A Kusuoka-Lyons-Victoir particle filter

Dan Crisan* Salvador Ortiz-Latorre†

February 18, 2013

Abstract

The aim of this paper is to introduce a new numerical algorithm for solving the continuous time non-linear filtering problem. In particular, we present a particle filter that combines the Kusuoka-Lyons-Victoir cubature method on Wiener space (KLV) [13], [18] to approximate the law of the signal with a minimal variance "thining" method, called the tree based branching algorithm (TBBA) to keep the size of the cubature tree constant in time. The novelty of our approach resides in the adaptation of the TBBA algorithm to *simultaneously* control the computational effort and incorporate the observation data into the system. We provide the rate of convergence of the approximating particle filter in terms of the computational effort (number of particles) and the discretization grid mesh. Finally, we test the performance of the new algorithm on a benchmark problem (the Beneš filter).

Keywords: Cubature on Wiener space; particle filters; TBBA

1 Introduction

The main goal of stochastic filtering is to estimate the state of a dynamical system based on partial observation. We model the dynamical system by a stochastic process $X = \{X_t\}_{t\geq 0}$, called the *signal*. We do not observe the signal directly. Instead, we make use of the information provided by observing another process $Y = \{Y_t\}_{t\geq 0}$, called the *observation process*. The information process is at each instant of time a functional of the signal until that time and some measurement noise , that is, $Y_t = \Gamma\left(\{X_s\}_{s\leq t}, W_t\right)$, where $\{W_t\}_{t\geq 0}$ is another stochastic process modelling the noise. In mathematical terms the problem reduces to compute the following conditional expectation, $\mathbb{E}[\varphi\left(X_t\right)|\mathcal{Y}_t] = \int \varphi\left(x\right)\pi_t\left(dx\right), \varphi \in \mathcal{H}$, where $\mathcal{H}$ is a suitable space of test functions and $\mathcal{Y}_t \triangleq \sigma\{Y_s, s \in [0,t]\}$ is the filtration generated by the observation process. In other words, we are interested in computing the conditional distribution of the signal X_t given $\mathcal{Y}_t$, which can be viewed as a probability measure valued process $\pi = \{\pi_t\}_{t\geq 0}$. With notable exceptions (such as the Kalman-Bucy filter and the Beneš filter), π is an infinite-dimensional process. One can not find an analytical computable expression for π and has to rely on numerical approximations for inference purposes. In the

*Department of Mathematics, Imperial College London, 180 Queen's Gate, London, SW7 2AZ, United Kingdom.

†Centre of Mathematics for Applications, Oslo University, P.O. Box 1053 Blindern, N-0316 Oslo, Norway.

numerical experiments below we will make use of the explicit solution for the Beneš filter to test the accuracy of our algorithm.

A number of different numerical methods for solving the filtering problem, ranging from the solution of partial differential equation to Wiener chaos expansions, see for example Chapter 8 in [1]. One of the most successful approaches, which is widely used in practice, is the class of the particle approximations. In this approach, the conditional distribution π_t is approximated by the empirical distribution of a system of random weighted particles. The classical particle filter, first introduced by Gordon et al. [10] use a correction mechanism that eliminates, at particular times, the particles with small weights and multiply the ones with bigger weights, maintaining the total number of particles in the system constant. However, this procedure adds some randomness to the system, diminishing the accuracy of the approximation. Hence, it is desirable to use a technique that minimizes this undesired effect. In [6], Crisan and Lyons introduced the tree based branching algorithm (TBBA). This algorithm satisfies a minimal variance property which allow to perform, in a sense optimally, the correction in the particle system. Another aspect of these branching particle filters is the choice of the resampling times. Most of the theoretical results assume fixed deterministic times of resampling and this is the approach that we will follow. Nevertheless, in practice these times are randomly selected in terms of some overall characteristics of the particle systems, see [9] and [8] for a theoretical study of this problem.

In the standard particle filter, the component particles follow the law of the signal. Usually, we model the signal by means of a stochastic differential equation (SDE) driven by a Brownian motion. A classical result, tells us that $\pi_t(\varphi)$ can be expressed as the expected value of a functional of the signal parametrised by the given observation path. Naturally, an efficient approximation of the law of the signal would give a good approximation of $\pi_t(\varphi)$. In recent years, Kusuoka [13],[14] and Lyons and Victoir [18] among others, have introduced high order schemes for solving SDEs, known as cubature on Wiener space or KLV methods. Surprisingly, these methods are essentially deterministic. They involve the construction of a discrete (deterministic) measure with support given by leaves of an n-ary tree, with the nodes being obtained by solving ordinary differential equations (ODEs). Unfortunately, this tree-like structure makes the number of ODEs to be solved increase exponentially. In order to counter this feature, the KLV cubature methods can be combined with a partial sampling procedure, particularly useful when the dimension of the SDE to solve is high or the final time is large. The use of cubature methods for solving the stochastic filtering problem has been suggested in [5] and [17], and the area of application of these methods is expanding continuously, see for instance [7], where they are used to solve backward SDEs.

