir:

$$\begin{aligned} |\mathbb{E}[f(X)] - \mathbb{E}[f(\hat{X})]| &\leq \mathbb{E}|\mathbb{E}[f(X) - f(\hat{X})|\hat{X}]| \\ &\leq C\mathbb{E}|\langle \Delta, \Delta \rangle| \leq C\|\Delta\|_2^2 \end{aligned} \quad (1.5)$$

En utilisant le théorème de Zador 1.1, on obtient un taux de convergence en  $O(N^{-\frac{1}{d}})$  dans le cas de l'inégalité (1.4), et moyennant une 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 (1.5).

L'intégration numérique par quantification est une méthode qui s'approche dans son principe des méthodes de Monte Carlo : 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 somme échantillonnée d'un nombre fini  $M$  d'échantillons iid:

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

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 off line, 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 on line 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.
- Le taux de convergence des estimateurs de Monte Carlo est en  $O(N^{-\frac{1}{2}})$ ; il est indépendant de la dimension. Pour les méthodes de quantification, la convergence, bien que dépendant de la dimension, reste intéressante en dimension 2 si elle est de type *ordre 0* et en dimension 4 si elle est de type *ordre 1*.

Il est enfin intéressant de voir qu'il est possible d'appliquer un raisonnement d'échantillonnage préférentiel dans l'intégration numérique par quantification.

**Definition 1.1 *Echantillonnage Préférentiel:*** C'est la procédure par laquelle on approche de manière empirique une mesure de probabilité difficile à simuler  $\nu$  en utilisant un échantillon iid  $(\xi_N, \dots, \xi_1)$  d'une autre mesure de probabilité  $\mu$  dite loi d'importance, plus facile à simuler. S'il existe une fonction  $m$  et une constante  $\bar{m}$  vérifiant:

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

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  $\mathbb{E}[f(X)]$ .

$$\begin{aligned} \mathbb{E}[f(X)] &= \mathbb{E}\left[f(Y) \frac{\mathbf{p}(Y)}{\mathbf{q}(Y)}\right] \approx \mathbb{E}\left[f(\hat{Y}) \frac{\mathbf{p}(\hat{Y})}{\mathbf{q}(\hat{Y})}\right], \\ \|\mathbb{E}[f(X)] - \mathbb{E}\left[f(\hat{Y}) \frac{\mathbf{p}(\hat{Y})}{\mathbf{q}(\hat{Y})}\right]\| &\leq C \|Y - \hat{Y}\|_2. \end{aligned} \tag{1.6}$$

### 1.3 Autres applications

La quantification a surtout 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. 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 part 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 les travaux de Pagès et Bally [3], où ils proposent un algorithme d'évaluation d'options américaines dans le cadre d'un modèle de Black et Scholes et estiment le temps d'arrêt optimal, la date d'exercice de l'option. Comme pour le problème d'intégration numérique, l'utilisation de quantifieurs stationnaires permet dans cette application d'améliorer l'estimation numérique à travers le passage à un ordre supérieur de convergence (cf Pagès et Bally [2]). Par ailleurs, Pagès et Pham [15] 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 portefeuille. Dans ces applications, la résolution numérique rétrograde est rendue possible grâce aux grilles de quantifications 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 par Pagès et Pham dans [14], 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 à celles 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 serait intéressant d'étudier le passage à des schémas numériques de type ordre 1, comme il a été suggéré dans [16] où l'utilisation de la propriété de stationnarité permet de définir des schémas récursifs d'ordre de convergence supérieur mais malheureusement pas implémentables. La question est donc:

- Comment définir un schéma numérique de premier ordre implémentable et quel serait l'impact alors sur l'amélioration prévue du taux de convergence des estimations de filtres?

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 caractéristique des modèles d'actifs à volatilité stochastique. 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 [3, 2] ou d'optimisation de portefeuille [15] 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 coeur des développements de cette thèse, nous y consacrerons toute la section qui suivra.

## 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 et dont on n'observe que des états bruités. Le filtrage s'inscrit dans une approche Bayésienne 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 connaît 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.

