The Gauss-Galerkin approximation method in nonlinear filtering

Fabien Campillo

The Gauss-Galerkin approximation method in nonlinear filtering

Fabien Campillo*

January 22, 2023

Abstract

We study an approximation method for the one-dimensional nonlinear filtering problem, with discrete time and continuous time observation. We first present the method applied to the Fokker-Planck equation. The convergence of the approximation is established. We finally present a numerical example.

Keywords: nonlinear filtering, moment method, particle approximation.

*This is the English translation of the paper: "*Fabien Campillo. La méthode d'approximation de Gauss-Galerkin en filtrage non linéaire. *RAIRO M2AN*, 20(2):203–223, 1986*" with some supplementary material, see Addendum page 23.*

1 Introduction

Usual methods for numerical solutions of partial differential equations typically involve a large number of space discretization points. Moreover, in their classical form, these methods use time-fixed discretization grids.

The method proposed by Donald A. Dawson [4] for the numerical solution of the Fokker-Planck equation, called the Gauss-Galerkin method, combines the Gauss quadrature and Galerkin approximation methods. This method, which can be considered as a particle method [10], has the double advantage of giving acceptable results even with a small number of unknown variables to calculate and a discretization grid able to adapt to the evolution of the solution of the partial differential equation considered. However, in its current form, the method is limited to the case of a single dimension of space.

We will study the behavior of this method, applied to the nonlinear filtering problem. In Section 2, we present the Gauss-Galerkin approximation method applied to the Fokker-Planck equation, and we establish a convergence result. The results in this section are a reworking and development of the work of Donald A. Dawson [4].

In Section 3, we first consider the nonlinear filtering problem with discrete time observation: we present the Gauss-Galerkin approximation and prove its convergence. We then

*Fabien.Campillo@inria.fr – website– MathNeuro, Inria Montpellier, France

consider the nonlinear filtering problem with continuous time observation: In this case, before introducing the approximation, we go back to the previous case by discretizing the observation equation.

We define the following spaces:

$\mathcal{M}_1(\mathbb{R})$	probability measures on $\mathbb{R}$,
$\mathcal{M}_+(\mathbb{R})$	non-negative measures on $\mathbb{R}$,
$\mathcal{M}(\mathbb{R})$	signed measures on $\mathbb{R}$,
$\mathcal{C}_{\mathrm{u}}(\mathbb{R})$	bounded and uniformly continuous functions $\mathbb{R} \to \mathbb{R}$,
$\mathcal{C}_{\mathrm{c}}^{\infty}(\mathbb{R})$	continuous functions $\mathbb{R} \to \mathbb{R}$ of class $\mathcal{C}^{\infty}$ with compact support,
$\mathcal{C}_{\mathrm{b}}^{2}(\mathbb{R})$	continuous and bounded functions $\mathbb{R} \to \mathbb{R}$ of class $\mathcal{C}^2$,
$\mathcal{C}[0,T]$	continuous functions $[0,T] \to \mathbb{R}$,
$\mathcal{P}_{2N-1}$	polynomial functions of degree at most $2N-1$.

2 Numerical solution of the Fokker-Planck equation

2.1 The Gauss-Galerkin approximation method

To introduce the Fokker-Planck equation, we consider the stochastic differential equation:

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t\,, \quad 0 \le t \le T\,, X_0 \sim \mu_0\,, \tag{1}$$

where $(X_t)_{t\le T}$ takes values in $\mathbb{R}$; $(W_t)_{t\le T}$ is a real standard Wiener process independent of X_0. Let $a(x) = \sigma^2(x)$, $a'(x) = \mathrm{d}a(x)/\mathrm{d}x$, $b'(x) = \mathrm{d}b(x)/\mathrm{d}x$, we make the following assumptions:

(Hl) b, $\sigma : \mathbb{R} \to \mathbb{R}$, are measurable and bounded applications;

(H2) $a' \in L^{\infty}(\mathbb{R})$ and there exists $\underline{a} > 0$ such that $a(x) \ge \underline{a}$, for all $x \in \mathbb{R}$;

(H3) b' is measurable bounded, and a' is continuous.

Under Assumptions (Hl)-(H2), Equation (1) admits a unique solution in the weak sense [12]. Hypothesis (H3) will be used in the following to demonstrate the convergence of the approximation.

Let $\mu_t \in \mathcal{M}_1(\mathbb{R})$ be the distribution law of X_t on $\mathbb{R}$, for all $0 \le t \le T$:

$$\langle \mu_t, \varphi \rangle = \mathbb{E}\big(\varphi(X_t)\big)\,, \quad \forall \varphi \in \mathcal{C}_{\mathrm{b}}^{2}(\mathbb{R})\,,$$

where:

$$\langle \mu_t, \varphi \rangle := \int_{\mathbb{R}} \varphi(x)\,\mu_t(\mathrm{d}x)\,.$$

It results from the Itô's formula that $(\mu_t)_{t\le T}$ is a solution of the Fokker-Planck equation (written in weak form) :

$$\langle \mu_t, \varphi \rangle = \langle \mu_0, \varphi \rangle + \int_0^t \langle \mu_s, \mathcal{L}\varphi \rangle\,\mathrm{d}s\,, \quad 0 \le t \le T\,, \quad \forall \varphi \in \mathcal{C}_{\mathrm{b}}^{2}(\mathbb{R})\,, \tag{2}$$

where $\mathcal{L}$ denotes the the infinitesimal generator of the Markov process X_t:

$$\mathcal{L}\varphi(x) := b(x)\,\varphi'(x) + \tfrac{1}{2}\,a(x)\,\varphi''(x)\,.$$

For $N \in \mathbb{N}$ given, the Gauss-Galerkin approximation method consists in approximating $(\mu_t)_{t\leq T}$ by a family of probability measures $(\mu_t^N)_{t\leq T}$ of the form :

$$\mu_t^N(\mathrm{d}x) = \sum_{i=1}^{N} w_t^{(i)}\,\delta_{x_t^{(i)}}(\mathrm{d}x) \in \mathcal{M}_1(\mathbb{R})\,.$$

The functions $t \to w_t^{(i)}$, $x_t^{(i)}$ are determined by posing :

$$\langle \mu_t^N, \pi\rangle = \langle \mu_0, \pi\rangle + \int_0^t \langle \mu_s^N, \mathcal{L}\pi\rangle\,\mathrm{d}s\,, \quad 0 \leq t \leq T\,, \quad \forall \pi \in \mathcal{P}_{2N-1}\,; \tag{3}$$

Note that:

$$\langle \mu_0^N, \pi\rangle = \langle \mu_0, \pi\rangle\,, \quad \text{for all } \pi \in \mathcal{P}_{2N-1}\,,$$

i.e. μ_0^N is the N-points Gauss-Christoffel approximation of the μ_0 (see Section 4.1.1).

2.2 Convergence of the approximation

Under an additional assumption, we will establish a convergence result.

Lemma 2.1 (Billingsley [1]) *Let $\mu \in \mathcal{M}_+(\mathbb{R})$ with finite moments of all orders $m_n = \langle \mu, x^n\rangle$. Suppose that the power series :*

$$\sum_{n\in\mathbb{N}} \frac{\theta^n}{n!}\,m_n$$

admits a strictly positive radius of convergence, then if $\nu \in \mathcal{M}_+(\mathbb{R})$ is s.t. $\langle \nu, x^n\rangle = m_n$ for all n then $\mu = \nu$. In this case, we say that the moment problem for μ is well posed. As of now, we make the abuse of notation x^n to designate the polynomial function $x \to x^n$.

Let:

$$\begin{aligned} m_n(t) &= \langle \mu_t, x^n\rangle\,, & \dot{m}_n(t) &= \mathrm{d}m_n(t)/\mathrm{d}t\,,\\ m_n^N(t) &= \langle \mu_t^N, x^n\rangle\,, & \dot{m}_n^N(t) &= \mathrm{d}m_n^N(t)/\mathrm{d}t\,. \end{aligned}$$

We make the additional hypothesis:

(H4) $\limsup_{n\to\infty} \left(\dfrac{m_{2n}(0)}{(2n)!}\right)^{\frac{1}{2n}} < \infty\,.$

This assumption ensures in particular the existence of moments of all orders for X_0, and thus for X_t, for all $0 \leq t \leq T$. Moreover, by using the Cauchy criterion on the convergence of series, (H4) implies that the power series $\sum_{n\in\mathbb{N}}(\theta^n/n!)\,m_n(0)$ has a strictly positive radius of convergence, so that according to Lemma 2.1, μ_0 is the only nonnegative measure on $\mathbb{R}$ admitting $(m_n(0))_{n\in\mathbb{N}}$ as moments.

Theorem 2.2 *Under assumptions (Hl)-(H4) the Gauss-Galerkin approximation is convergent:*

$$\mu_t^N \underset{N\to\infty}{\Longrightarrow} \mu_t\,, \quad t \geq 0\,.$$

To prove this theorem, we use several lemmas. In the following we will reason for $t \in [0,T]$; all results will be true for any $T > 0$

Lemma 2.3 *There exist real numbers K_n, K'_n which do not depend on N such that :*

(i) $|m_n^N(t)| \leq K_n$, *for all* (n,N) *s.t.* $n \leq 2N-1$, $0 \leq t \leq T$,

(ii) $|\dot{m}_n^N(t)| \leq K'_n$, *for all* (n,N) *s.t.* $n \leq 2N-1$, $0 \leq t \leq T$,

(iii) the power series $\sum_{n\in\mathbb{N}}(\theta^n/n!)\,K_n$ *has a strictly positive radius of convergence.*

Proof We show that there exist K_n and K'_n such that :

(4) $\quad |m_n(t)| \leq K_n\,, \quad$ for all $n\,,\ 0 \leq t \leq T\,,$

(5) $\quad |\dot{m}_n(t)| \leq K'_n\,, \quad$ for all $n\,,\ 0 \leq t \leq T\,,$

(6) the power series $\sum_{n\in\mathbb{N}}(\theta^n/n!)\,K_n$ has a strictly positive radius of convergence.

