数理解析研究所講究録 1462 巻 2006 年 79-99

Quantization Methods in Filtering and Applications to Partially Observed Stochastic Volatility Models

Huyên PHAM
Laboratoire de Probabilités et
Modèles Aléatoires
CNRS, UMR 7599
Université Paris 7
e-mail: pham@math.jussieu.fr
and CREST

October 12, 2005

Abstract

We present some recent developments on optimal quantization methods for discrete-time nonlinear filtering. We analyse first the quantization algorithm for the filter given a fixed observation, and then the quantization of the filter process. This last study is motivated by dynamic optimization problems under partial information arising for example in finance in the pricing of American options under partially observed stochastic volatility models. Several numerical illustrations are carried out, emphasizing the convergence and the stability of the approximate filter.

Key words : Nonlinear filtering, Markov chain, quantization, stochastic gradient descent, optimal stopping.

MSC Classification (2000) : 60G35, 65C20, 65N50, 60G40.

1 Introduction

Optimal quantization of random vectors consists in finding the best approximation in L^p-norm of a random vector by a discrete random vector taking at most N values. This was originally developed in the 50's in the context of information theory where the basic motivation was to transmit efficiently a continuous stationary signal by means of a finite number of codes (or quantizers) . More recently, the quantization approach was applied to various fields, and notably to numerical probability, where it appears as an efficient spatial discretization method for solving multi-dimensional problems arising typically in finance.

In this article, we present recent developments of optimal quantization methods applied to the nonlinear filtering problem. This is formally the situation where we face a stochastic system whose evolution is governed by a hidden process that we observe only through some noise. Filtering is a traditional problem in probability and statistics, and occurs in particular in a financial context where we can observe stock price but not its volatility. Mathematically, the problem is to recover the optimal filter, i.e. the conditional law of the hidden process (the signal) given the past observations. For instance, standard filtering problems in finance are the estimation of the law of the volatility given the past price observations, and then the pricing of derivatives and portfolio optimization in a partially observed stochastic volatility model. Except some very specific cases like the Kalman-Bucy model, there is no explicit solution for the filter and one has to resort on numerical methods. The various approaches proposed in the literature (particle methods, grid methods) rely basically on the principle of finding a finite-dimensinal representation of the filter. We present here the quantization approach for filtering introduced in [7], which is a grid method, and where one searchs for grids that fit optimally according to L^2-norm to the distribution of the signal. In these numerical methods, the filter is approximated for a given fixed set of observations. However, in many applications arising in dynamic optimization under partial observation, one need to approximate the filter process where randomness is due to past observation process. We then present a quantization approach of the filter process introduced in [9] and give a numerical application to the problem of optimal stopping under partial observation.

The paper is organized as follows. Section 2 formulates and recalls some preliminaries on the filtering problem. In Section 3, we present the quantization method for approximating the filter given a fixed observation. Section 4 illustrates the results with numerical experiments. In Section 5, we introduce the quantization approach for the approximation of the filter process, and we deal in the last Section 6 with a numerical application to the optimal stopping problem in a partially observed stochastic volatility model.

2 Filtering problem for discrete observations

2.1 General framework

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 $\mathbb{R}^q$, such that the pair (X_k, Y_k) is a Markov chain on the probability space $(\Omega, \mathbb{P})$ 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 (called sometimes local likelihood function) :

$$y' \longmapsto g_k(X_{k-1}, Y_{k-1}, X_k, y').$$

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 $\mathcal{F}_k^Y$:

$$\Pi_k(dx) = \mathbb{P}\left[X_k \in dx | \mathcal{F}_k^Y\right], \quad k \in \mathbb{N}.$$

A typical case of signal-observation model is given by the following scheme :

$$\begin{aligned} X_k &= F_k(X_{k-1}, \varepsilon_k), \\ Y_k &= G_k(X_{k-1}, Y_{k-1}, X_k, \eta_k), \end{aligned}$$

for some measurable functions F_k, G_k, and where (ε_k) and (η_k) are two white noises. For example in finance, $(X_k)_k$ is the unobservable return and/or volatility of stock price S while $Y_k = \ln S_k$ is the logarithm of the observed price process

$$Y_k = Y_{k-1} + b(X_{k-1}) + \sigma(X_{k-1})\eta_k,$$

and g_k is explicit once the density of the white noise η_k is specified.

2.2 Filter evolution

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. From Markov property and Bayes formula, the filter process Π_k valued in $\mathcal{P}(E)$ satisfies the filtering forward equation :

$$\begin{aligned} \Pi_0 &= \mu, \\ \Pi_k(dx') &= \frac{\Pi_{k-1}H_k(dx')}{\Pi_{k-1}H_k(E)} := \frac{\int_E \Pi_{k-1}(dx)H_k(x, dx')}{\int_E \Pi_{k-1}H_k(dx')} \end{aligned} \tag{2.1}$$

where H_k is the prediction-updating transition kernel given by :

$$H_k(x, dx') = g_k(x, Y_{k-1}, x', Y_k)P_k(x, dx').$$

Hence, the calculation from Π_{k-1} to Π_k is done in two steps : a first step of prediction, which uses the a priori information from the signal transition probability P_k, and a second step of

correction and updating, which uses the a posteriori information given by the observation at time k via the local likelihood function g_k. We denote the relation (2.1) (which is nonlinear due the normalization) from Π_{k-1} to Π_k by :

$$\Pi_k = \bar{G}_k(\Pi_{k-1}, Y_{k-1}, Y_k).$$

Given a fixed set of observations, the filtering problem consists in solving or simulating by approximation this filtering equation valued in the infinite dimensional space $\mathcal{P}(E)$ (when the state space E is continuous). We distinguish essentially three types of methods :

- Analytical methods where one tries to solve analytically the forward equation. This is explicitly possible when X and Y are Gaussian processes leading to the well-known Kalman filter, which is a finite-dimensional filter completely characterized by its mean and variance. Some extensions are derived with the so-called extended Kalman filter. We also cite recent work in [3], which derives an explicit filter of infinite dimension.

- Monte-Carlo methods : this approach consists basically in approximating Π_k by a sequence of empirical measures associated to $\bar{N}$ interacting particles at each time k and simulate according to the prediction/updating mechanism given by the transition kernel H_k.

- Grid methods : this consists in approximating the transition kernel H_k by a transition matrix $\hat{H}_k$, which leads in turn to an approximate forward equation valued in a finite dimensional space.

We develop in the two next sections optimal quantization methods that belong to grid methods. It is also of interest to approximate not only the filter for a given set of observations, but more generally the filter process viewed as a random measure function of the uncertainty of the observation process. This will be developed in the two last sections where a quantization approach is introduced for approximating the filter process with applications to partial observation problems.

3 Approximate filter by optimal quantization (Fixed observation)

3.1 Short background on optimal vector quantization

The basic idea of (quadratic) quantization is to replace an $\mathbb{R}^d$-valued random vector $X \in L^2(\mathbb{P}, \mathbb{R}^d)$, with probability law $\mathbb{P}_X$, by a random vector taking at most N values in order to minimize the induced L^2-error. For this, consider a grid $x = \{x^1, \ldots, x^N\}$ of N points in $\mathbb{R}^d$ (we shall often identify such a grid with a N-tuple in $\mathbb{R}^d$), and its Voronoi tesselations, that is Borel partitions $C_1(x)$, ..., $C_N(x)$ of $\mathbb{R}^d$ satisfying :

$$C_i(x) \subset \left\{ u \in \mathbb{R}^d : |u - x^i| = \min_{j=1,\ldots,N} |u - x^j| \right\}, \quad i = 1, \ldots, N.$$

Then, one defines the x-Voronoi quantization of X as the closest neighbour projection of X on the grid x :

$$\hat{X}^x = \mathrm{Proj}_x(X) := \sum_{i=1}^{N} x^i 1_{C_i(x)}(X),$$

whose discrete probability law $\mathbb{P}_{\hat{X}}$ is characterized by :

$$\hat{p}_i := \mathbb{P}_{\hat{X}}(x^i) = \mathbb{P}_X(C_i(x)), \quad i = 1, \ldots, N.$$