## 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:

$$\begin{cases} X_k = F_k(X_{k-1}, \varepsilon_k), & X_0 \text{ de loi } \mu_0 \text{ connue à priori,} \\ Y_k = G_k(X_k, \eta_k), & k \geq 1. \end{cases} \quad (2.7)$$

où  $\varepsilon_k$  et  $\eta_k$  désignent des variables aléatoires connues, indépendantes entre elles et indépendantes de  $X_0$ .  $\varepsilon_k$  modélise l'innovation du processus caché,  $\eta_k$  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  $Y_{0:n} = (Y_0, \dots, Y_n)$  est donnée par l'espérance conditionnelle:

$$\mathbb{E}[X_n | Y_{0:n}].$$

De manière plus générale, l'information la plus riche disponible à travers les observations  $Y_{0:n}$  est donnée par la loi conditionnelle  $\mathcal{L}(X_n | Y_{0:n})$ . 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:

$$\Pi_{y,n}(dx) = \mathbb{P}[X_n \in (dx) | Y_0 = y_0, \dots, Y_n = y_n],$$

ou

$$\Pi_{y,n}f = \mathbb{E}[f(X_n) | Y_0 = y_0, \dots, Y_n = y_n].$$

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

$$\Pi_{Y,n}(dx) = \mathbb{P}[X_n \in (dx) | Y_0, \dots, Y_n] \quad \text{ou} \quad \Pi_{Y,n}f = \mathbb{E}[f(X_n) | Y_0, \dots, Y_n].$$

### 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  $F_k : \mathbb{R}^d \times \mathbb{R}^q \rightarrow \mathbb{R}^d$ , est une fonction Borélienne et  $(\varepsilon_k)_{1 \leq k \leq n}$  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  $\mu_0$  de  $X_0$  est supposée connue à priori. Par ailleurs, on désigne par  $\mathbf{P}_k(x, dx')$  le noyau de transition de  $X_k$  à  $X_{k+1}$ , et on notera 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  $G_k : \mathbb{R}^{d'} \times \mathbb{R}^d \times \mathbb{R}^{q'} \rightarrow \mathbb{R}^{d'}$  est une fonction borélienne et  $(\eta_k)$  est une suite de variables aléatoires iid à valeurs dans  $\mathbb{R}^{q'}$  et indépendantes 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:

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

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

- Le processus  $(X_k, Y_k)$  est une chaîne 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  $\eta_k$  et  $X_{k-1}$  et  $\eta_{k-1}$ . Par conséquent, la loi jointe  $\mathcal{L}(X_0, \dots, X_n, Y_0, \dots, Y_n)$  s'écrit:

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

et par la formule de Bayes on établit la formule de Kallianpur Striebel [11]:

$$\mathbb{E}[f(X_{0:n})|Y_{1:n} = y_{1:n}] = \frac{\int \dots \int f(x_{0:n})\mu_0(dx_0) \prod_{k=1}^n g_k(x_k, y_k)\mathbf{P}_k(x_{k-1}, dx_k)}{\int \dots \int \mu_0(dx_0) \prod_{k=1}^n g_k(x_k, y_k)\mathbf{P}_k(x_{k-1}, dx_k)} \quad (2.8)$$

Plus particulièrement, on définit le filtre évalué à 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, on confondra la densité conditionnelle  $g_k$  avec la fonction de vraisemblance associée:

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

la dépendance en  $y_k$  sera implicite.

### 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.

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

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

$$\Pi_{k+1|k}(dx') = \int \Pi_k(dx) \mathbf{P}_k(x, dx'). \quad (2.9)$$

**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:

$$\Pi_{k+1}(dx) = \frac{g_{k+1}(x) \Pi_{k+1|k}(dx)}{\int g_{k+1}(x) \Pi_{k+1|k}(dx)}. \quad (2.10)$$

### 2.1.3 Le modèle d'états discret

La formulation récursive du calcul du filtre par les equations (2.9) et (2.10) rend la résolution du problème numériquement évidente dans le cadre d'un signal à espace d'états discret. En effet, si pour tout  $0 \leq k \leq n$ ,  $X_k(\Omega) = \{x_k^1, \dots, x_k^{N_k}\}$  et si  $\mathbf{P}_k$  désigne la matrice de transition de la chaîne de Markov signal, d'après le paragraphe précédent, le filtre est calculé séquentiellement en introduisant les opérateurs matriciels  $H_{y,k}$ :

$$\begin{aligned} H_{y,k}f(x_{k-1}^i) &= \mathbb{E}[f(X_k)g_k(X_k, y_k) | X_{k-1} = x_{k-1}^i], \\ &= \sum_{j=1}^{N_k} f(x_k^j) \mathbf{P}_{k-1}^{ij} g_k(x_k^j, y_k). \end{aligned}$$