Suppose that $|m_{2n-2}(t)| \leq K_{2n-2}$, for all $0 \leq t \leq T$, taking $\varphi(x) = x^{2n}$ in (2) leads to:

$$m_{2n}(t) \leq m_{2n}(0) + c\int_0^t \left(m_{2n}(s) + n^2\,K_{2n-2}\right)\mathrm{d}s\,.$$

Then using Gronwall's lemma:

$$m_{2n}(t) \leq c\left(m_{2n}(0) + n^2\,K_{2n-2}\right),$$

where c denotes a constant that depends on T, but not on n. Let K_0 be such that $|m_0(t)| \leq K_0$, for all $0 \leq t \leq T$, we define by recurrence :

$$K_{2n} = c\left(m_{2n}(0) + n^2\,K_{2n-2}\right), \tag{7}$$

then $|m_{2n}(t)| \leq K_{2n}$, for all n, $0 \leq t \leq T$. Moreover:

$$|x|^{2n-1} \leq \tfrac{1}{2}\left(\frac{x^{2n}}{2n} + 2\,n\,x^{2n-2}\right),$$

then we can choose:

$$K_{2n-1} = \tfrac{1}{2}\left(\frac{K_{2n}}{2n} + 2\,n\,K_{2n-2}\right),$$

and (4) is thus proved. By explicitly writing K_{2n} from (7) we can show (6). (5) is verified without difficulty. To establish the lemma it suffices to note that the above argument remains valid for moments $m_n^N(t)$ with the same constants K_n and K'_n. □

Lemma 2.4 *There exists a family of distribution laws $(\nu_t\,;\,0 \leq t \leq T)$, and a subsequence $(\nu_t^N\,;\,0 \leq t \leq T)_{n\in\mathbb{N}}$ extracted from $(\nu_t^N\,;\,0 \leq t \leq T)_{n\in\mathbb{N}}$, such that :*

$$\nu_t^N \underset{N\to\infty}{\Longrightarrow} \nu_t\,, \quad 0 \leq t \leq T\,.$$

Proof According to Lemma 2.3 (i)-(ii), for all n fixed, the family $(m_n^N(\,\cdot\,)\,;\, N > (n+1)/2)$ is bounded and equicontinuous in $\mathcal{C}[0,T]$, and therefore relatively compact. By a Cantor diagonalization procedure we show that there exists an increasing sequence of integers $(N_{n'})_{n'\in\mathbb{N}}$ and functions $m_n^* \in \mathcal{C}[0,T]$, such that:

$$m_n^{N_{n'}}(\,\cdot\,) \xrightarrow[n'\to\infty]{} m_n^*(\,\cdot\,) \quad \text{in } \mathcal{C}[0,T]\,,\ \forall n \in \mathbb{N}\,. \tag{8}$$

Moreover, we consider the following result [11]: Given a sequence of real numbers $(m_p)_{p\in\mathbb{N}}$, a necessary and sufficient condition for there to exist a non-negative measure which admits $(m_p)_{p\in\mathbb{N}}$ for moments, is that

$$\forall P \in \mathbb{N}\,,\ C_0, C_1, \ldots, C_P \in \mathbb{R}\ :\quad \Big(\sum_{p=0}^{P} C_p\, x^p \geq 0\,,\ \forall x \in \mathbb{R}\Big) \Rightarrow \Big(\sum_{p=0}^{P} C_p\, m_p \geq 0\Big)\,.$$

This last property is satisfied by $(m_n^{N_p}(t))_{n\in\mathbb{N}}$, so is preserved at the limit $p \to \infty$. According to (8), there exists a nonnegative measure ν_t which admits $(m_n^*(t))_{n\in\mathbb{N}}$ for moments, for all n and $0 \leq t \leq T$. Hence for $t \leq T$:

$$\langle \mu_t^{N_p}, x^n \rangle \xrightarrow[p\to\infty]{} \langle \nu_t, x^n \rangle \tag{9}$$

According to Lemma 2.3, the power series $\sum_{n\in\mathbb{N}} (\theta^n/n!)\, m_n^*(t)$ has a strictly positive radius of convergence ν_t is the only law on $\mathbb{R}$ which verifies (9) (see [1]), which makes it possible to assert that $\nu_t^{N_p} \Rightarrow \nu_t$ as $p \to \infty$ [2, p. 181]. □

Lemma 2.5 *Under the assumptions* (Hl)-(H3)*, the Fokker-Planck equation* (2) *has a unique solution* $t \to \mu_t$*, a function with values in* $\mathcal{M}_+(\mathbb{R})$*.*

Proof Using Itô's formula, we can easily verify that the law of X_t solves (2), hence the existence of a solution is proved. Let $\psi(\,\cdot\,,\,\cdot\,) \in \mathcal{C}_{\rm b}^{1,2}(\mathbb{R}_+ \times \mathbb{R})$ and $\tilde{\mu}$ a solution of (2) with values in $\mathcal{M}(\mathbb{R})$. Then,

$$\big\langle \tilde{\mu}_t, \psi(t,\,\cdot\,)\big\rangle = \big\langle \tilde{\mu}_0, \psi(0,\,\cdot\,)\big\rangle + \int_0^t \big\langle \tilde{\mu}_s, \partial_s\psi(s,\,\cdot\,) + \mathcal{L}\psi(s,\,\cdot\,)\big\rangle\, \mathrm{d}s\,. \tag{10}$$

Furthermore, we consider the backward partial differential equation:

$$\frac{\partial v(s,x)}{\partial s} + \mathcal{L}v(s,x) = 0\,,\ s < t\,,\ v(t,x) = \bar{v}(x)\,,\ \forall x \in \mathbb{R} \tag{11}$$

$(v'(s,x) := \partial v(s,x)/\partial s)$. According to the assumptions made, and using regularity theorems for solutions of parabolic PDEs [8] we have : for all $\bar{v} \in \mathcal{C}_{\rm c}^{\infty}(\mathbb{R})$, (11) admits a solution $v \in \mathcal{C}_{\rm b}^{1,2}([0,t]\times\mathbb{R})$. After taking the difference between two solutions, to prove uniqueness it suffices to check that if $\mu_0 = 0$ then $\tilde{\mu}_t = 0$ for $t \geq 0$.

Let $t \geq 0$ and $\bar{v} \in \mathcal{C}_{\rm c}^{\infty}(\mathbb{R})$, by (11) we associate to $\tilde{v}$ an application $v \in \mathcal{C}_{\rm b}^{1,2}([0,t]\times\mathbb{R})$. From (10), with $\mu_0 = 0$, and (11):

$$\big\langle \tilde{\mu}_t, v(t,\,\cdot\,)\big\rangle = \int_0^t \big\langle \tilde{\mu}_s, \partial_s v(s,\,\cdot\,) + \mathcal{L}v(s,\,\cdot\,)\big\rangle\, \mathrm{d}s = 0\,,$$

so that $\langle \tilde{\mu}_t, v(t)\rangle = \langle \tilde{\mu}_t, \bar{v}\rangle = 0$ for all $\bar{v} \in \mathcal{C}_{\rm c}^{\infty}(\mathbb{R})$, hence $\tilde{\mu}_t = 0$. □

Proof of Theorem 2.2 If we establish that

$$\langle \nu_t, \varphi \rangle = \langle \nu_0, \varphi \rangle + \int_0^t \langle \nu_s, \mathcal{L}\varphi \rangle \, \mathrm{d}s \,, \;\; t \leq T \,, \;\; \forall \varphi \in \mathcal{C}_{\mathrm{b}}^2(\mathbb{R}) \,, \tag{12}$$

where $(\nu_t)_{t \leq T}$ is the limit of a subsequence whose existence is guaranteed by Lemma 2.4, then by Lemma 2.5 we have $\nu_t = \mu_t$, for all $0 \leq t \leq T$. We deduce that a subsequence of μ_t^N converges to μ_t. But by redoing the demonstration, by uniqueness of the limit we show that the whole sequence converges. So we have to show that (12) is verified for all $\varphi \in \mathcal{C}_{\mathrm{b}}^2(\mathbb{R})$. We will consider several steps.

Step 1: Suppose that φ is a polynomial function of degree d. For all $N \geq (d+1)/2$:

$$\langle \nu_t^N, \varphi \rangle = \langle \mu_0^N, \varphi \rangle + \int_0^t \langle \nu_s^N, \mathcal{L}\varphi \rangle \, \mathrm{d}s \,,$$

so (12) is obtained by dominated convergence when $N \to \infty$.

Step 2: Suppose that $\varphi(x) = e^{\mathbf{i}\,\theta\,x}\,\pi(x)$, where $\theta \in \mathbb{R}$, π polynomial function, and $\mathbf{i}^2 = -1$. Let us first take $\varphi(x) = e^{\mathbf{i}\,\theta\,x}$, $|\theta| \leq \theta_1$, where θ_1 is the radius of convergence given by Lemma 2.3-(*iii*). Let $\varphi_n(x) = \sum_{k=0}^n (\mathbf{i}\theta x)^k/k!$, φ_n verifies (12), so when $n \to \infty$ we get $\Phi(\theta) = 0$ for all $|\theta| \leq \theta_1$ where:

$$\Phi(\theta) := \langle \nu_t, \varphi \rangle - \langle \mu_0, \varphi \rangle - \int_0^t \langle \nu_s, \mathcal{L}\varphi \rangle \mathrm{d}s \,.$$

Thus for all $j \geq 1$, $\Phi^{(j)}(\theta) = 0, |\theta| \leq \theta_1$, where $\Phi^{(j)}$ is the jth derivative of Φ w.r.t. θ, we deduce that (12) is true for any φ of the form $e^{\mathbf{i}\theta x}\,\pi(x)$, $|\theta| \leq \theta_1$, π polynomial function. Using the inequality :

$$\left| \exp\big(\mathbf{i}\,(\theta+\theta_1)\,x\big) + \exp\big(\mathbf{i}\,\theta_1 x\big) \sum_{k=0}^n \frac{(\mathbf{i}\,\theta\,x)^k}{k!} \right| \leq c\, \frac{|\theta|^{n+1}}{(n+1)!}\, |x|^{n+1} \,,$$

and by the same argument, we show that (12) is verified for any φ of the form form $e^{\mathbf{i}\theta x}\,\pi(x)$, $|\theta| \leq 2\,\theta_1$, π polynomial function, and recursively for all $\theta \in \mathbb{R}$. Thus, Step 2 is proved.

Step 3: Suppose that $\varphi \in \mathcal{C}^2(\mathbb{R})$ with compact support. There exists φ_n of the form:

$$\varphi_n(x) = \sum_{k=-n}^n a_k^n \exp(\mathbf{i}\, b_k^n\, x) \quad \text{such that} \quad \|\varphi^{(j)} - \varphi_n^{(j)}\|_\infty \underset{n\to\infty}{\longrightarrow} 0 \quad (j = 0, 1, 2) \,,$$

where $\varphi^{(j)}$ is the jth derivative of φ, (12) is verified for φ_n for all n, and therefore, by taking the limit $n \to \infty$, also for φ.