In the sequel, we often drop the exponent x in $\hat{X}^x$ when there is no ambiguity, and we say that $\hat{X}$ is a quantizer of X. The L^2-error induced by this projection, called L^2-quantization error, is $\|X - \hat{X}\|_2$. As a function of the N-tuple (grid) $x = (x^1, \ldots, x^N) \in (\mathbb{R}^d)^N$, the square of the L^2-quantization error, called distorsion, is written as :

$$D_N^X(x) = \|X - \hat{X}\|_2^2 = \int \min_{i=1,\ldots,N} |u - x^i|^2 \mathbb{P}_X(du). \tag{3.2}$$

First, notice by definition of the closest neighbour projection that the L^2-quantization error is the minimum of L^2-error $\|X - Y\|_2$ among all random variables Y taking values in the grid x. Then, two questions arise naturally : for fixed N, is there an optimal grid x^* which minimizes the L^2-quantization error (or equivalently the distorsion), and how does this minimum behave when N goes to infinity? The latter question is answered by the so-called Zador theorem :

Theorem 3.1 *(see [4])*
Assume that $X \in L^{2+\varepsilon}(\mathbb{P}, \mathbb{R}^d)$ for some $\varepsilon > 0$ and set f the Radon-Nykodim density of $\mathbb{P}_X$ in its decomposition with respect to the Lebesgue measure λ_d on $\mathbb{R}^d$. Then,

$$\lim_N N^{\frac{2}{d}} \min_x \|X - \hat{X}^x\|_2^2 = J_d \|f\|_{\frac{d}{d+2}},$$

where $\|f\|_r = (\int |f|^r d\lambda_d)^{1/r}$ for $r > 0$, and J_d is a constant depending on d, corresponding to the uniform distribution on $[0,1]^d$.

Remark 3.1 In dimensions $d = 1$ and 2, $J_1 = \frac{1}{12}$ and $J_2 = \frac{5}{18\sqrt{3}}$. For $d \geq 3$, $J_d \sim \frac{d}{2\pi e}$ as d goes to infinity.

The optimal N-quantization problem that consists in determining a grid x^*, which minimizes the L^2-quantization error, relies on the property that the distorsion is continuously differentiable at any N-tuple having pairwise distinct components, with a gradient obtained by formal differentiation in (3.2) :

$$\nabla D_N^X(x) = 2\left(\int_{C_i(x)} (x^i - u) P_X(du)\right)_{1 \leq i \leq N}. \tag{3.3}$$

A quantizer $\hat{X} = \hat{X}^x$ is said stationary if the associated N-tuple x satisfies

$$\nabla D_N^X(x) = 0.$$

An optimal quantizer is a stationary quantizer. From (3.3), we have the useful property of stationary quantizers :

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

The integral representation (3.3) of ∇D_N^X suggests, as soon as independent copies of X can be simulated, to implement a stochastic gradient algorithm (descent), in order to get numerically a stationary quantizer. In our context, this leads to the Kohonen algorithm or competitive learning vector quantization (CLVQ) algorithm, which also provides as a byproduct an estimation of the weights $\hat{p}_i$ of the Voronoi tesselations associated to the stationary quantizer. We refer to [8] for a description and discussion of the algorithm. Optimal grids and their companion parameters, i.e. weights of the Voronoi tesselation and distorsion, for the normal distribution are available and downloadable on the webpages of Gilles Pagès or Jacques Printems.

3.2 Filter quantization approximation for a fixed observation

We are in the framework where the signal state space is continuous, say $\mathbb{R}^d$. We show how one can apply quantization methods for providing a numerical approximation of the filter, given a fixed set of observations. This is achieved in three steps.

Step 1. We assume that for each time k, the random vector X_k is square integrable and simulatable. Then, for each k, we apply an optimal vector quantization of the random vector X_k. We denote

$$\hat{X}_k \;=\; \mathrm{Proj}_{x_k}(X_k) \;:=\; \sum_{i=1}^{N_k} x_k^i \mathbf{1}_{C_i(x_k)}(X_k),$$

the associated quantizer on the optimal grid $x_k = (x_k^1, \ldots, x_k^{N_k})$ of size N_k points in $\mathbb{R}^d$.

Step 2. : *Marginal quantization of the Markov chain* (X_k).
This consists in approximating the distribution law of the Markov chain $(X_k)_{0\le k\le n}$ as follows :

$$\begin{array}{rcl} \text{approximate law } \mu \text{ of } X_0 & \text{by} & \text{law } \hat{\mu} \text{ of } \hat{X}_0 \\ \text{approximate law } P_k \text{ of } X_k|X_{k-1} & \text{by} & \text{law } \hat{P}_k \text{ of } \hat{X}_k|\hat{X}_{k-1}, \quad k=1,\ldots,n. \end{array}$$

In other words, $\hat{\mu}$ is the weight vector $(\hat{p}_0^i)$ given by :

$$\hat{p}_0^i \;=\; \mathbb{P}\left[\hat{X}_0 = x_0^i\right], \quad i = 1, \ldots, N_0.$$

and for $k = 1, \ldots, n$, $\hat{P}_k = (\hat{p}_k^{ij})$ is the probability transition matrix :

$$\hat{p}_k^{ij} \;=\; \mathbb{P}\left[\hat{X}_k = x_k^j | \hat{X}_{k-1} = x_{k-1}^i\right],$$

for $i = 1, \ldots, N_{k-1}$, $j = 1, \ldots, N_k$. These transition weights

$$\begin{aligned} \hat{p}_0^i &= \mathbb{P}_{X_0}[C_i(x_0)] \\ \hat{p}_k^{ij} &= \frac{\mathbb{P}_{X_{k-1},X_k}[C_i(x_{k-1}), C_j(x_k)]}{\mathbb{P}_{X_{k-1}}[C_i(x_{k-1})]}, \end{aligned}$$

are estimated by Monte-Carlo simulations of X_k, $k = 0, \ldots, n$. They can also be obtained as a byproduct of the Kohonen algorithm.

Step 3. : *Filter approximation for a fixed observation.*
We are given a set of observations $(Y_0, \ldots, Y_n)$ fixed to $(y_0, \ldots, y_n)$. For each $k = 1, \ldots, n$, we consider the approximation of the prediction-updating transition kernel H_k by the transition matrix $\hat{H}_k = (\hat{H}_k^{ij})$ defined as :

$$\hat{H}_k^{ij} = g_k(x_{k-1}^i, y_{k-1}, x_k^j, y_k)\hat{p}_k^{ij}, \quad i = 1, \ldots, N_{k-1},\ j = 1, \ldots, N_k.$$

We then approximate the filter Π_k by the discrete probability measure $\hat{\Pi}_k$ on the grid x_k : $\hat{\Pi}_k = \sum_{i=1}^{N_k} \hat{\Pi}_k^l \delta_{x_k^i}$, that is defined by the approximate forward equation :

$$\begin{aligned} \hat{\Pi}_0 &= \hat{\mu} \\ \hat{\Pi}_k &= \frac{\hat{\Pi}_{k-1}\hat{H}_k}{(\hat{\Pi}_{k-1}\hat{H}_k)(x_k)}. \end{aligned}$$

The weights $(\hat{\Pi}_k^l)$, $k = 0, \ldots, n$, $l = 1, \ldots, N_k$, are then computed as :

$$\begin{aligned} \hat{\Pi}_0^i &= \hat{p}_0^i, \quad i = 1, \ldots, N_0, \\ \hat{\Pi}_k^j &= \frac{\sum_{i=1}^{N_{k-1}} \hat{H}_k^{ij}\hat{\Pi}_{k-1}^i}{\sum_{j=1}^{N_k}\sum_{i=1}^{N_{k-1}} \hat{H}_k^{ij}\hat{\Pi}_{k-1}^i}, \quad k = 1, \ldots, n, \quad j = 1, \ldots, N_k. \end{aligned}$$

From a practical viewpoint, the above procedure is implemented as follows :

Phase 1. *Off-line computations (the most demanding)* : Optimal quantization of the signal. Notice that this phase does not depend on the observations. We need to :
- Specify the size N_k of the grids for $k = 1, \ldots, n$ given a total number of points $N = N_0 + \ldots + N_n$.
- Process optimal grids (by Kohonen algorithm) and the associated transition weights $(\hat{p}_k^{ij})$.