In this paper we present a new numerical algorithm to solve the nonlinear stochastic filtering problem. This algorithm is based on a combination of the Kusuoka-Lyons-Victoir (KLV) method and the tree based branching algorithm (TBBA). The KLV method is used to compute a high order approximation of the law of the signal X, whilst the TBBA is used to partially prune the KLV tree in a coherent manner. In our approach the weights of the TBBA are computed taken into account both, the cubature weights (the weights of the discrete measure) and the likelihood weights. In this way, we can simultaneously control the computational effort at each time step and mitigate the sample degeneracy.

The paper is organised as follows. In the next section we introduce some basic notation on multi-indices and vector fields necessary to present the cubature on Wiener space. In Section 3, we introduce the basic results on cubature, in particular we give the main bound on the local and global error of the method. Section 4 is devoted to the detailed description of the filtering problem. In addition, we also introduce the Crisan-Ghazali approach to apply cubature methods on filtering. In Section 5, we recall the TBBA algorithm, recall the basic properties of the random variables generated by the method and describe in detail the construction of the associated trees. In Section 6 we introduce the new algorithm and prove the convergence of the particle approximation. A variation of the algorithm where the likelihood weights are not taking into account when pruning the KLV-tree is also introduced. Finally, in Section 7 we test the new algorithm on the Beneš filter.

2 Basic notation and preliminaries

Here we introduce some basic notation on vector fields and multi-indices used to present the Stratonovich Taylor expansion and the main results on cubature.

2.1 Multi-indices

Let $p \in \mathbb{N}$, and let $\mathcal{A}$ be the set of all multi-indices with values in $\{0, ..., p\}$, that is, $\mathcal{A} \triangleq \{\varnothing\} \cup \bigcup_{k=1}^{\infty} \{0, ..., p\}^k$. For any (non-empty) multi-index $\alpha = (\alpha_1, ..., \alpha_k) \in \mathcal{A}$, define its length by $|\alpha| = k$ and its degree by $\|\alpha\| = k + card\{j : \alpha_j = 0\}$. We also define the subsets of $\mathcal{A}, \mathcal{A}(j) \triangleq \{\alpha \in \mathcal{A} : \|\alpha\| \leq j\}$ and $\mathcal{A}_1(j) \triangleq \{\alpha \in \mathcal{A} \backslash \{\varnothing, (0)\} : \|\alpha\| \leq j\}$. We will write $-\alpha = (\alpha_2, ..., \alpha_k)$ and $\alpha- = (\alpha_1, ..., \alpha_{k-1})$. Given two multi-indices $\alpha = (\alpha_1, ..., \alpha_k)$ and $\beta = (\beta_1, ..., \beta_l)$ we define their concatenation as $\alpha * \beta = (\alpha_1, ..., \alpha_k, \beta_1, ..., \beta_l)$. For any $\alpha = (\alpha_1, ..., \alpha_k) \in \mathcal{A}$, we also define $\alpha[i]$, the truncated index of length $i = 1, ..., k$, by $\alpha[i] \triangleq (\alpha_1, ..., \alpha_i)$.

2.2 Vector fields

Let $C_b^\infty\left(\mathbb{R}^d; \mathbb{R}^d\right)$ denote the space of $\mathbb{R}^d$-valued infinitely differentiable bounded functions defined on $\mathbb{R}^d$ whose derivatives of any order are bounded. Recall that $V \in C_b^\infty\left(\mathbb{R}^d; \mathbb{R}^d\right)$ can be viewed as a vector field (or a first order differential operator) on $\mathbb{R}^d$, i.e., $V(f) = \sum_{j=1}^d V^j \frac{\partial}{\partial x_j} f$, where V^j is the jth coordinate function of V and $f \in C_b^\infty\left(\mathbb{R}^d; \mathbb{R}\right)$. Given $V, W \in C_b^\infty\left(\mathbb{R}^d; \mathbb{R}^d\right)$, the composition operator is defined by $V \circ W(f) = \sum_{j=1}^d V^j \frac{\partial}{\partial x_j} \left(\sum_{i=1}^d W^i \frac{\partial}{\partial x_i} f\right)$, for $f \in C_b^\infty\left(\mathbb{R}^d; \mathbb{R}\right)$.We also define the Lie bracket of vector fields by $[V, W](f) = V \circ W(f) - W \circ V(f)$.

Given a family of vector fields $\mathbb{V} = \{V_0, V_1, ..., V_p\} \in C_b^\infty\left(\mathbb{R}^d; \mathbb{R}^d\right), p \in \mathbb{N}$, we define the vector field concatenation $V_{[\alpha]}, \alpha \in \mathcal{A}$, as follows: $V_{[\varnothing]} = 0, V_{[i]} = V_i, V_{[\alpha * i]} = [V_\alpha, V_i], i = 0, ..., p$. Note that V_α will stand for the usual composition of vector fields, that is, $V_\alpha = V_{\alpha_1} \circ \cdots \circ V_{\alpha_k}$.