Step 4: Suppose $\varphi \in \mathcal{C}_{\rm b}^2(\mathbb{R})$. Let $\tau_n \in \mathcal{C}_{\rm b}^2(\mathbb{R})$ s.t.

$$\begin{aligned} &0 \le \tau_n^{(j)} \le 1\,,\ j = 0,1,2\,,\\ &\tau_n = 1 \text{ on } [-n,n]\,,\\ &\tau_n = 0 \text{ on } (-\infty,-n-1] \cup (n+1,\infty)\,. \end{aligned}$$

Then we can apply the previous step to $\varphi_n := \tau_n\,\varphi$ and by dominated convergence ($n \to \infty$) we prove that φ satisfies (12), which ends the proof. □

Remark 2.6 *We proved the conservation of the Cauchy criterion :*

$$\textit{if} \quad \lim_{n\to\infty} \Big(\frac{m_{2n}(0)}{(2\,n)!}\Big)^{\frac{1}{2n}} < \infty\,,\ \textit{then} \quad \overline{\lim}_{n\to\infty} \Big(\frac{m_{2n}(t)}{(2\,n)!}\Big)^{\frac{1}{2n}} < \infty\,,\ \forall t \in [0,T]\,.$$

This result will be used to prove convergence in the case of nonlinear filtering.

3 Numerical solution of the Zakai equation

3.1 Filtering with discrete time observation

We consider the system :

$$\begin{aligned} \mathrm{d}X_t &= b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t\,,\ X_0 \sim \mu_0\,,\\ y_k &= h(X_{t_k}) + v_k\,, \end{aligned}$$

where $0 \le t \le T$, $0 < t_1 < \cdots < t_K = T$ is a sequence of given instants, to simplify we take:

$$t_k = k\,\Delta\,,\ \text{with } \Delta = \frac{T}{K} \text{ for some } K \in \mathbb{N}\,.$$

$(X_t)_{t\le T}$, $(W_t)_{t\le T}$, $(y_k)_{k\le K}$ et $(v_k)_{k\le K}$ are processes with values in $\mathbb{R}$; $(v_k)_{k\le K}$ is a sequence of independent Gaussian variables, $v_k \sim N(0,R)$; $(W_t)_{t\le T}$ is a standard standard Wiener process independent of $(v_k)_{k\le K}$; X_0 is independent of $(W_t)_{t\le T}$ and $(v_k)_{k\le K}$. Note that the case where the observation y_k takes values in $\mathbb{R}^d$ is treated in exactly the same way.

Let us assume Hypotheses (Hl)-(H4) satisfied, as well as the hypothesis :

(H5) $h : \mathbb{R} \to \mathbb{R}$ is measurable and bounded.

X_t describes the evolution of a physical system, and y_k its discrete time observation. The filtering problem consists in determining η_t, the conditional law of X_t given $(y_k)_{k;t_k\le t} = (y_1,\ldots,y_{\lfloor t/\Delta\rfloor})$, that is:

$$\langle \nu_t, \varphi\rangle = \mathbb{E}\big(\varphi(X_t)\big|y_1,\ldots,y_{\lfloor t/\Delta\rfloor}\big)\,,\quad \forall \varphi \in \mathcal{C}_{\rm b}^2(\mathbb{R})\,,$$

where $\lfloor t/\Delta\rfloor$ is the integer part of t/Δ. Between two moments of observation, i.e. $t_{k-1} < t < t_k$, the evolution of η_t is described by the (weak form of the) Fokker-Planck equation :

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle \eta_t\,,\,\varphi\rangle = \langle \eta_t\,,\,\mathcal{L}\varphi\rangle\,,\quad \forall \varphi \in \mathcal{C}_{\rm b}^2(\mathbb{R})\,.$$

At the time of observation $t = t_k$, by using the Bayes formula:

$$\langle \eta_{t_k} \, , \, \varphi \rangle = \frac{\big\langle \eta_{t_k^-} \, , \, f(\,\cdot\,, y_k)\, \varphi \big\rangle}{\big\langle \eta_{t_k^-} \, , \, f(\,\cdot\,, y_k) \big\rangle} \, ,$$

where

$$\langle \eta_{t_k^-}, \varphi \rangle := \lim_{\substack{t \to t_k \\ t < t_k}} \langle \eta_t, \varphi \rangle \, ,$$

and $f(x, y)$ is the local likelihood function:

$$f(x, y) := \exp\Big(\tfrac{1}{R}\, h(x)\, y - \tfrac{1}{2R}\, h(x)^2 \Big) \, .$$

Thus $(\eta_t)_{t \leq T}$ is a solution of the equation :

$$\langle \eta_t, \varphi \rangle = \langle \mu_0, \varphi \rangle + \int_0^t \langle \eta_s, \mathcal{L}\varphi \rangle \, \mathrm{d}s + \sum_{k=1}^{\lfloor t/\Delta \rfloor} \left\{ \frac{\big\langle \eta_{t_k^-}, f(\,\cdot\,, y_k)\, \varphi \big\rangle}{\big\langle \eta_{t_k^-}, f(\,\cdot\,, y_k) \big\rangle} - \big\langle \eta_{t_k^-}, \varphi \big\rangle \right\} \, , \ \forall \varphi \in \mathcal{C}_{\mathrm{b}}^2(\mathbb{R}) \, . \tag{13}$$

We propose to approximate η_t by a probability measure of the form:

$$\eta_t^N(\mathrm{d}x) = \sum_{i=1}^N w_t^{(i)}\, \delta_{x_t^{(i)}}(\mathrm{d}x) \, ,$$

where the stochastic processes $w_t^{(i)}$ and $x_t^{(i)}$ are determined by posing:

$$\langle \eta_t^N, \pi \rangle = \langle \mu_0, \pi \rangle + \int_0^t \langle \eta_s^N, \mathcal{L}\pi \rangle \, \mathrm{d}s + \sum_{k=1}^{\lfloor t/\Delta \rfloor} \left\{ \frac{\big\langle \eta_{t_k^-}^N, f(\,\cdot\,, y_k)\, \pi \big\rangle}{\big\langle \eta_{t_k^-}^N, f(\,\cdot\,, y_k) \big\rangle} - \big\langle \eta_{t_k^-}^N, \pi \big\rangle \right\} \, , \ \forall \pi \in \mathcal{P}_{2N-1} \, . \tag{14}$$

Theorem 3.1 *Under assumptions* (Hl)-(H5)*, for any given trajectory* $(y_1, \ldots, y_K)$*, the Gauss-Galerkin approximation is convergent :* $\eta_t^N \Rightarrow \eta_t$ *as* $N \to \infty$*, for all* $0 \leq t \leq T$*.*

Proof Let $m_n(t) = \langle \eta_t, x^n \rangle$, let's assume that the hypotheses:

$$\eta_t^N \Rightarrow \eta_t \, , \ \text{as } N \to \infty \, , \tag{15}$$

$$\overline{\lim_{n \to \infty}} \Big(\frac{m_{2n}(t)}{(2n)!} \Big)^{\frac{1}{2n}} < \infty \tag{16}$$

are verified for $t = t_{k-1}$; we will show that (15)-(16) are verified for $t \in [t_{k-1}, t_k]$. To prove the theorem it will be enough for us to establish (15)-(16) for $t = 0$.

For $t \in (t_{k-1}, t_k)$, the evolution of η_t is described by the Fokker-Planck equation, we deduce from Theorem 2.2, and from (16) in $t = t_{k-1}$, that (15) is satisfied for all $t \in (t_{k-1}, t_k)$. Since:

$$\langle \eta_{t_k}^N, \varphi \rangle = \frac{\langle \eta_{t_k^-}^N, f(\,\cdot\,, y_k)\ \varphi \rangle}{\langle \eta_{t_k^-}^N, f(\,\cdot\,, y_k) \rangle}\,,$$

we deduce that (15) is also true for $t = t_k$. Moreover,

$$\begin{aligned}\overline{\lim_{n\to\infty}} \Big(\frac{m_{2n}(t_k)}{(2n)!} \Big)^{\frac{1}{2n}} &= \overline{\lim_{n\to\infty}} \Big(\frac{1}{(2n)!} \frac{\langle \eta_{t_k^-}, f(\,\cdot\,, y_k)\, x^{2n} \rangle}{\langle \eta_{t_k^-}, f(\,\cdot\,, y_k) \rangle} \Big)^{\frac{1}{2n}} \\ &\leq \overline{\lim_{n\to\infty}} \Big(\frac{m_{2n}(t_k^-)}{(2n)!} \Big)^{\frac{1}{2n}}\,.\end{aligned}$$

Using (15), for $t = t_{k-1}$, and Remark 2.6, we show that the latter expression is finite. We deduce that (16) is true for $t = t_k$. To end the demonstration, we just need to check (15)-(16) for $t = 0$. From Equation (14), η_0^N is the Gauss-Christoffel approximation of $\eta_0 = \mu_0$, and the convergence $\eta_0^N \Rightarrow \eta_0$ can be deduced from Theorem 2.2. By Moreover (16) in $t = 0$ is exactly Hypothesis (H4). □

3.2 Filtering with continuous time observation

We consider the nonlinear system :

$$\text{(17)} \qquad \begin{cases} dX_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t\,, & X_0 \sim \mu_0\,, \\ dY_t = h(X_t)\,\mathrm{d}t + \mathrm{d}V_t\,, & Y_0 = 0\,, \end{cases}$$

for $0 \leq t \leq T$. The assumptions of the previous sections are assumed to be satisfied. The observation $(Y_t)_{t\leq T}$ with values in $\mathbb{R}$, is here in continuous time, $(V_t)_{t\leq T}$ is a standard Wiener process independent of X_0 and $(W_t)_{t\leq T}$.