A special case of interest : a stationary signal. In this usual case in filtering model, we only need to compute the optimal grid $x^* = \{x^1, \ldots, x^{\bar{N}}\}$ of the stationary distribution μ of X_0, with size $\bar{N} = N/(n+1)$. Then, $x_k = x^*$, $k = 0, \ldots, n$, are the optimal grids for each X_k. We estimate the probability $\hat{\mu}$ of $\hat{X}_0 = \mathrm{Proj}_{x^*}(X_0)$, and we only have to estimate one single transition matrix :

$$\hat{p}_k^{ij} = \hat{p}_0^{ij} = \mathbb{P}[\hat{X}_0 = x^j | \hat{X}_0 = x^i], \quad 0 \le i, j \le \bar{N},\ k = 1, \ldots, n.$$

From a computational viewpoint, the size of the parameters to be stored is divided by a factor n (or the available quantization size for the distribution of X_0 and the transition matrix is multiplied by n).

Phase 2. *On-line computations* : given an observation vector $y = (y_0, \ldots, y_n)$, we compute the quantized prediction-updating transition matrix $(\hat{H}_k)$, $k = 1, \ldots, n$. Finally, we compute the quantized filter $(\hat{\Pi}_k)$, $k = 1, \ldots, n$, by the approximate forward equation, and for every (needed) test function ϕ :

$$\hat{\Pi}_n\varphi = \sum_{i=1}^{N_n} \varphi(x_n^i)\hat{\Pi}_n^i$$

This phase is almost instantaneous.

3.3 Error and convergence analysis

We denote

$$BL_1(\mathbb{R}^d) = \Big\{\phi : \mathbb{R}^d \to \mathbb{R},\ \phi \text{ bounded by 1} \\ \text{and } \phi \text{ Lipschitz with } [\phi]_{Lip} := \sup_{x\neq x'} \frac{|\phi(x)-\phi(x')|}{|x-x'|} \leq 1\Big\}.$$

For any $\Pi \in \mathcal{P}(\mathbb{R}^d)$, we denote

$$\Pi\phi = \int \phi(x)\Pi(dx), \quad \forall \phi \in BL_1(\mathbb{R}^d)$$

We make essentially two conditions on the signal-observation model. We assume a Lipschitz condition on the probability transitions of the signal :

(A1) The probability transitions P_k, $k = 1, \ldots, n$, are Lipschitz with ratio $[P_k]_{Lip}$, i.e. for any Lipschitz function ϕ on $\mathbb{R}^d$, with ratio $[\phi]_{Lip}$, we have:

$$|P_k\phi(x) - P_k\phi(\hat{x})| \leq [P_k]_{Lip}[\phi]_{Lip}|x-\hat{x}|, \quad \forall x, \widehat{x} \in \mathbb{R}^d.$$

We then set $[P]_{Lip} := \max_{k=1,\ldots,n}[P_k]_{Lip}$.

We also assume a Lipschitz condition on the updating observation functions :

(A2)
- (i) The functions g_k, $k = 1, \ldots, n$, are bounded.
  $\|g\|_\infty := \max_{k=1,\ldots,n} \|g_k\|_\infty$
- (ii) There exists $[g_k]_{Lip}$, $k = 1, \ldots, n$, s.t. $\forall x, x', \hat{x}, \hat{x}' \in \mathbb{R}^d, y, y' \in \mathbb{R}^q$
  $|g_k(x,y,x',y') - g_k(\hat{x},y,\hat{x}',y')| \leq [g_k]_{Lip}(|x-\hat{x}| + |x'-\hat{x}'|)$.
  $[g]_{Lip} := \max_{k=1,\ldots,n}[g_k]_{Lip}$.

We then have the following error bound for the approximation of the filter by quantization.

Theorem 3.2 *Under* **(A1)** *and* **(A2)**, *given a fixed observation* $(Y_0, \ldots, Y_n) = (y_0, \ldots, y_n)$, *we have :*

$$\sup_{\phi \in BL_1(\mathbb{R}^d)} \left|\Pi_n\phi - \hat{\Pi}_n\phi\right| \leq \frac{\|g\|_\infty^n}{\gamma_n(y)} \sum_{k=0}^{n} A_{n,k}\|X_k - \hat{X}_k\|_2, \tag{3.4}$$

where $\gamma_n(y)$ *is the density of* $(Y_1, \ldots, Y_n)$ *at* $y = (y_1, \ldots, y_n)$ *:*

$$\gamma_n(y) = \mathbb{E}\left[\prod_{k=1}^{n} g_k(X_{k-1}, y_{k-1}, X_k, y_k)\right]$$

and

$$A_{n,k} = [P]_{Lip}^{n-k+1} + \frac{2[g]_{Lip}}{\|g\|_\infty}\left(\frac{[P]_{Lip}+1}{[P]_{Lip}-1}([P]_{Lip}^{n-k+1}-1)+2\right).$$

Elements of proof. We give a sketch of the proof of this theorem.
Step 1. Backward representation of the filter : We consider the unnormalized filter (π_k) given by the unnormalized forward linear equation :

$$\pi_0 = \mu, \quad \pi_k = \pi_{k-1} H_k, \quad k = 1, \dots, n,$$

so that

$$\Pi_n = \frac{\pi_n}{\pi_n(E)} \quad \text{and} \quad \pi_n = \mu H_1 \dots H_n$$

From this symmetric expression, we introduce the transition kernels given by the backward equation :

$$K_n = Id, \quad K_k = H_{k+1} K_{k+1}, \quad k = 0, \dots, n-1,$$

so that

$$\pi_n = \mu K_0$$

Similarly, the quantized filter is expressed in a backward induction :

$$\hat{\Pi}_n = \frac{\hat{\pi}_n}{\hat{\pi}_n(E)},$$

where

$$\hat{\pi}_n = \hat{\mu} \hat{K}_0$$

and

$$\hat{K}_n = Id, \quad \hat{K}_k = \hat{H}_{k+1} \hat{K}_{k+1}, \quad k = 0, \dots, n-1.$$

Step 2. Error approximation of the unnormalized filter : We write

$$\begin{aligned} |\pi_n \phi - \hat{\pi}_n \phi| = \left| \mu K_0 \phi - \hat{\mu} \hat{K}_0 \phi \right| &= \left| \mathbb{E}\left[K_0 \phi(X_0) \right] - \mathbb{E}\left[\hat{K}_0 \phi(\hat{X}_0) \right] \right| \\ &\leq \left\| K_0 \phi(X_0) - \hat{K}_0 \phi(\hat{X}_0) \right\|_2 \end{aligned}$$

From the backward formula on K_k and $\hat{K}_k$, we derive an estimation of

$$\left\| K_k \phi(X_k) - \hat{K}_k \phi(\hat{X}_k) \right\|_2$$

in terms of the quantization error $\left\| X_k - \hat{X}_k \right\|_2$ by a backward induction : this uses

- Lipschitz condition **(A1)**, **(A2)**
- L^2-contraction property of conditional expectation and the fact that $\hat{X}_k$ is $\sigma(X_k)$-measurable
- Gronwall's lemma.