Ainsi,

$$\begin{aligned} \pi_{y,0}f &= \mathbb{E}[f(X_0)], \\ \pi_{y,k}f &= \pi_{y,k-1}H_{y,k}f. \end{aligned}$$

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:

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

Finalement, en normalisant:

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

#### 2.1.4 Filtrage de Kalman

Quand on sort du cadre discret précédent, l'évaluation exacte du filtre en utilisant les équations (2.9) et (2.10) devient plus délicate, car elle implique le calcul successif d'intégrales. Le modèle d'état appelé de Kalman Bucy constitue un des rares modèles à espace d'états continu où une formulation explicite du filtre est possible. On considère:

$$\begin{cases} X_k = \rho_k X_{k-1} + \theta_k \varepsilon_{k+1}, & X_0 \sim \mathcal{N}(m_0, \Sigma_0), \\ Y_k = X_k + \alpha_k \eta_k, \\ \varepsilon_k \text{ et } \eta_k \text{ iid } \sim \mathcal{N}(0, I_d), \\ \rho_k, \theta_k, \alpha_k \in \mathcal{M}_d(\mathbb{R}). \end{cases} \quad (2.11)$$

Dans ce cas particulier, il est établi que le filtre  $\Pi_k$  ainsi que la prédiction  $\Pi_{k+1|k}$  sont gaussiens de lois respectives  $\mathcal{N}(m_k, \Sigma_k)$  et  $\mathcal{N}(m_{k+1|k}, \Sigma_{k+1|k})$ . Les paramètres  $m_k$ ,  $\Sigma_k$ ,  $m_{k+1|k}$  et  $\Sigma_{k+1|k}$  sont connus récursivement par l'algorithme suivant (Cf. [8]),  $k = 1, \dots, n$ :

$$\begin{aligned} m_{k+1|k} &= \rho_k m_k, \\ \Sigma_{k+1|k} &= \rho_k \Sigma_k \rho_k' + \theta_k \theta_k', \\ m_k &= m_{k|k-1} + K_k (Y_k - m_{k|k-1}), \\ \Sigma_k &= (I - K_k) \Sigma_{k|k-1}, \\ K_k &= \Sigma_{k|k-1} (\Sigma_{k|k-1} + \alpha_k \alpha_k')^{-1}. \end{aligned} \quad (2.12)$$

## 2.2 Les méthodes d'approximation

Dans un cadre plus général, le modèle d'états est non linéaire ou non gaussien et dans la majorité de ces cas (Cf [6, 12]), une solution explicite n'est pas disponible. On parle alors de problème de filtrage non linéaire. Plusieurs méthodes numériques d'approximation ont été introduites pour résoudre de tels problèmes. Bien que d'approches différentes, elles s'appuient toutes sur le principe de trouver une représentation fini-dimensionnelle de la loi objectif  $\Pi_k$ . Dans ce qui suit, nous présentons succinctement trois méthodes d'approximation numériques.

### 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:

$$\begin{cases} X_{k+1} = F_k(X_k, \varepsilon_{k+1}) \\ Y_k = G_k(X_k) + \alpha_k \eta_k \end{cases} \quad (2.13)$$

Pour ce système, le processus solution  $(\Pi_k)$  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 récursif du type (2.12). La loi du filtre  $\Pi_k$  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:

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

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

$$\begin{aligned} m_{k+1|k} &= F_k(m_k, 0), \\ \Sigma_{k+1|k} &= D_x F_k(m_k, 0)\Sigma_k D_x F_k(m_k, 0)' + D_\varepsilon F_k(m_k, 0)D_\varepsilon F_k(m_k, 0)', \\ m_k &= m_{k|k-1} + K_k (Y_k - G_k(m_{k|k-1})), \\ \Sigma_k &= (I - K_k DG_k(m_{k|k-1})) \Sigma_{k|k-1}, \\ K_k &= \Sigma_{k|k-1} DG_k(m_{k|k-1})' (DG_k(m_{k|k-1})\Sigma_{k|k-1} DG_k(m_{k|k-1})' + \alpha_k \alpha_k')^{-1}. \end{aligned} \quad (2.14)$$

### 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 définir ainsi des 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  $\hat{X}_k(\Omega) = \Gamma_k = \{x_k^1, \dots, x_k^{N_k}\}$ . On peut alors approcher le filtre récursivement en définissant l'estimateur  $\hat{\pi}_{y,n}$  par l'algorithme inspiré du modèle d'états discret ([17, 14]):

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

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:

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

Le choix du quantifieur, notamment de la grille  $\Gamma_k$  et de la partition associée constituent un point crucial dans la qualité de l'estimation. D'après [14] en effet, un choix de quantifieur  $L^2$ -optimal nous permet d'établir un taux de convergence vers zero via le théorème de Zador. Ceci est possible à travers l'inéquation de contrôle de l'erreur sur le filtre par l'erreur de quantification:

**Theorem 2.2** *Supposons  $\mathbf{P}_k$  est Lipschitzien et  $f$  is bornée Lipschitzienne continue, alors il existe une suite positive de constntes  $(C_j^n)_{0 \leq j \leq n}$  telles que:*

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

### 2.2.3 Méthodes particulières

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 générées par des simulations iid selon  $\mu_0$ . 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:

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

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 (2.9) et (2.10). En effet, étant donné un système de particules  $(X_k^i)_{1 \leq i \leq M}$  iid selon  $\mathcal{L}(X_{0:k})$ , les échantillons  $(X_{k+1}^i)_{1 \leq i \leq M}$  simulés selon la transition  $\mathbf{P}_k$  sont iid selon  $\mathcal{L}(X_{0:k+1})$  et selon (2.9) 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 (2.10) 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 du problème de dégénérescence des pondérations [17, 13]. Pour y remédier, la solution 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 [9, 5]. Ce type d'algorithmes avec rééchantillonnage séquentiel (**SIR** ou *Sequential Importance Resampling*) multiplie les particules à fortes pondérations et élimine les autres. Différents types de convergence de ces méthodes sont établis dans [4, 5]. Il est cependant utile de mentionner que cette solution peut s'avérer insuffisante dans plusieurs cas pratiques, dans la mesure où elle pose le problème de la dégénérescence des particules. Ce problème apparaît quand la population des particules s'appauvrit et se concentre sur un seul point.

### 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 peut être considéré sous deux angles de vue différents. D'une part l'approfondissement des méthodes de filtrage par quantification déjà consruites pas Pagès et Pham dans [14], en utilisant une approche de développement au premier ordre; et la compraison numérique de l'approche par grille à l'approche Monte Carlo des filtres particuliers à travers plusieurs modèles d'états. D'autre part, nous nous sommes intéressés à l'évaluation même du filtre par quantification soit à travers la quantification des observations ou la quantification du filtre même. Le but dans la première approche était d'élaborer un algorithme de calcul plus rapide, tout en établissant une majoration 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 inobservée...

### References

- [1] V. Bally and G. Pagès. A quantization algorithm for solving discrete time multi-dimensional optimal stopping problems. *Bernoulli*, 9:1003–1049, 2003.
- [2] V. Bally, G. Pagès, and J. Printems. First order schemes in the numerical quantization method. *Mathematical finance*, 13(1):1–16, 2003.

- [3] 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.
- [4] D. Crisan and A. Doucet. A Survey of Convergence Results on Particle Filtering Methods for Practitioners. *IEEE Transactions on signal processing*, 50:736–746, 2002.
- [5] A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo methods for filtering*. Springer, 1 edition, 1996.
- [6] V. Genon-Catalot. A non linear explicit filter. *Statist. and Prob. letters*, 61:145–154, 2003.
- [7] Allen Gersho and Robert M. Gray. *Vector Quantization and Signal Compression*. The Kluwer International Series in Engineering and Computer Science. Springer, 2000.
- [8] F. Le Gland. *Introduction au filtrage en temps discret. Filtre de Kalman, Modèles de Markov cachés*. IRISA/INRIA, 2002-2003.
- [9] 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.
- [10] S. Graf and H. Luschgy. *Foundations of quantization for probability distributions*. Lecture Notes in Mathematics. Springer, 2000.
- [11] 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.
- [12] T.J Lyons and T.S Salisbury, editors. *Numerical Methods and Stochastics*, USA, 1999. American Mathematical Society.
- [13] N. Oudjane. *Stabilité et approximations particulières en filtrage non linéaire : Application au pistage*. PhD thesis, Université de Rennes, 2000.
- [14] G. Pagès and H. Pham. Optimal quantization methods for nonlinear filtering with discrete time observations. To appear in *Bernoulli*, 2003.
- [15] G. Pagès, H. Pham, and J. Printems. An optimal Markovian quantization algorithm for multidimensional stochastic problems. *Stochastics and Dynamics*, 4:501–545, 2004.
- [16] 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.
- [17] 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.