The filtering problem consists in determining ν_t the conditional distribution of X_t given $\mathcal{F}_t := \sigma(Y_s; s \leq t)$, that is:

$$\langle \nu_t, \varphi \rangle = \mathbb{E}\big(\varphi(X_t)\big|Y_s\,,\, 0 \leq s \leq t\big)\,, \quad \forall \varphi \in \mathcal{C}_{\mathrm{b}}^2(\mathbb{R})\,.$$

To characterize ν_t, we can use the method of the reference probability. Let $\mathring{\mathbb{P}}$ be the law determined by :

$$\frac{\mathrm{d}\mathring{\mathbb{P}}}{\mathrm{d}\mathbb{P}} = Z_T^{-1}\,, \quad \text{with} \quad Z_t = \exp \int_0^t \Big(h(X_s)\,\mathrm{d}Y_s - \tfrac{1}{2}\, h(X_s)^2\,\mathrm{d}s \Big)\,.$$

The computation of the conditional distribution of X_t given $\mathcal{F}_t$ under $\mathbb{P}$, is related to an expression computed under $\mathring{\mathbb{P}}$ by the Kallianpur-Striebel formula :

$$\mathbb{E}\big(\varphi(X_t)\big|\mathcal{F}_t\big) = \frac{\mathring{\mathbb{E}}\big(\varphi(X_t)\, Z_t\big|\mathcal{F}_t\big)}{\mathring{\mathbb{E}}(Z_t|\mathcal{F}_t)}\,.$$

We define $\tilde{\nu}_t$ the unnormalized conditional distribution of X_t given $\mathcal{F}_t$, by posing:

$$\langle \tilde{\nu}_t, \varphi \rangle := \overset{\circ}{\mathbb{E}}\big(\varphi(X_t)\, Z_t \big| \mathcal{F}_t\big)\,.$$

$\tilde{\nu}_t$ is a solution of the (weak form) Zakai equation:

$$\langle \tilde{\nu}_t, \varphi \rangle = \langle \mu_0, \varphi \rangle + \int_0^t \langle \tilde{\nu}_s, \mathcal{L}\varphi \rangle \,\mathrm{d}s + \int_0^t \langle \tilde{\nu}_s, h\,\varphi \rangle \,\mathrm{d}Y_s\,,\ \ \forall \varphi \in \mathcal{C}_{\mathrm{b}}^2(\mathbb{R})\,. \tag{18}$$

We can determine $\tilde{\nu}_t^N$ the Gauss-Galerkin approximation of $\tilde{\nu}_t$, but after discretization in time, the equation of $\tilde{\nu}_t^N$ involves only discrete time observations. It is therefore preferable to discretize the observation equation in (17) directly:

$$y_k = h(X_{t_k}) + v_k\,,$$

with $t_k = k\Delta$, and where $v_k := (V_{t_{k+1}} - V_{t_k})/\Delta$ and y_k is the approximation of $(Y_{t_{k+1}} - Y_{t_k})/\Delta$.

We define $\mathcal{F}_t^\Delta := \sigma(y_1, \ldots, y_{\lfloor t/\Delta \rfloor})$ et ${}^\Delta\nu_t$ the conditional distribution of X_t given $\mathcal{F}_t^\Delta$. As we saw in Section 3.1, the evolution of $({}^\Delta\nu_t)$ is described by the equation:

$$\begin{aligned}\langle {}^\Delta\nu_t, \varphi \rangle = \langle \mu_0, \varphi \rangle &+ \int_0^t \langle {}^\Delta\nu_s, \mathcal{L}\varphi \rangle \,\mathrm{d}s \\ &+ \sum_{k=1}^{\lfloor t/\Delta \rfloor} \left\{ \frac{\big\langle {}^\Delta\nu_{t_k^-}\,,\, f_\Delta(\,\cdot\,, y_k)\,\varphi \big\rangle}{\big\langle {}^\Delta\nu_{t_k^-}\,,\, f_\Delta(\,\cdot\,, y_k) \big\rangle} - \big\langle {}^\Delta\nu_{t_k^-}\,,\, \varphi \big\rangle \right\}\,, \quad \forall \varphi \in C_b^2(\mathbb{R})\,,\end{aligned} \tag{19}$$

with:

$$f_\Delta(x, y) := \exp\Big(h(x)\, y\, \Delta - \tfrac{1}{2}\, h(x)^2\, \Delta\Big)\,.$$

We have the following result:

Theorem 3.2 *In addition to assumptions (Hl)-(H5), suppose that $h \in \mathcal{C}_{\mathrm{b}}^2(\mathbb{R})$, then for any observed trajectory $(Y_s)_{s \le t}$:*

$${}^\Delta\nu_t \underset{\Delta \to 0}{\Longrightarrow} \nu_t\,, \quad 0 \le t \le T$$

(provided that ν_t is defined in "robust form" cf. for example [9]).

Proof Consider a probability space $(\Omega, \mathcal{F}, \overset{\circ}{\mathbb{P}})$ and the following SDE on this space:

$$\begin{cases} \mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t\,, \\ \mathrm{d}Y_t = \mathrm{d}\overset{\circ}{V}_t\,, \end{cases}$$

where $(W_t, \overset{\circ}{V}_t)$ is a standard Wiener process with values in $\mathbb{R} \times \mathbb{R}$ independent from X_0 and $Y_0 = 0$.

For Y we adopt the canonical representation $(\mathcal{C}[0,T], \mathcal{B}, \mathcal{W}, Y)$, i.e. $(\mathcal{C}[0,T], \mathcal{B})$ is the space of continuous functions $[0,T] \to \mathbb{R}$ equipped with the Borel σ-algebra $\mathcal{B}$, $\mathcal{W}$ is

the Wiener measure on this space and Y is the canonical process: for all $\omega \in \mathcal{C}[0,T]$, $Y_t(\omega) := \omega(t)$. Moreover, let $\bar{\mathbb{P}}$ be the marginal distribution of X on a space $(\bar{\Omega}, \bar{\mathbb{P}}, \bar{\mathcal{F}})$.

Under $\overset{\circ}{\mathbb{P}}$, X and Y are independent:

$$\overset{\circ}{\mathbb{P}}(\mathrm{d}X, \mathrm{d}Y) = \bar{\mathbb{P}}(\mathrm{d}X) \times \mathcal{W}(\mathrm{d}Y)\,. \tag{20}$$

Let

$$\Delta_K := \frac{T}{K}\,,$$

and $t_k^K := k\,\Delta_K$ which we will denote t_k. Define also:

$$h^K(t,x) := h(t_k)\,, \quad \text{for } t \in [t_k, t_{k+1})\,.$$

Consider the following $\overset{\circ}{\mathbb{P}}$-exponential martingales:

$$\begin{cases} Z_t := \exp \displaystyle\int_0^t \Big(h(X_s)\,\mathrm{d}Y_s - \tfrac{1}{2}\,h(X_s)^2\,\mathrm{d}s \Big)\,, \\ Z_t^K := \exp \displaystyle\int_0^t \Big(h^K(X_s)\,\mathrm{d}Y_s - \tfrac{1}{2}\,h^K(X_s)^2\,\mathrm{d}s \Big)\,. \end{cases}$$

Let:

$$\mathrm{d}M_s := \mathcal{L}h(X_s)\,\mathrm{d}s + (h'\,\sigma)(X_s)\,\mathrm{d}W_s\,,$$

integration by part in the Itô integral leads to:

$$Z_t = \exp\Big(h(X_t)\,Y_t - \int_0^t \big(h(X_s)\,\mathrm{d}M_s - \tfrac{1}{2}\,h(X_s)^2\,\mathrm{d}s \big) \Big)\,. \tag{21}$$

In addition, as $t \to h^K(t,x)$ is piecewise constant:

$$\begin{aligned} Z_t^K = \exp\Big(& h(X_{t_k})\,(Y_t - Y_{t_k}) - \tfrac{1}{2} h(X_{t_k})^2\,(t - t_k) \\ & + \sum_{j=0}^{k-1} \big\{ h(X_{t_j})\,(Y_{t_{j+1}} - Y_{t_j}) - \tfrac{1}{2}\,h(X_{t_j})^2\,\Delta_k \big\} \Big)\,, \quad \text{for all } t \in [t_k, t_{k+1})\,. \end{aligned} \tag{22}$$

Representations (21) and (22) allow to consider Z_t and Z_t^K for any fixed trajectory of Y.

We define the distribution:

$$\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\overset{\circ}{\mathbb{P}}} := Z_T\,, \qquad \frac{\mathrm{d}\mathbb{P}^K}{\mathrm{d}\overset{\circ}{\mathbb{P}}} := Z_T^K\,, \tag{23}$$

and $\mathbb{E}$, $\mathbb{E}^K$ the associated expectations. Under $\mathbb{P}$ (resp. $\mathbb{P}^K$), (X,Y) admits the representation:

$$\begin{cases} \mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t\,, \\ \mathrm{d}Y_t = h(X_t)\,\mathrm{d}t + \mathrm{d}V_t\,, \end{cases} \qquad \big(\text{resp. } \mathrm{d}Y_t = h^K(X_t)\,\mathrm{d}t + \mathrm{d}V_t^K\big)$$

where V (resp. V^K) is a $\mathbb{P}$ standard Wiener process (resp. $\mathbb{P}^K$ standard Wiener process) defined by:

$$V_t := \overset{\circ}{V}_t + \int_0^t h(X_s)\,\mathrm{d}s \qquad \left(\text{resp. } V_t^K := \overset{\circ}{V}_t + \int_0^t h^K(s, X_s)\,\mathrm{d}s\right).$$

Consider now the system with discrete time observation:

$$\begin{cases} \mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t\,, \\ y_k^K = h(X_{t_k}) + v_k^K\,, \end{cases}$$

with

$$v_k^K := \tfrac{1}{\Delta_k}\left(V_{t_{k+1}} - V_{t_k}\right).$$

Clearly, under $\mathbb{P}^K$, the conditional distribution of X_t given $\sigma(y_k^K\,;\, k \text{ s.t. } t_k \le t)$ is equal to the conditional distribution of X_t given $\mathcal{F}_t := \sigma(Y_s\,;\, s \le t)$. Our goal is therefore to demonstrate the convergence of expressions $\mathbb{E}^K\big(\varphi(X_t)\big|\mathcal{F}_t\big)$ for any continuous and bounded function φ.