Step 3. Error approximation of the (normalized) filter : We write

$$\sup_{\phi\in BL_1(\mathbb{R}^d)} \left|\Pi_n\phi - \hat{\Pi}_n\phi\right| = \sup_{\phi\in BL_1(\mathbb{R}^d)} \left|\frac{\pi_n\phi}{\pi_n(E)} - \frac{\hat{\pi}_n\phi}{\hat{\pi}_n(E)}\right| \leq \frac{2\sup_{\phi\in BL_1(\mathbb{R}^d)}|\pi_n\phi - \hat{\pi}_n\phi|}{\pi_n(E)\vee\hat{\pi}_n(E)},$$

and we notice that $\pi_n(E) = \gamma_n(y)$ is the density of $(Y_1,\ldots,Y_n)$. □

Remark 3.2 *Convergence of the quantized filter.* If the grids are chosen optimally at each time $k = 0,\ldots,n$, then in view of Zador's theorem, we get a bound for the rate of convergence of the quantized filter :

$$\sup_{\phi\in BL_1(\mathbb{R}^d)} \left|<\Pi_n,\phi> - <\hat{\Pi}_n,\phi>\right| \leq \frac{\|g\|_\infty^n}{\gamma_n(y)}\sum_{k=0}^n A_{n,k}\, C(\mathbb{P}_{X_k},d)\frac{1}{N_k^{\frac{1}{d}}}. \qquad (3.5)$$

Consequently :

- Given a total number of points N, we may optimally dispatch the number of points for each time grid, i.e. find $(N_0,\ldots,N_k)$ s.t. $N_0+\ldots+N_n = N$ and minimizing the r.h.s. of (3.5).

- For a fixed horizon n, we have the convergence of the quantized filter, i.e. $\hat{\Pi}_n$ converges to Π_n as $\min_{0\leq k\leq n} N_k$ goes to infinity.

- When n goes to infinity, the convergence of the filter is satisfied typically in the case of discretized diffusion on $[0,T]$ with Euler scheme of step T/n :

$$X_{k+1} = X_k + b(X_k)\frac{T}{n} + \sigma(X_k)\sqrt{\frac{T}{n}}\varepsilon_{k+1}.$$

Under Lipschitz condition on b and σ, we have :

$$[P]_{Lip} = 1+\frac{c}{n}$$

for some constant c. Then if we simply assign $N_k = \bar{N} = N/(n+1)$ points at each grid, (3.5) provides a rate of convergence of order

$$\frac{\|g\|_\infty^n}{\gamma_n(y)}\,\frac{n+1}{\bar{N}^{1/d}}$$

This is to be compared with the rate of convergence obtained by Monte-Carlo methods using $\bar{N}$ interacting particles :

$$\left(\frac{\|g\|_\infty^n}{\gamma_n(y)}\right)^n \frac{1}{\bar{N}^{1/2}}$$

Remark 3.3 *Extensions : first-order schemes.* In the method described in paragraph 3.2, we approximated the probability transition P_k as follows : for any Borel function ϕ

$$P_k\phi(X_k) = \mathbb{E}\left[\phi(X_{k+1})\middle|\, X_k\right] \simeq \hat{P}_k\phi(\hat{X}_k) := \mathbb{E}\left[\phi(\hat{X}_{k+1})\middle|\,\hat{X}_k\right]$$

This is a piecewise constant approximation of the conditional expectation at the centers x_{k+1}^j and x_k^i of the tesselations of $\hat{X}_{k+1}$ and $\hat{X}_k$, and is called zero-order scheme. This suggests to consider linear interpolation based on Taylor expansion around the centers of the tesselations, which leads to correction terms in the transition weights $\hat{P}_k$, and to so-called first order scheme for quantization. The main interest is that thanks to stationary property of optimal grids : $\mathbb{E}[X_k|\hat{X}_k] = \hat{X}_k$, we expect to get an estimation error with terms

$$\left\|X_k - \hat{X}_k\right\|_2^2$$

instead of $\|X_k - \hat{X}_k\|_2$ for zero order scheme as in (3.4). Consequently, we should get an improved rate of convergence. These first-order schemes are developed in [11].

3.4 Application : Pricing of European options under partial information

We give a direct application of the above quantization procedure for the calculation of European options under partial observation. Namely, let us consider (X_k), $k = 0, \ldots, n$, the return and/or volatility process of the stock price. (Y_k), $k = 0, \ldots, n$, is the (Logarithm) of the stock price process. We denote $\mathcal{F}_k = \sigma(X_j, Y_j, 0 \leq j \leq k)$, $k = 0, \ldots, n$, the complete information and $\mathcal{F}_k^Y = \sigma(Y_j, 0 \leq j \leq k)$, $k = 0, \ldots, n$ the partial information, i.e. when one does not observe return/volatility but only price process. In this model, we are given an European option of payoff $h(Y_n)$ and more generally $h(X_n, Y_n)$. Its price under complete information is given at time k by :

$$U_k \;=\; \mathbb{E}\left[h(X_n, Y_n)|\mathcal{F}_k\right] \;=\; v_k(X_k, Y_k),$$

for some Borel function v_k by the Markov property of (X, Y). (Here, we supposed that $\mathbb{P}$ is already a risk-neutral probability measure). The function v may be easily computed by different methods, e.g. quantization or Monte-Carlo. On the other hand, the price of the European option under partial information at time k is given by :

$$U_k^Y \;=\; \mathbb{E}\left[h(X_n, Y_n)|\mathcal{F}_k^Y\right].$$

By the law of iterated condition expectation, it is written as :

$$\begin{aligned} U_k^Y \;=\; \mathbb{E}\left[h(X_n, Y_n)|\mathcal{F}_k^Y\right] \;&=\; \mathbb{E}\left[v_k(X_k, Y_k)|\mathcal{F}_k^Y\right] \\ &=\; \int v_k(x, Y_k)\Pi_k(dx) \;=:\; \Pi_k v_k(., Y_k) \end{aligned}$$

Given an observation $(Y_0, \ldots, Y_k) = (y_0, \ldots, y_k)$, this is approximated by the explicit formula :

$$\hat{\Pi}_k v_k(., y_k) \;:=\; \sum_{i=1}^{N_k} v_k(x_k^i, y_k)\hat{\Pi}_k^i,$$

where $\hat{\Pi}_k$ is the quantized filter.

4 Numerical experiments (Fixed observation)

4.1 Kalman-Bucy model

We first illustrate the filtering quantization method in paragraph 3.2 with the Kalman Bucy model. This linear Gaussian model for the signal-observation process is described by

$$\begin{aligned} X_{k+1} &= \rho X_k + \theta \varepsilon_k, \quad X_0 \rightsquigarrow \mathcal{N}(0, \Sigma_0) \\ Y_k &= X_k + \gamma \eta_k, \end{aligned}$$

where ρ and θ are constant matrices of appropriate dimensions, and (ε_k), (η_k) are independent Gaussian noises : $\varepsilon_k \rightsquigarrow \mathcal{N}(0, I_d)$, $\eta_k \rightsquigarrow \mathcal{N}(0, I_d)$. In this case, the filter is explicit :

$$\Pi_n \rightsquigarrow \mathcal{N}(m_n, \Sigma_n),$$

where m_n and Σ_n are computed by forward induction, see e.g. [2].

We perform numerical tests with parameters chosen so that the signal X_k is stationary, i.e. $X_k \rightsquigarrow \mathcal{N}(0, \Sigma_0)$ for all k. In that case, we can work with a single grid at each time k. Actually, we start with the optimal (prestored) grid for $\mathcal{N}(0, I_d)$ and make an homothety of Σ_0. We put the same number $\bar{N}$ of points at each time grid.

4.1.1 Test 1 : convergence of the filter at a fixed instant n when the number of points $\bar{N}$ of each grid goes to infinity

The approximate filter $\hat{\Pi}_n$ is computed on the test functions ϕ_i, i.e. $\Pi_n \phi_i$, for

$$\begin{aligned} \phi_1(x) &= x_d, \quad x = (x_1, \ldots, x_d) \\ \phi_2(x) &= |x|^2, \quad \phi_3(x) = \exp(-|x_d|), \end{aligned}$$

in signal dimension $d = 1$ and 3. The following figures show the convergence and the rate of convergence of the approximated filter by quantization, when $\bar{N}$ goes to infinity. The theoretical convergence rate $\bar{N}^{\frac{1}{d}}$ (see Remark 3.2) is then in ln-scale 1 in dimension 1 and $1/3$ in dimension 3, and is consistent to what we found by numerical experiments. Actually, we even get better rate of convergence with the numerical tests.

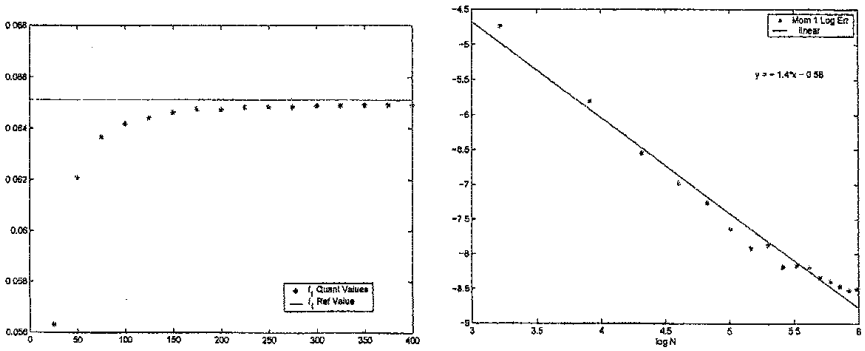

Figure 1a : dim1 - : Convergence and Convergence Rate (in a ln-scale) of $\hat{\Pi}_n \phi_1$ as N grows, for $n = 15$ and $\phi_1(x) = x$.

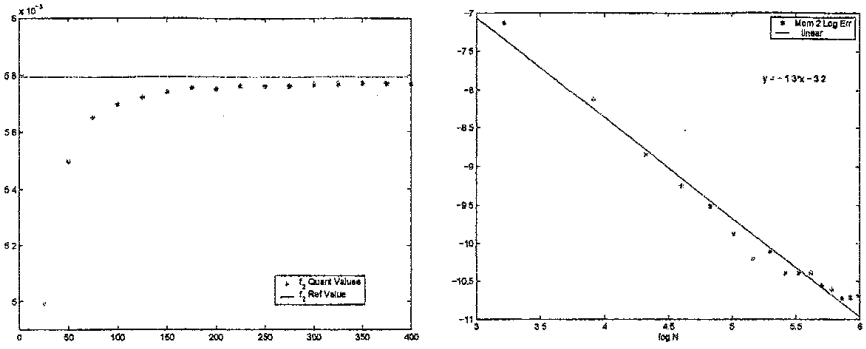

Figure 1b : dim1 - : Convergence and Convergence Rate (in a ln-scale) of $\hat{\Pi}_n\phi_2$ as N grows, for $n = 15$ and $\phi_2(x) = x^2$.

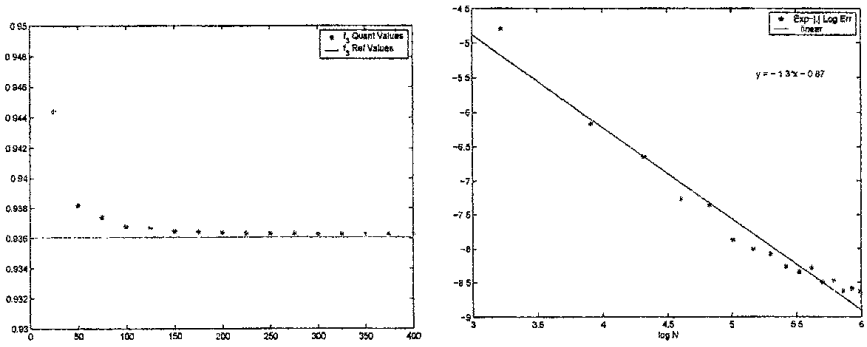

Figure 1c : dim1 - : Convergence and Convergence Rate (in a ln-scale) of $\hat{\Pi}_n\phi_3$ as N grows, for $n = 15$ and $\phi_3(x) = \exp(-|x|)$.

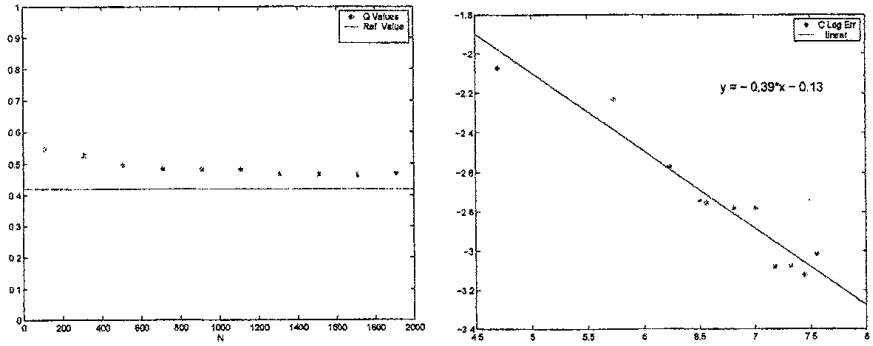

Figure 2a : dim3 - : Convergence and Convergence Rate (in a ln-scale) of $\hat{\Pi}_n\phi_1$ as N grows, for $n = 15$ and $\phi_1(x) = x_3$.

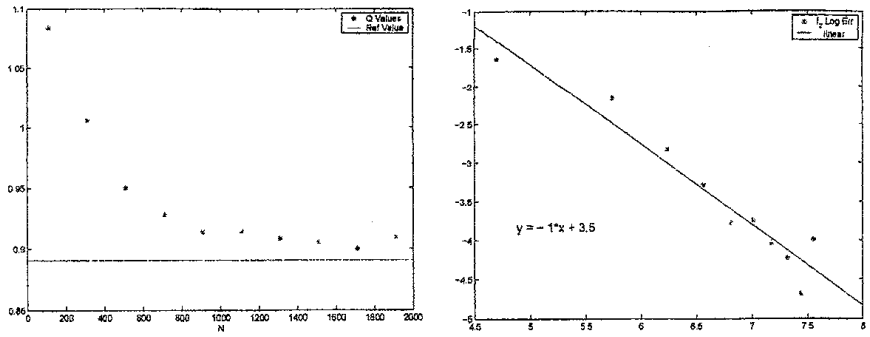

Figure 2b : dim3 - : Convergence and Convergence Rate (in a ln-scale) of $\hat{\Pi}_n\phi_2$ as N grows, for $n = 15$ and $\phi_2(x) = |x|^2$.

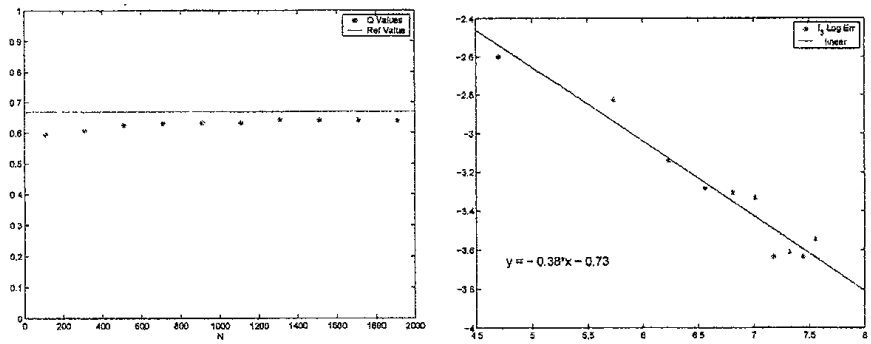

Figure 2c : dim3 - : Convergence and Convergence Rate (in a ln-scale) of $\hat{\Pi}_n\phi_3$ as N grows, for $n = 15$ and $\phi_3(x) = \exp(-|x_3|)$.

4.1.2 Test 2 : Stability of the filter for a fixed grid size $\bar{N}$ as n goes to infinity.

We perform numerical tests for a signal in dimension 2. The following figures show the stability of the quantized filter when the horizon n goes to infinity.