Thanks to the Kallianpur-Striebel formula, (23) gives:

$$\mathbb{E}\big(\varphi(X_t)\big|\mathcal{F}_t\big) = \frac{\mathbb{E}\big(\varphi(X_t)\,Z_t\big|\mathcal{F}_t\big)}{\mathbb{E}\big(Z_t\big|\mathcal{F}_t\big)}\,, \quad \mathbb{E}^K\big(\varphi(X_t)\big|\mathcal{F}_t\big) = \frac{\mathbb{E}\big(\varphi(X_t)\,Z_t^K\big|\mathcal{F}_t\big)}{\mathbb{E}\big(Z_t^K\big|\mathcal{F}_t\big)}\,, \quad \overset{\circ}{\mathbb{P}}\text{-a.s.}\,.$$

But, according to (20):

$$\mathbb{E}\big(\varphi(X_t)\,Z_t\big|\mathcal{F}_t\big) = \bar{\mathbb{E}}\big(\varphi(X_t)\,Z_t\big)\,, \quad \mathbb{E}\big(\varphi(X_t)\,Z_t^K\big|\mathcal{F}_t\big) = \bar{\mathbb{E}}\big(\varphi(X_t)\,Z_t^K\big)\,, \quad \mathcal{W}\text{-a.s.}\,.$$

For a given trajectory $(Y_s\,;\, s \le t)$ of the observation process, it is thus necessary to prove:

$$\mathbb{E}\big(\varphi(X_t)\,Z_t^K\big|\mathcal{F}_t\big) \xrightarrow[K\to\infty]{} \mathbb{E}\big(\varphi(X_t)\,Z_t\big|\mathcal{F}_t\big) \quad \text{a.s.}\,.$$

Since φ is bounded, it is sufficient to show that:

$$Z_t^K \xrightarrow[K\to\infty]{} Z_t \quad \text{in } L^1(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}}) \tag{24}$$

For any given t, Z_t^K and Z_t are positive random variables with mean 1, for all K, so a sufficient condition for (24) is:

$$Z_t^K \xrightarrow[K\to\infty]{} Z_t \quad \text{in } \bar{\mathbb{P}}\text{-probability}\,,$$

this result can be deduced from definitions (21) and (22) of Z_t and Z_t^K, which completes the proof of Theorem 3.2. □

We use the Gauss-Galerkin method to approximate ${}^{\Delta}\nu_t$ by a probability measure ${}^{\Delta}\nu_t^N$ of the form:

$${}^{\Delta}\nu_t^N(\mathrm{d}x) = \sum_{i=1}^N w_t^{(i)}\,\delta_{x_t^{(i)}}(\mathrm{d}x)\,,$$

where the stochastic stochastic processes $(w_t^{(i)})_{t\leq T}$ and $(x_t^{(i)})_{t\leq T}$, which depend on Δ and N, are determined by posing:

$$\langle {}^{\Delta}\nu_t^N, \pi\rangle = \langle \mu_0, \pi\rangle + \int_0^t \langle {}^{\Delta}\nu_s^N, \mathcal{L}\pi\rangle \, \mathrm{d}s + \sum_{k=1}^{\lfloor t/\Delta\rfloor} \frac{\left\langle {}^{\Delta}\nu_{t_k^-}^N \, , \, \big(f_\Delta(\,\cdot\,, y_k) - 1\big)\,\pi\right\rangle}{\left\langle {}^{\Delta}\nu_{t_k^-}^N \, , \, f_\Delta(\,\cdot\,, y_k)\right\rangle} \, , \ \forall \pi \in \mathcal{P}_{2N-1} \, .$$

According to Theorem 3.1, for any Δ, we have the following convergence:

$$ {}^{\Delta}\nu_t^N(\omega) \underset{N\to\infty}{\Longrightarrow} {}^{\Delta}\nu_t(\omega), \quad \text{for almost all } \omega, \text{ and } 0 \leq t \leq T \, . \tag{25}$$

Let $(f_p)_{p\in\mathbb{N}}$ be a dense sequence in $\mathcal{C}_{\mathrm{u}}(\mathbb{R})$, the set of bounded and uniformly continuous functions. We define :

$$d(\mu,\nu) := \sum_{p\in\mathbb{N}} \frac{1}{2^p} \frac{\big|\langle \mu, f_p\rangle - \langle \nu, f_p\rangle\big|}{\|f_p\|_\infty} \, ,$$

with $\|f\|_\infty = \sup\{|f(x)|; x\in\mathbb{R}\}$; $d(\,\cdot\,,\,\cdot\,)$ is a metric on $\mathcal{M}_+(\mathbb{R})$, which induces a topology equivalent to the one induced by the weak convergence of measures [12]. Thus (25) and Theorem 3.2 implies that for all $\Delta > 0$ we can associate $N(\Delta) \in \mathbb{N}$ such that:

$$ {}^{\Delta}\nu_t^{N(\Delta)}(\omega) \underset{\Delta\to 0}{\Longrightarrow} \nu_t(\omega) \, , \quad \text{for all } \omega \text{ a.s. and } 0 \leq t \leq T \, .$$

This last convergence result is not entirely satisfactory, we do not know how to explicitly choose $N(\Delta)$, but as in practice the observation equation is always in discrete time, for a given discretization step Δ the convergence (25) is satisfactory.

We could have obtained a root mean square convergence in the case of continuous time observations, however this is not of great interest.

To obtain a convergence for each observed trajectory, one could think of using the "robust form" of the Zakai equation, this was not feasible, because the multiplication by $\exp(-h(x)\,Y_t)$ brings out of the space of polynomials of degree at most equal to $2\,N-1$.

4 Numerical study

4.1 Presentation of the algorithm

We will use the following notations:

$(w_k^{(1:N)}, x_k^{(1:N)})$	*will denote*	$(w_k^{(i)}, x_k^{(i)})_{i=1,\dots,N}$,
$w_{0:K}^{(i)}$	"	$(w_k^{(i)})_{k=0,\dots,K}$,
$w_k^{(1:N)} = \tilde{w}_k^{(1:N)}$	"	$w_k^{(i)} = \tilde{w}_k^{(i)}$ *for* $i = 1,\dots,N$,
$i = 1:N$	"	$i = 1,\dots,N$, *etc.*

4.1.1 A reminder on Gauss-Christoffel quadrature methods

All the results of this section come from Wheeler [13] and Gautschi [6]. Given $N \in \mathbb{N}$ and a nonnegative measure $\nu \in \mathcal{M}_+(\mathbb{R})$, we want to find $(w_{1:N}, x_{1:N})$ such that:

$$\sum_{i=1}^{N} w_i \, \pi(x_i) = \langle \nu, \pi \rangle \,, \quad \forall \pi \in \mathcal{P}_{2N-1} \,. \tag{26}$$

It is well known that, if $x \to \nu((-\infty, x])$ admits at least N increasing points, then (26) admits a unique solution, where the particle x_i are two by two distinct and the weights are strictly positive, $w_i > 0$. The empirical measure:

$$\nu^N \stackrel{\text{def}}{=} \sum_{i=1}^{N} w_i \, \delta_{x_i}$$

is the Gauss-Christoffel approximation of ν; ν^N and ν have the same $2N$ first moments:

$$\langle \nu^N, x^p \rangle = m_p := \langle \nu, x^p \rangle \,, \quad p = 0 : 2N-1 \,.$$

To compute $(w_{1:N}, x_{1:N})$ from the moments $m_{0:2N-1}$ we use a classical method. We introduce $\boldsymbol{\pi}_{0:2N-1}$, the family of orthogonal polynomial functions relative to the measure ν, i.e. $\boldsymbol{\pi}_p$ is of degree p and $\langle \nu, \boldsymbol{\pi}_p \, \boldsymbol{\pi}_q \rangle = 0$ if $p \neq q$. These polynomial functions are defined up to a multiplicative constant, we can decide for example that the coefficient of the highest degree monomial in $\boldsymbol{\pi}_p$ is 1. In this case the family $\boldsymbol{\pi}_{0:2N-1}$ satisfies a recurrence relation of the form:

$$\begin{aligned}
\boldsymbol{\pi}_{-1}(x) &= 0 \,, && \text{(by convention)} \,, \\
\boldsymbol{\pi}_0(x) &= 1 \,, && \\
\boldsymbol{\pi}_{p+1}(x) &= (x - \alpha_p) \, \boldsymbol{\pi}_p(x) - \beta_p \, \boldsymbol{\pi}_{p-1}(x) \,, && p = 0 : 2N-2 \,,
\end{aligned}$$

for all $x \in \mathbb{R}$, for some $(\alpha_{0:2N-2}, \beta_{0:2N-2}) \in \mathbb{R}^{2(2N-1)}$ with $\beta_p > 0$ for $p \geqq 1$ and $\beta_0 = 0$.

The calculation of $(w_{1:N}, x_{1:N})$ is reduced to the calculation of the coefficients $(\alpha_{0:N-1}$, $\beta_{0:N-1})$, with by convention $\beta_0 = 0$, in the following way, let:

$$J_N = \begin{pmatrix}
\alpha_0 & \sqrt{\beta_1} & & & \\
\sqrt{\beta_1} & \alpha_1 & \sqrt{\beta_2} & & (0) \\
& \ddots & \ddots & \ddots & \\
& & \sqrt{\beta_{N-2}} & \alpha_{N-2} & \sqrt{\beta_{N-1}} \\
& (0) & & \sqrt{\beta_{N-1}} & \alpha_{N-1}
\end{pmatrix} . \tag{27}$$

J_N has N real eigenvalues $\lambda_{1:N}$, two by two distinct; let $\boldsymbol{v}_{1:N}$ the respectively associated orthonormal eigenvectors. We have the following result :

$$(w_{1:N}, x_{1:N}) = (\boldsymbol{v}^2_{1,1:N}, \lambda_{1:N}) \tag{28}$$

where $\boldsymbol{v}_{1,i}$ denotes the first component of the vector $\boldsymbol{v}_i$.

Next, we have to notice that using the standard moments $m_p = \langle \nu, x^p \rangle$ is numerically not a good idea, it leads to ill-conditioned algorithms. A classical method is instead to use *modified moments*, that is:

$$\tilde{m}_p \stackrel{\text{def}}{=} \langle \nu, \tilde{\boldsymbol{\pi}}_p \rangle \,, \quad p = 0 : 2N-1 \,.$$

where $\tilde{\boldsymbol{\pi}}_{0:2N-1}$ is a *given* basis of $\mathcal{P}_{2N-1}$ formed by orthogonal vectors, these kind of polynomial functions are defined by a recurrence:

$$\begin{aligned} \tilde{\boldsymbol{\pi}}_{-1}(x) &= 0 \,, && \text{(by convention)} \,, \\ \tilde{\boldsymbol{\pi}}_{0}(x) &= 1 \,, \\ \tilde{\boldsymbol{\pi}}_{p+1}(x) &= (x - \tilde{\alpha}_p)\, \tilde{\boldsymbol{\pi}}_p(x) - \tilde{\beta}_p\, \tilde{\boldsymbol{\pi}}_{p-1}(x) \,, && p = 1 : 2N-2 \,, \end{aligned}$$

(for all $x \in \mathbb{R}$) where the recurrence coefficients $(\tilde{\alpha}_{0:2N-2}, \tilde{\beta}_{0:2N-2})$ are given with $\tilde{\beta}_p > 0$ for $p \geqq 1$ and $\tilde{\beta}_0 \equiv 0$. In practice, we can use the Hermite polynomials.

Initialization

$\sigma_{-1,0} \leftarrow 0$

$\sigma_{0,p} \leftarrow \tilde{m}_p \,, \quad p = 0 : 2N-1$

$\alpha_0 \leftarrow \tilde{\alpha}_0 + \tilde{m}_1 / \tilde{m}_0$

$\beta_0 \leftarrow 0$

Iterations

for $p = 1 : N-1$ **do**

  for $q = p : 2N-p+1$ **do**

    $\sigma_{p,q} \leftarrow \sigma_{p-1,q+1} - (\alpha_{p-1} - \tilde{\alpha}_q)\, \sigma_{p-1,q} - \beta_{p-1}\, \sigma_{p-2,q} + \tilde{\beta}_q\, \sigma_{p-1,q-1}$

  end for

  $\alpha_p \leftarrow \tilde{\alpha}_p - \dfrac{\sigma_{p-1,p}}{\sigma_{p-1,p-1}} + \dfrac{\sigma_{p,p+1}}{\sigma_{p,p}}$

  $\beta_p \leftarrow \dfrac{\sigma_{p,p}}{\sigma_{p-1,p-1}}$

end for

Algorithm 1: *This modified Chebyshev algorithm allows us to compute* $(\alpha_{0:N-1},\, \beta_{0:N-1})$ *from* $(\tilde{m}_{0:2N-1}, \tilde{\alpha}_{0:2N-2},\, \tilde{\beta}_{0:2N-2})$*, see [13].*

Finally, the computation of $(\alpha_{0:N-1}, \beta_{0:N-1})$ from $(\tilde{m}_{0:2N-1}, \tilde{\alpha}_{0:2N-2}, \tilde{\beta}_{0:2N-2})$ is performed using the modified Chebyshev Algorithm 1.

4.1.2 Fokker-Planck equation

We first consider the approximation algorithm of the Fokker-Planck equation. The practical implementation of this algorithm requires a time discretization of the equation:

$$\langle \mu_t^N, \tilde{\boldsymbol{\pi}}_p \rangle = \langle \mu_0, \tilde{\boldsymbol{\pi}}_p \rangle + \int_0^t \langle \mu_s^N, \mathcal{L}\tilde{\boldsymbol{\pi}}_p \rangle \, \mathrm{d}s \,, \quad 0 \leq t \leq T \,, \quad p = 0 : 2N-1 \,, \tag{29}$$

where $\tilde{\boldsymbol{\pi}}_{0:2N-1}$ denotes a basis of $\mathcal{P}_{2N-1}$. In $t = 0$, (29) leads to the fact that μ_0^N is the Gauss-Christoffel approximation of μ_0 so that μ_0 can be replaced by μ_0^N.

All time discretization schemes could be considered, but in order to simplify the presentation we will use the Euler scheme with a time step $\delta = T/L$, with $L \in \mathbb{N}$. In order to simplify the notations, in the case of nonlinear filtering discussed later, we will assume that L is a multiple of K, so that the observation instants $t_k = k\,\Delta$ are included in $(\ell\delta)_{\ell=0:L}$. Also to simplify the notation, $\mu^N_{\ell\delta}$, $w^{(i)}_{\ell\delta}$ (etc.) will be noted μ^N_ℓ, $w^{(i)}_\ell$ (etc.).

The time-discretized equation (29) is thus written:

$$
\begin{aligned}
&\langle \mu^N_0, \tilde{\boldsymbol{\pi}}_p \rangle \leftarrow \langle \mu_0, \tilde{\boldsymbol{\pi}}_p \rangle\,, \quad p = 0 : 2N-1 \\
&\textbf{for } \ell = 1 : L \textbf{ do} \\
&\quad \langle \mu^N_\ell, \tilde{\boldsymbol{\pi}}_p \rangle \leftarrow \langle \mu^N_{\ell-1}, \tilde{\boldsymbol{\pi}}_p + \mathcal{L}\tilde{\boldsymbol{\pi}}_p\,\delta \rangle\,, \quad p = 0 : 2N-1 \\
&\textbf{end for}
\end{aligned}
\tag{30}
$$

where:

$$
\mu^N_\ell(\mathrm{d}x) = \sum_{i=1}^N w^{(i)}_\ell\, \delta_{x^{(i)}_\ell}(\mathrm{d}x)
$$

is the approximation of μ^N_t at time $t = \ell\,\delta$.

From $(w^{(1:N)}_{\ell-1}, x^{(1:N)}_{\ell-1})$, the recurrence (30) allows us to approximate the modified moments of μ^N_ℓ with respect to the basis $\tilde{\boldsymbol{\pi}}_{0:2N-1}$, that is :

$$
\tilde{m}_p(\ell) \simeq \tilde{m}_p(\ell-1) + \delta \sum_{i=1}^N w^{(i)}_{\ell-1}\, \mathcal{L}\tilde{\boldsymbol{\pi}}_p(x^{(i)}_{\ell-1}) \quad p = 0 : 2N-1\,.
$$

Given $\tilde{m}_{0:2N-1}(\ell)$, we now want to calculate $(w^{(1:N)}_\ell, x^{(1:N)}_\ell)$ such that :

$$
\sum_{i=1}^N w^{(i)}_\ell\, \tilde{\boldsymbol{\pi}}_p(x^{(i)}_\ell) = \tilde{m}_p(\ell)\,, \quad p = 0 : 2N-1\,. \tag{31}
$$

To solve this problem, we use the Gauss-Christoffel quadrature method presented in Section 4.1.1. The Gauss-Galerkin approximation algorithm, for the Fokker-Planck equation, is given by see Algorithm 2.

4.1.3 Nonlinear filtering equation

For the nonlinear filtering problem, we have to solve numerically an equation of the form (cf. (14) et (18)) :

$$
\begin{aligned}
\langle \nu^N_t, \tilde{\boldsymbol{\pi}}_p \rangle = \langle \mu_0, \tilde{\boldsymbol{\pi}}_p \rangle &+ \int_0^t \langle \nu^N_s, \mathcal{L}\tilde{\boldsymbol{\pi}}_p \rangle\, \mathrm{d}s \\
&+ \sum_{k=1}^{\lfloor t/\Delta \rfloor} \left\{ \frac{\left\langle \nu^N_{t_k^-}\,,\, f(\,\cdot\,, y_k)\,\tilde{\boldsymbol{\pi}}_p \right\rangle}{\left\langle \nu^N_{t_k^-}\,,\, f(\,\cdot\,, y_k) \right\rangle} - \left\langle \nu^N_{t_k^-}\,,\, \tilde{\boldsymbol{\pi}}_p \right\rangle \right\}\,, \quad p = 0 : 2N-1\,.
\end{aligned}
\tag{32}
$$

Inputs

$\tilde{\alpha}_p\,,\ \tilde{\beta}_p\,, \qquad p = 0 : 2N-2$

$\tilde{m}_p(0) := \langle \mu_0, \tilde{\boldsymbol{\pi}}_p \rangle\,, \qquad p = 0 : 2N-1$

Iterations

for $\ell = 0 : L$ **do**

Computation of $(\alpha_{0:N-1},\, \beta_{0:N-1})$ from $(\tilde{m}_{0:2N-1}(\ell), \tilde{\alpha}_{0:2N-2},\, \tilde{\beta}_{0:2N-2})$ (cf. Algo. 1)

Computation of eigenvalues and orthonormal eigenvectors of J_N defined in (27)

Computation of $(w_\ell^{(1:N)}, x_\ell^{(1:N)})$ from (28)

$\tilde{m}_p(\ell+1) \leftarrow \tilde{m}_p(\ell) + \delta \sum_{i=1}^N w_\ell^{(i)}\, \mathcal{L}\tilde{\boldsymbol{\pi}}_p(x_\ell^{(i)}), \quad p = 0 : 2N-1$

end for

Algorithm 2: *The Gauss-Galerkin approximation algorithm for the Fokker-Planck equation (2) presented with an Euler scheme (any other scheme can be used, see Section 4.1.4).*

Equation (32) after discretization using the Euler scheme, is written :

$\langle v_0^N, \tilde{\boldsymbol{\pi}}_p \rangle \leftarrow \langle \mu_0, \tilde{\boldsymbol{\pi}}_p \rangle$ for $p = 0 : 2N-1$

for $\ell = 1 : L$ **do**

$\langle \nu_\ell^N\,,\ \tilde{\boldsymbol{\pi}}_p \rangle \leftarrow \langle \nu_{\ell-1}^N\,,\ \tilde{\boldsymbol{\pi}}_p + \Delta\, \mathcal{L}\tilde{\boldsymbol{\pi}}_p \rangle$ for $p = 0 : 2N-1$ {prediction}

if (ℓ modulo $\frac{L}{K}$) $= 0$ **then**

$k \leftarrow \ell\, K/L$ {observation index}

$\langle \nu_\ell^N, \tilde{\boldsymbol{\pi}}_p \rangle \leftarrow \dfrac{\langle \nu_\ell^N\,,\ f(\,\cdot\,, y_k)\, \tilde{\boldsymbol{\pi}}_p \rangle}{\langle \nu_\ell^N\,,\ f(\,\cdot\,, y_k) \rangle}$ for $p = 0 : 2N-1$ {correction}

end if

end for

The complete algorithm is then equivalent to the one presented for the of Fokker-Planck.

4.1.4 Numerical tools

For the approximation of the Fokker-Planck equation and of the prediction part of the nonlinear filter, we use a Runge-Kutta algorithm of order 2; one could of course use more efficient schemes if the nature of the considered problem requires it.

In practice, the basis $\tilde{\boldsymbol{\pi}}_{0:2N-1}$ of $\mathcal{P}_{2N-1}$ used is that of the Hermite polynomial functions. For the computation of the eigenvalues of J_N, we used a variant of the QL algorithm for symmetric and tridiagonal matrices from the EISPACK software library [5].

4.2 Example

We present an example of application of the Gauss-Galerkin method in nonlinear filtering. The computations have been done on a VAX 730 computer in double precision FORTRAN 77. The approximation method applied to the Fokker-Planck equation on nonlinear examples gave good results up to $N = 10$ (N: number of Gauss points). Beyond that, we run into problems of ill-conditioning. For the filtering problem, we first tested the method on linear examples. We compared the results obtained with those given by a Kalman-Bucy

filter. Here again, we obtained good results, even with very few Gauss points ($N = 3$ or 4). We now present a numerical example; let us consider the nonlinear filtering problem:

$$\text{(33)} \qquad \begin{cases} \mathrm{d}X_t = -X_t\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t\,, & X_0 \sim N(0,1)\,, \\ \mathrm{d}Y_t = \exp(\mathbf{i}\,X_t)\,\mathrm{d}t + \rho\,\mathrm{d}V_t\,, & Y_0 = 0\,, \end{cases}$$

$0 \le t \le T$. The standard Wiener process $(V_t)_{t\le T}$ and the observation process $(Y_t)_{t\le T}$ take values in the complex plan ($\mathbf{i}^2 = -1$); $(W_t)_{t\le T}$ is a real standard Wiener process independent of $(V_t)_{t\le T}$; X_0 is independent of $(W_t)_{t\le T}$ and $(V_t)_{t\le T}$. Let ν_t be the the conditional distribution of X_t given $\mathcal{F}_t = \sigma(Y_s; s \le t)$. We implemented three methods of approximation of ν_t:

GGA **Gauss-Galerkin approximation :** ν_t is approximated by a distribution law of the form $\nu_t^{\mathsf{GGA}} = \sum_{i=1}^N w_t^{(i)}\,\delta_{x_t^{(i)}}$, where N is the number of Gauss points. The calculation of ν_t^{GGA} was presented in Section 4.1.3.