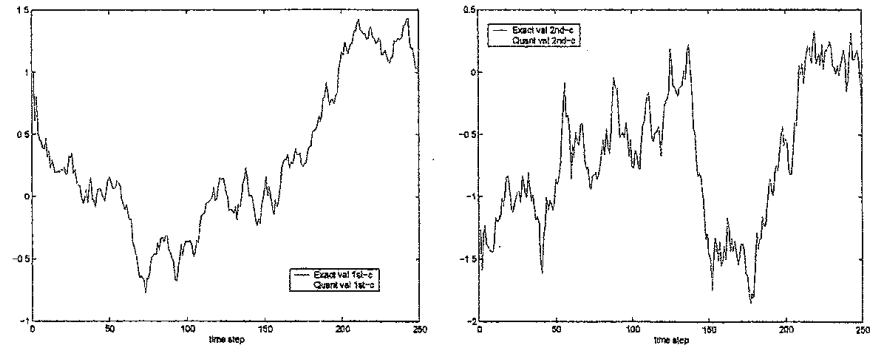

Figure 3 : dim 2 - : Error dependence of $\hat{\Pi}_n\phi$ as n grows, for $\bar{N} = 600$, $\phi(x) = x_1$ and $\phi(x) = x_2$.

4.2 A stochastic volatility model

We now consider the ARCH model :

$$\begin{aligned} X_{k+1} &= \rho X_k + \varepsilon_k, \quad X_0 \rightsquigarrow \mathcal{N}(0, \Sigma_0) \\ Y_k &= \sigma(X_k)\eta_k, \quad \sigma(x) = \gamma + |x|, \end{aligned}$$

where (ε_k) are (η_k) independent Gaussian noises. This model is popular in finance, as a discretization of stochastic volatility model, where X is the volatility and Y the price process. Unlike Kalman-Bucy model, there is no explicit reference formula for the filter. The following figures show the convergence of the quantized filter when $\bar{N}$ goes to infinity.

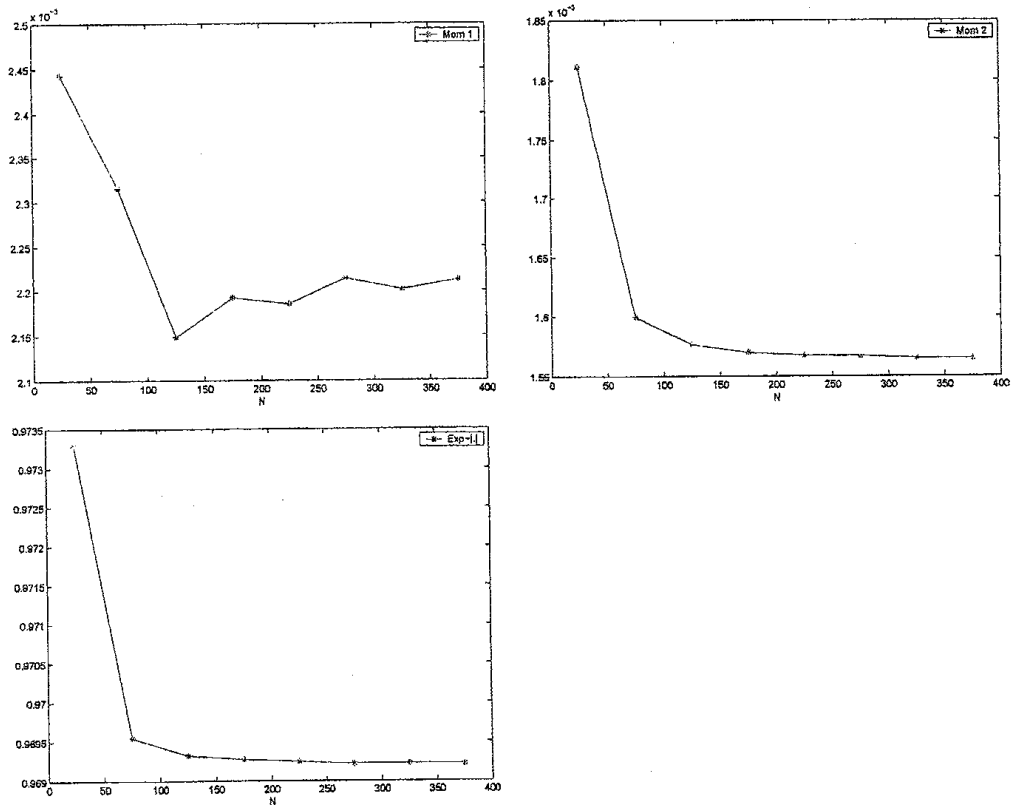

Figure 4 : SV model : filter values of $\hat{\Pi}_n\phi$ at fixed n as $\bar{N}$ grows, for $\phi(x) = x$, $|x|^2$, and $\exp(-|x|)$.

More numerical illustrations are investigated in Sellami's thesis [12] with in particular comparison to Monte-Carlo particle methods.

5 Approximation of the filter process by quantization

In the quantization algorithm described in the previous section, we need for each new set of observations, to compute again the quantized transition matrices $(\hat{H}_k)$ via the local

likelihood observation function g_k. This on-line phase may be removed by an off-line preprocessing of the observations, as suggested in [6]. One can then store in addition to the signal quantizers, the local likelihood functions precomputed on the signal-observation grids. This observation quantization approach is developed in [11], where error estimation are provided and numerically illustrated. If we stress the dependence of the filter on the observation : $\Pi_n(Y_0,\ldots,Y_n)$, then it is approximated by $\hat{\Pi}_n(\hat{Y}_0,\ldots,\hat{Y}_n)$, where $\hat{Y}_k$ is a quantizer of Y_k. However, we notice that the size of the look-up tables for the filter may be very large. For instance, if $\hat{Y}_k$ takes M values, then at time n, the random filter $\hat{\Pi}_n(\hat{Y}_0,\ldots,\hat{Y}_n)$ would take M^n values in $\mathcal{P}(E)$, which may explode for a long horizon n. This makes computations untractable when solving dynamic optimization problems under partial information, even if E is already a finite state space.

In order to overcome this numerical difficulty, we present a quantization approach introduced in [9] and based on the Markov property of the pair filter-observation (Π_k, Y_k) with respect to the observation filtration $(\mathcal{F}_k^Y)$. Indeed, by denoting $(\mathcal{F}_k)$ 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 $\mathcal{P}(E)\times\mathbb{R}^q$:

$$\begin{aligned}
&\mathbb{E}\left[\varphi(\Pi_{k+1},Y_{k+1})|\,\mathcal{F}_k^Y\right]\\
&= \mathbb{E}\left[\mathbb{E}[\varphi(\bar{G}_{k+1}(\Pi_k,Y_k,Y_{k+1}),Y_{k+1})|\mathcal{F}_k]\big|\,\mathcal{F}_k^Y\right]\\
&= \mathbb{E}\left[\int \varphi(\bar{G}_{k+1}(\Pi_k,Y_k,y'),y')P_{k+1}(X_k,dx')q_{k+1}(X_k,Y_k,x',dy')\,\middle|\,\mathcal{F}_k^Y\right]\\
&= \int \varphi(\bar{G}_{k+1}(\Pi_k,Y_k,y'),y')P_{k+1}(x,dx')q_{k+1}(x,Y_k,x',dy')\Pi_k(dx). \qquad (5.6)
\end{aligned}$$

This shows the Markov property of (Z_k) with probability transition R_k (from time $k-1$ to k) given by :

$$R_k\varphi(\pi,y) = \int \varphi(\bar{G}_k(\pi,y,y'),y')Q_k(\pi,y,dy'), \qquad (5.7)$$

where $Q_k(\pi,y,dy')$ is the law of Y_k conditional on $(\Pi_{k-1},Y_{k-1})=(\pi,y)$ with density :

$$y' \longrightarrow \int g_k(x,y,x',y')P_k(x,dx')\pi(dx).$$

We consider a framework with finite state space : $E=\{x^1,\ldots,x^m\}$, so that the filter Π_k is a random discrete probability measure identified with a random vector $\Pi_k=(\Pi_k^i)_{1\le i\le m}$ in the simplex K_m of $\mathbb{R}^m$:

$$K_m = \left\{\Pi\in\mathbb{R}_+^m \,:\, \sum_{i=1}^m \Pi^i = 1\right\} \simeq \mathcal{P}(E).$$