FD **Finite Differences :** we use a finite difference scheme in space, in order to solve numerically the Zakai equation for the unnormalized conditional density of ν_t. ν_t is thus approximated by a law ν_t^{FD} of the form $\nu_t^{\mathsf{FD}}(\mathrm{d}x) = p(t,x)\,\mathrm{d}x$; for details of this method cf. Le Gland [7].

EKF **Extended Kalman filter :** ν_t is approximated by the Gaussian distribution $\nu_t^{\mathsf{EKF}} = N(\hat{X}_t^{\mathsf{EKF}}, Q_t^{\mathsf{EKF}})$ where $\hat{X}_t^{\mathsf{EKF}}$ and Q_t^{EKF} are the outputs of the extended Kalman filter associated to (33).

Remarks 4.1 (*i*) *These three methods are in fact implemented after discretization in time of the system* (33).

(*ii*) *The initial condition X_0 as well as the Wiener processes (in discrete time: the Gaussian white noise) W_t and V_t have been simulated on a computer.*

(*iii*) *The* FD *method is used as a reference method: we will compare the conditional moments computed by* GGA *with those computed by* FD. *However,* FD *has the disadvantage that it cannot be applied in a simple way in the case where the support of ν_t does not remain, when t varies, in a bounded and fixed domain of $\mathbb{R}$. Indeed in* FD *the conditional density $p(t,x)$ is computed on a domain $[-M, M]$ fixed in advance.*

For the simulation we take $T = 10$ and $\Delta = 0.01$. In a first set of simulations we take $N = 10$ and $\rho = 0.5$, see Figures 1-4. In a second set of simulations we take $N = 2$ and $\rho = 1$, see Figure 5. In view of the numerical examples (two of which are presented at the end of the section) we can make several observations:

(*i*) The estimators $\hat{X}_t^{\mathsf{method}} := \langle \nu_t^{\mathsf{method}}, x\rangle$ ($\mathsf{method} = \mathsf{GGA}, \mathsf{FD}, \mathsf{EKF}$) of X_t, given by the three methods, are equivalent (cf. Fig. 1). On the other hand, contrary to GGA, EKF gives a poor estimate of the conditional variance $Q(t) := \langle \nu_t, x^2\rangle - \langle \nu_t, x\rangle^2$ (cf. Fig. 2).

(*ii*) GGA correctly follows the evolution of the conditional moments for the first set of parameters ($N = 10$, $\rho = 0.5$), the first 14 moments are estimated in a satisfactory way).

(*iii*) Even for a small number of Gauss points ($N = 2$ in the second set of parameters), GGA gives significant results (cf. Fig. 5).

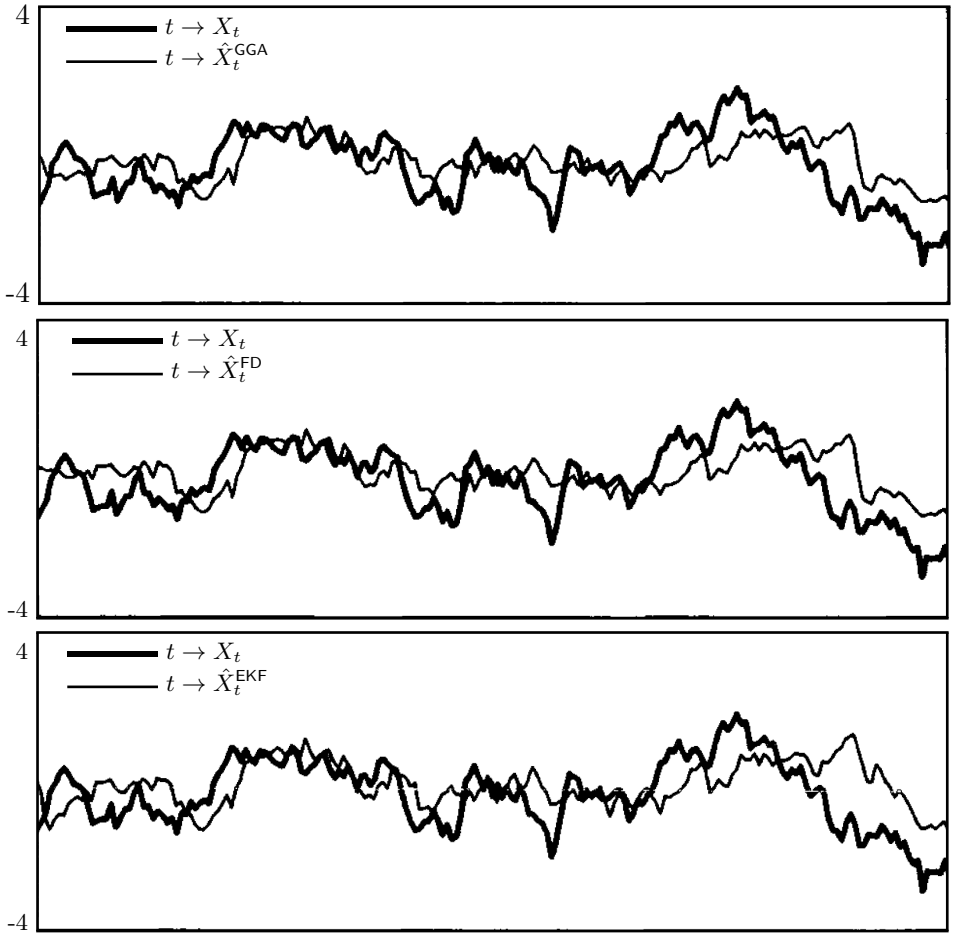

Figure 1: *First set of parameters ($N = 10$, $\rho = 2$); comparison of the real state trajectory $t \to X_t$ and of the estimators $t \to \hat{X}_t^{\mathsf{method}} := \langle \nu_t^{\mathsf{method}}, x \rangle$ with* $\mathsf{method} = \mathsf{GGA}, \mathsf{FD}, \mathsf{EKF}$.

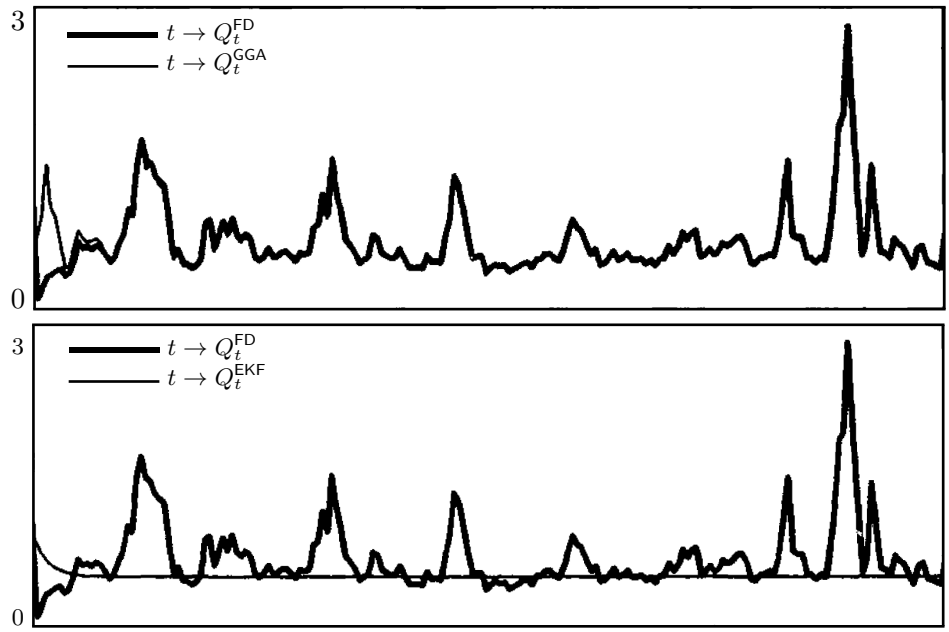

Figure 2: *First set of parameters ($N = 10$, $\rho = 2$); comparison of conditional variances, $t \to Q^{\mathsf{method}}(t) := \langle \nu_t^{\mathsf{method}}, x^2 \rangle - \langle \nu_t^{\mathsf{method}}, x \rangle^2$ with* method = GGA, FD, EKF.

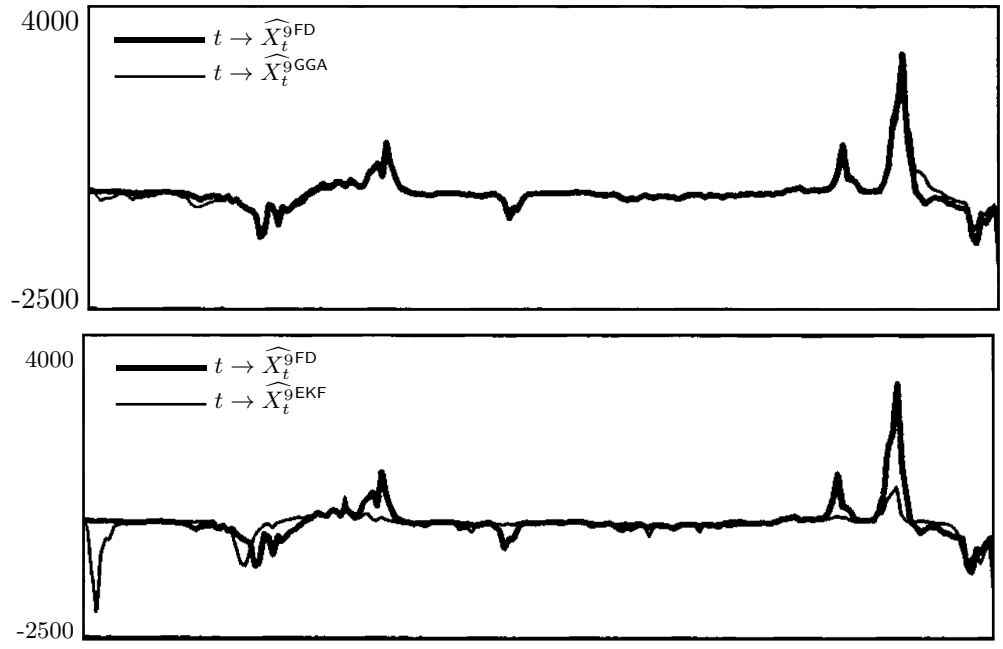

Figure 3: *First set of parameters ($N = 10$, $\rho = 2$); comparison of conditional moments of order 9, $t \to \widehat{X_t^9}^{\mathsf{method}} = \langle \nu_t^{\mathsf{method}}, x^9 \rangle$ with* method = GGA, FD, EKF.