The idea is to apply a marginal quantization of the Markov chain $(Z_k)=(\Pi_k,Y_k)$ valued in $K_m\times\mathbb{R}^q$. Hence, for each $k=0,\ldots,n$, we denote

$$\hat{Z}_k = \mathrm{Proj}_{z_k}(Z_k) := \sum_{i=1}^{N_k} z_k^i \mathbf{1}_{C_i(z_k)}(Z_k),$$

the associated quantizer on the optimal grid $z_k = (z_k^1, \ldots, z_k^{N_k})$ of size N_k points in $K_m \times \mathbb{R}^q$. The probability transition R_k of the Markov chain (Z_k) is approximated by the transition matrix $\hat{R}_k = (\hat{r}_k^{ij})$:

$$\hat{r}_k^{ij} = \mathbb{P}\left[\hat{Z}_k = z_k^j \middle| \hat{Z}_{k-1} = z_{k-1}^i\right].$$

The optimal grids z_k and the associated transition weights $\hat{r}_k^{ij}$ are processed and estimated by the Kohonen algorithm. This is based on the Monte-Carlo simulations of (Z_k), which rely themselves, from (5.7), on the following simulation procedure of the probability transition R_k :

- simulate X_{k-1} with probability law Π_{k-1}, and then X_k according to the probability transition P_k
- simulate Y_k according to the probability transition $g_k(X_{k-1}, Y_{k-1}, X_k)dy'$
- compute Π_k by the forward filtering (finite-dimensional) equation

$$\Pi_k = \bar{G}_k(\Pi_{k-1}Y_{k-1}, Y_k) = \frac{\Pi_{k-1}H_k}{\Pi_{k-1}H_k(E)}.$$

6 Application : pricing of American options under partial observation

6.1 Optimal stopping problem under partial observation

Given a bounded measurable function h on $E \times \mathbb{R}^q$, and a horizon n, we denote for any $k = 0, \ldots, n$, $\mathcal{T}_{k,n}^Y$ as the set of $(\mathcal{F}_k^Y)$-stopping times valued in $\{k, \ldots, n\}$, and we consider the following optimal stopping problem under partial observation :

$$U_k = \operatorname*{ess\,sup}_{\tau \in \mathcal{T}_{k,n}^Y} \mathbb{E}\left[h(X_\tau, Y_\tau) \middle| \mathcal{F}_k^Y\right]. \tag{6.8}$$

By using the law of iterated conditional expectation and the definition of the filter, we notice that problem (6.8) may be reduced to a complete observation model with state variable the $(\mathcal{F}_k^Y)$-adapted process (Z_k) :

$$\begin{aligned} U_k &= \operatorname*{ess\,sup}_{\tau \in \mathcal{T}_{k,n}^Y} \mathbb{E}\left[\sum_{j=k}^n 1_{\tau=j}\mathbb{E}[h(X_j, Y_j)|\mathcal{F}_j^Y] \middle| \mathcal{F}_k^Y\right] \\ &= \operatorname*{ess\,sup}_{\tau \in \mathcal{T}_{k,n}^Y} \mathbb{E}\left[\sum_{j=k}^n 1_{\tau=j}\Pi_j h(., Y_j) \middle| \mathcal{F}_k^Y\right] \\ &= \operatorname*{ess\,sup}_{\tau \in \mathcal{T}_{k,n}^Y} \mathbb{E}\left[\Pi_\tau h(., Y_j) \middle| \mathcal{F}_k^Y\right] = \operatorname*{ess\,sup}_{\tau \in \mathcal{T}_{k,n}^Y} \mathbb{E}\left[\tilde{h}(Z_\tau) \middle| \mathcal{F}_k^Y\right], \end{aligned}$$

with the notation :

$$\tilde{h}(z) = \pi h(., y) = \sum_{i=1}^m h(x^i, y)\pi^i, \quad \forall z = (\pi, y),\ \pi = (\pi^i)_i \in K_m, y \in \mathbb{R}^q.$$

By the $(\mathcal{F}_k^Y)$-Markov property of (Z_k) and the dynamic programming principle, we have $U_k = u_k(Z_k)$ where functions u_k are calculated in backward induction by :

$$\begin{aligned} u_n(z) &= \tilde{h}(z) \\ u_k(z) &= \max\left\{\tilde{h}(z)\,,\ \mathbb{E}\left[u_{k+1}(Z_{k+1})\middle|\, Z_k = z\right]\right\}. \end{aligned}$$

Following [1], we provide a quantization approximation of $U_k = u_k(Z_k)$ by $\hat{U}_k = \hat{u}_k(\hat{Z}_k)$, where $(\hat{Z}_k)$ is a marginal quantization of (Z_k) on grid z_k, as described in the previous section, and functions $\hat{u}_k$ are explicitly computed in recursive form by :

$$\begin{aligned} \hat{u}_n(z) &= \tilde{h}(z) \\ \hat{u}_k(z) &= \max\left\{\tilde{h}(z)\,,\ \mathbb{E}\left[\hat{u}_{k+1}(\hat{Z}_{k+1})\middle|\, \hat{Z}_k = z\right]\right\}. \end{aligned}$$

From an algorithmic viewpoint, this reads as :

$$\begin{aligned} \hat{u}_n(z_n^i) &= \tilde{h}(z_n^i), \quad i = 1, \ldots, N_n \\ \hat{u}_k(z_k^i) &= \max\left\{\tilde{h}(z_k^i)\,,\ \sum_{j=1}^{N_{k+1}} \hat{r}_{k+1}^{ij}\hat{u}_{k+1}(z_{k+1}^j)\right\}, \quad i = 1, \ldots, N_k. \end{aligned}$$

L^1-error estimation $\|U_k - \hat{U}_k\|_1$ in terms of quantization error $\|Z_k - \hat{Z}_k\|_2$ is stated in [9].

6.2 Numerical illustration : Bermudean options in a partially observed stochastic volatility model

We consider an observable stock (logarithm) price $Y_k = \ln S_k$, with dynamics given by :

$$Y_{k+1} = Y_k + \left(r - \frac{1}{2}X_k^2\right)\delta + X_k\sqrt{\delta}\varepsilon_{k+1}$$

where (ε_k) is a sequence of Gaussian white noise, and (X_k) is the unobservable volatility process. $\delta = \frac{T}{n}$ is the time step from an Euler scheme. We assume that (X_k) is a Markov chain approximation à la Kushner [5] with spatial step Δ and with $m = 3$ states of a mean-reverting process :

$$dX_t = \lambda(x_0 - X_t)dt + \eta dW_t.$$

In this context of a partially observed stochastic volatility model, we consider a Bermudean put option with payoff $h(y) = (\kappa - e^y)_+$, and with price :

$$u_0 = \sup_{\tau \in \mathcal{T}_{0,n}^Y} \mathbb{E}\left[e^{-r\tau\delta}h(Y_\tau)\right]. \tag{6.9}$$

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 $\bar{N}$ fixed for each time period with step $\delta = \frac{1}{n}$.

	$E[\Pi_n^1]$	$E[\Pi_n^2]$	$E[\Pi_n^3]$	Relative error (%)
Monte Carlo	0.287608	0.422833	0.289558	
Quant. with $\bar{N} = 300$	0.301651	0.421725	0.276624	0.898
Quant. with $\bar{N} = 600$	0.301604	0.421458	0.276938	0.886
Quant. with $\bar{N} = 900$	0.301598	0.421316	0.277086	0.881
Quant. with $\bar{N} = 1200$	0.301618	0.42122	0.277162	0.879
Quant. with $\bar{N} = 1500$	0.301605	0.421205	0.27719	0.878

Table 1: Comparison of quantized filter value to its Monte Carlo estimation

We first compare in Table 1 the filter expectation at the final date computed with a time step size $\delta = 1/5$ and by using the optimal quantization method with increasing grid size $\bar{N}$, and with 10^6 Monte Carlo iterations.

We observe that besides the very low error level, the absolute error (plotted in Figure 5) and the relative error are decreasing as the grid size grows.

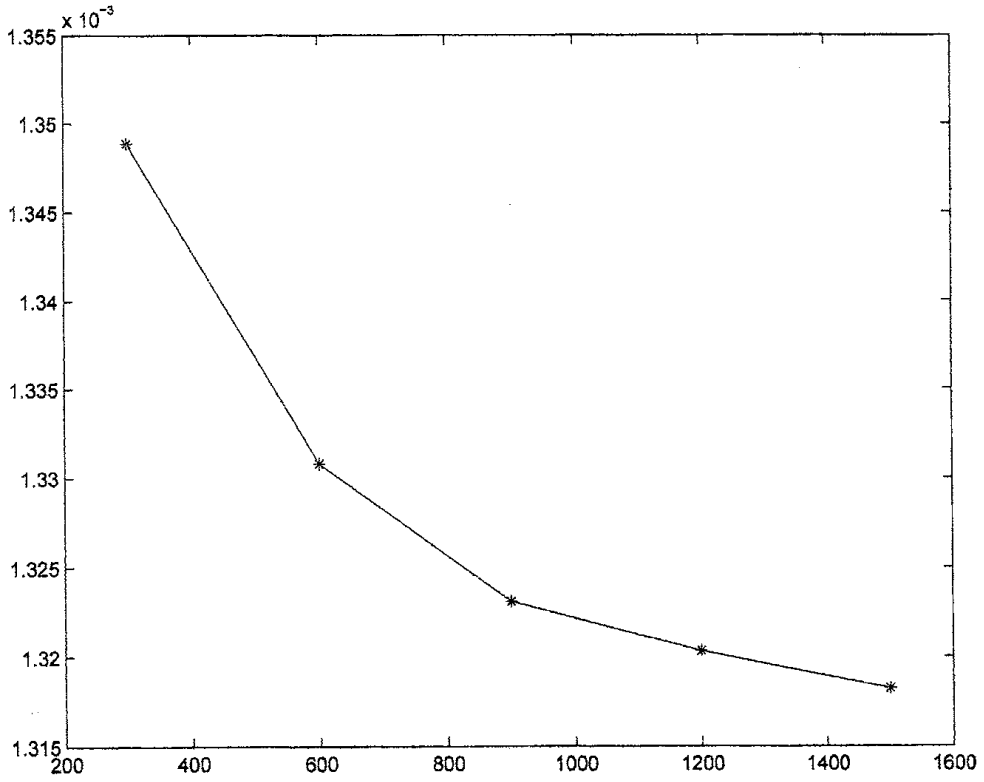

Figure 5 : Filter error convergence as N grows

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 6, where we performed option pricing over grids of size $\bar{N}_{\Pi,Y} = 1500$ in case of partial observation. The total observation price is given by the same pricing algorithm carried out on $\bar{N}_{X,Y} = 45$ points for the product grid of (X_k, Y_k). For fixed n, the rate of convergence for the approximation of the value function under partial observation is of order $\bar{N}_{\Pi,Y}^{1/(m-1+d)}$ where $\bar{N}_{\Pi,Y}$ 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 [1], we also know that the rate of convergence for the approximation of the value function under full observation is of order $m \times \bar{N}_Y$ where $\bar{N}_{X,Y} = m \times \bar{N}_Y$ 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 $\bar{N}_Y \sim \bar{N}_{\Pi,Y}^{1/3}$.

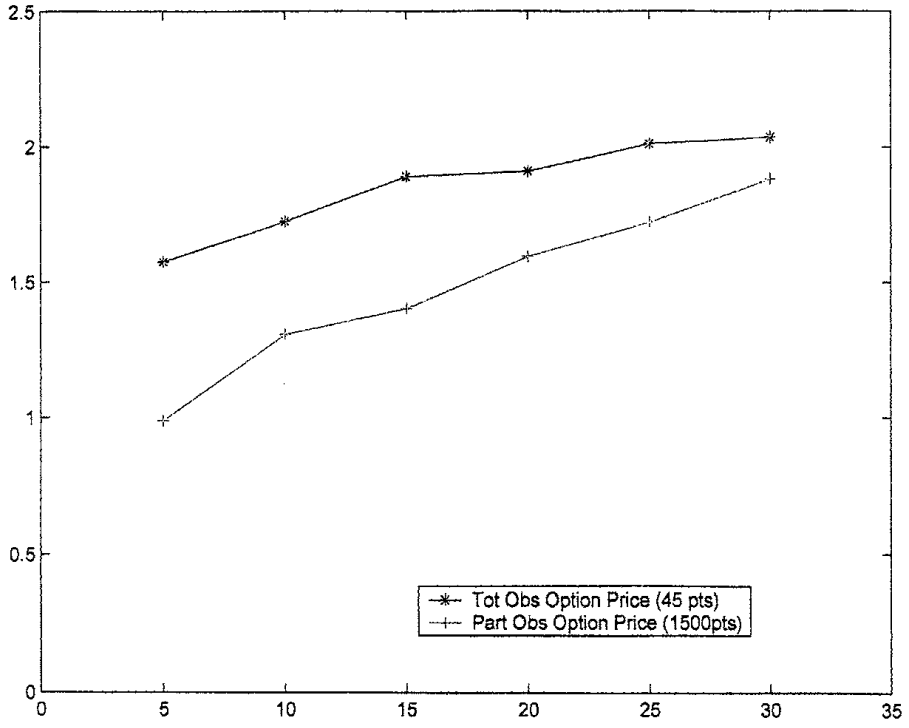

Figure 6 : Partial and total observation option prices as $\delta \to 0$

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 2 and 3).

n	4	8	16
Tot. Obs. ($\bar{N}_{X,Y} = 30$)	1.45863	1.75689	1.77642
Part. Obs. ($\bar{N}_{\Pi,Y} = 1000$)	0.921729	1.13898	1.47089
Variation	0.53	0.61	0.30

Table 2: American option price for embedded filtrations - First Example

n	5	10	20
Tot. Obs. ($\bar{N}_{X,Y} = 45$)	1.57506	1.72595	1.91208
Part. Obs. ($\bar{N}_{\Pi,Y} = 1500$)	0.988531	1.30616	1.59632
Variation	0.58	0.42	0.31

Table 3: American option price for embedded filtrations - Second Example

References

[1] Bally V. and G. Pagès (2003) : "A quantization algorithm for solving discrete time multi-dimensional optimal stopping problems", *Bernoulli*, **9**, 1003-1049.

[2] Elliott R., Aggoun L. and J. Moore (1995) : *Hidden Markov models, estimation and control*, Springer Verlag.

[3] Genon Catalot V. (2003) : "A non linear explicit filter", *Stat. Prob. Lett.*, **61**, 145-154.

[4] Graf S. and H. Luschgy (2000) : :*Foundations of quantization for random vectors*, Lecture Notes in Mathematics n^{0}1730, Springer, Berlin, 230 pp.

[5] Kushner H.J. and P. Dupuis (1992) : *Numerical Methods for Stochastic Control Problems in Continuous Time*, Springer, New York.

[6] Newton N. (2001) : "Approximations for nonlinear filters based on quantization", *Monte Carlo Methods and Appl*, **7**, 311-320.

[7] Pagès G. and H. Pham (2005) : "Optimal quantization methods for nonlinear filtering with discrete-time observations", *Bernoulli*, **11**.

[8] Pagès G., Pham H. and J. Printems (2004) : "Optimal quantization methods and applications to numerical problems in finance", *Handbook of computational and numerical methods in finance*, ed. S. Rachev, Birkhauser.

[9] Pham H., Runggaldier W. and A. Sellami (2005) : "Approximation by quantization of the filter process and applications to optimal stopping problems under partial observation", *Monte Carlo Methods and Applications*, **11**, 57-82.

[10] Sellami A. (2004) : "Nonlinear filtering with observation quantization", Preprint, Laboratoire de Probabilités et modèles aléatoires, Universités Paris 6 & 7 (France).

[11] Sellami A. (2005) : "Quantization based filtering method using first order approximation", Preprint PMA-1009.

[12] Sellami A. (2005) : Méthodes de quantification optimale en filtrage et applications en finance, Phd Thesis, University Paris Dauphine.