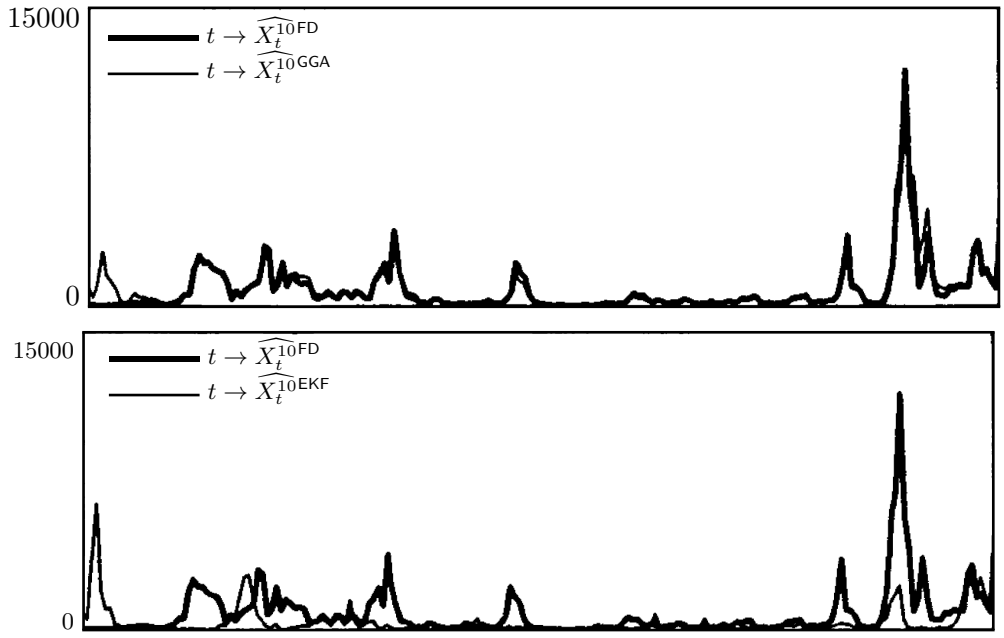

Figure 4: *First set of parameters ($N = 10$, $\rho = 2$); comparison of conditional moments of order 10,* $t \to \widehat{X_t^{10}}^{\mathsf{method}} = \langle \nu_t^{\mathsf{method}}, x^{10} \rangle$ *with* method = GGA, FD, EKF.

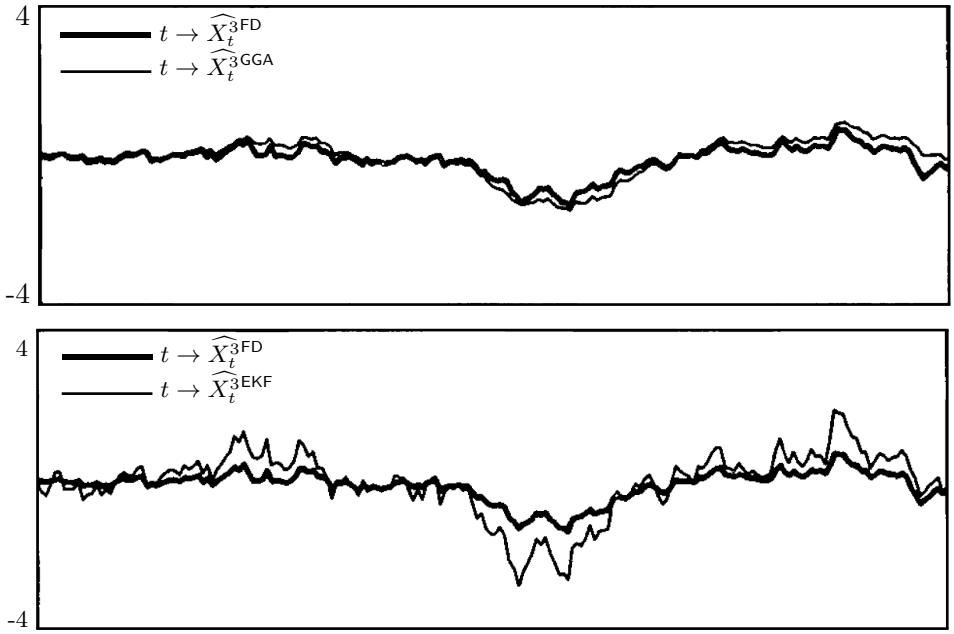

Figure 5: *Second set of parameters ($N = 2$, $\rho = 1$); comparison of conditional moments of order 3,* $t \to \widehat{X_t^{3}}^{\mathsf{method}} = \langle \nu_t^{\mathsf{method}}, x^{3} \rangle$ *with* method = GGA, FD, EKF.

References

[1] Patrick Billingsley. *Probability and Measure*. John Wiley & Sons, 1979.

[2] Leo Breiman. *Probability*. Classics in Applied Mathematics. Addison Wesley, Philadelphia, 1968. First edition in 1968.

[3] Fabien Campillo. *Filtrage et Détection de Ruptures de Processus Partiellement Observées*. PhD thesis, Thèse de Troisième Cycle, Université de Provence, Marseille, 1984.

[4] Donald A. Dawson. Galerkin approximation of nonlinear Markov processes. In *Statistics and related topics (Ottawa, Ont., 1980)*, pages 317–339. North-Holland, Amsterdam, 1981.

[5] Burton S. Garbow, James M. Boyle, Jack Dongarra, and Cleve B. Moler. *Matrix Eigensystem Routines - EISPACK Guide Extension*. Lecture Notes in Computer Science (LNCS, volume 51). Springer Verlag, 1977.

[6] Walter Gautschi. On generating orthogonal polynomials. *SIAM Journal on Scientific and Statistical Computing*, 3(3):289–317, 1982.

[7] François Le Gland. *Estimation de Paramètres dans les Processus Stochastiques, en Observation Incomplète — Applications à un Problème de Radio–Astronomie*. PhD thesis, Thèse de Docteur–Ingénieur, Université de Paris IX – Dauphine, 1981.

[8] Olga A. Ladyzhenskaya and Nina N. Uraltseva. *Linear and quasilinear elliptic equations*. Academic Press, 1968.

[9] Étienne Pardoux. Équations du filtrage non linéaire de la prédiction et du lissage. *Stochastics*, 6(3-4):193–231, 1982.

[10] Pierre-Arnaud Raviart. An analysis of particle methods. In F. Brezzi, editor, *Numerical Methods in Fluid Dynamics – Lectures given at the 3rd 1983 Session of the Centro Internationale Matematico Estivo (CIME) held at Como, Italy, July 7-15, 1983*, Lecture Notes in Mathematics vol. 1127. Springer Verlag, Berlin, 1985.

[11] James A. Shohat and Jacob D.Tamarkin. *The Problem of Moments*. American Mathematical Society, 1950.

[12] Daniel W. Stroock and S. R. Srinivasa Varadhan. *Multidimensional Diffusion Processes*. Springer-Verlag, 1979.

[13] John C. Wheeler. Modified moments and Gaussian quadratures. *Rocky Mountain Journal of Mathematics*, 4(2):287–296, 1974.

Addendum

This is the English translation of the paper [B]. A number of typos (many...) have been corrected, some notations and demonstrations have been clarified. Also some elements from the original works [3], that was not detailed or present in the 1986 version, such as the proof of Theorem 3.2, are developed here in order to obtain a self-contained version.

This article contains, to my knowledge, the first occurrence of the term **"particle approximation"** in the context of nonlinear filtering. Indeed, the conditional law η_t of the state given the observations is approximated by an empirical law of the form:

$$\eta_t(\mathrm{d}x) \simeq \eta_t^N(\mathrm{d}x) = \sum_{i=1}^N w_t^{(i)}\, \delta_{x_t^{(i)}}(\mathrm{d}x)\,,$$

where $w_t^{(i)} \geq 0$, $\sum_{i=1}^N w_t^{(i)} = 1$, and $\delta_{x_t^{(i)}}(\mathrm{d}x)$ is the Dirac measure on the particle $x_t^{(i)}$.

I coined this term in reference to the recent work at the time of Pierre-Arnaud Raviart on the approximation of solutions of first-order PDEs: "the exact solution is approximated by a linear combination of Dirac measures in the space variables" [C] and [10].

In the following years we proposed another particle approximation method in nonlinear filtering limited to the noise-free state equation case. In this case, the infinitesimal generator $\mathcal{L}$ is of first order making it possible to use the particle approximation methods proposed by P.A. Raviart [C]. Although proposed in a rather limited case, the proposed approximation method constitutes one of the premises of what will be called later "particle filtering" or "sequential Monte Carlo". In our approach, a crucial step was however missing, the famous bootstrap step ! This idea, in the context of nonlinear filtering, came to the table later, in the beginning of the 90s [E-F].

[A] F. Campillo. *La méthode d'approximation de Gauss-Galerkin – Application à l'équation du filtrage non linéaire*, Master Thesis, Université de Provence, 1982 [PDF]

[B] F. Campillo. *La méthode d'approximation de Gauss-Galerkin en filtrage non linéaire.* RAIRO M2AN, 20(2):203–223, 1986. [PDF]

[C] P.A. Raviart, *Particle approximation of first order systems*, Journal of Computational Mathematics, 1(4):50-61, 1986.

[D] F. Campillo, F. Legland, *Approximation particulaire en filtrage non linéaire. Application à la trajectographie*, 22ème Congrès National d'Analyse Numérique, Loctudy, 1990. [PDF]

[E] P. Del Moral, J.C. Noyer, G. Rigal, G. Salut, *Traitement non-linéaire du signal par réseau particulaire: Application radar*, 14ème Colloque sur le Traitement du Signal et des Images (GRETSI), Juan les Pins 1993.

[F] N.J Gordon, D.J. Salmond, A.F.M. Smith, *Novel approach to nonlinear/non–Gaussian Bayesian state estimation*, IEE Proceedings, Part F, 2(140):107–113, 1993.

In [B], I regrettably forgot to thank Walter Gautschi. Summer of 1984, a few months before the defense of my thesis, I indeed needed some additional elements concerning the Gauss quadrature methods using orthogonal polynomial functions. As Walter Gautschi was visiting Europe, I had invited him to Marseille. He completely clarified the situation for me. To thank him I proposed him to visit Aix-en-Provence... but my car broke down on the highway, Walter Gautschi finally got to visit Aix-en-Provence in record time ! I warmly thank him.