2.3 Stratonovich Taylor expansion

Consider the probability space $(\Omega, \mathcal{F}, P) = (C_0([0,T], \mathbb{R}^p), \mathcal{B}(C_0([0,T], \mathbb{R}^p)), \mathbb{P})$, where $C_0([0,T], \mathbb{R}^p)$ is the space of $\mathbb{R}^p$-valued continuous functions starting at 0, $\mathcal{B}(C_0([0,T], \mathbb{R}^p))$ its Borel σ-algebra and $\mathbb{P}$ the Wiener Measure. Also consider the coordinate mapping process $B_t^j(\omega) = \omega^j(t), t \in [0,T], \omega \in \Omega$, which under $\mathbb{P}$ is a Brownian motion starting at 0. For $\omega \in \Omega$, we make the convention $\omega^0(t) = t$ and $B_t^0(\omega) = t$.

Let $X_{t,x}$ be the unique solution of the following d-dimensional stochastic differential equation

$$dX_{t,x} = V_0(X_{t,x})dt + \sum_{j=1}^{p} V_j(X_{t,x}) \circ dB_t^j, \quad X_{0,x} = x, \tag{2.1}$$

where $\mathbb{V} = \{V_0, V_1, ..., V_p\} \in C_b^\infty(\mathbb{R}^d; \mathbb{R}^d), x \in \mathbb{R}^d$. This equation is written in Stratonovich form and has an Itô equivalent form given by

$$dX_{t,x} = \tilde{V}_0(X_{t,x})dt + \sum_{j=1}^{p} V_j(X_{t,x})\, dB_t^j, \quad X_{0,x} = x,$$

where $V_0^i \triangleq \tilde{V}_0^i - \frac{1}{2}\sum_{j=1}^{p}\sum_{k=1}^{d} V_j^k \frac{\partial V_j^i}{\partial x_k}, i = 1, ..., d$. Given a multi-index $\alpha \in \mathcal{A}$, we define the Stratonovich iterated integrals as follows

$$J_\alpha(f)_{0,t} \triangleq \begin{cases} f(s) & \text{if} \quad |\alpha| = 1 \\ \int_0^t J_{\alpha-}(f)_{0,u}\, du & \text{if} \quad k \geq 1,\ \alpha_k = 0 \\ \int_0^t J_{\alpha-}(f)_{0,u} \circ dB^{\alpha_k} & \text{if} \quad k \geq 1,\ \alpha_k \neq 0 \end{cases}.$$

Given f, a sufficiently smooth function, and $X_{t,x}$, the solution of (2.1), we can expand $f(X_{t,x})$ in terms of iterated Stratonovich integrals. The precise statement is as follows.

Lemma 2.1 (Stratonovich-Taylor expansion) *Let $f \in C_b^\infty(\mathbb{R}^d; \mathbb{R}), m \in \mathbb{N}$. Then,*

$$f(X_{t,x}) = \sum_{\alpha \in \mathcal{A}(m)} V_\alpha f(x)\, J_\alpha(1)_{0,t} + R_m(t,x,f).$$

The remainder process $R_m(t,x,f)$ satisfies

$$\sup_{x\in\mathbb{R}^d} \sqrt{\mathbb{E}_{\mathbb{P}}[(R_m(t,x,f))^2]} \leq C \sum_{j=m+1}^{m+2} t^{j/2} \sup_{\alpha\in\mathcal{A}(j)\backslash\mathcal{A}(j-1)} \|V_\alpha f\|_\infty,$$

where $C = C(m)$ is a positive constant that only depends on m.

3 Cubature method on Wiener space

Cubature formulas are classical methods of numerical approximation of integrals over finite dimensional spaces with respect to positive measures. Let μ be a positive measure on $\mathbb{R}^d$ with finite moments up to order $m \in \mathbb{N}$. A cubature approximation μ^m for μ of degree m is a finite sequence of points $x_1, ..., x_n$ in the support of μ and positive weights $\lambda_1, ..., \lambda_n$ such that

$$\mu(p) \triangleq \int_{\mathbb{R}^d} p(x)\mu(dx) = \sum_{i=1}^{n} \lambda_i p(x_i) \triangleq \mu^m(p),$$

where p is any element of the space of polynomials in d variables and of degree less than or equal to m. If f is a regular enough function, $\mu^m(f)$ will be a good approximation of $\mu(f)$ as long as the approximation of f by polynomials is good. In order to make precise the previous statement, one relies on the Taylor expansion.

The cubature method on Wiener space is an infinite dimensional extension of cubature methods on $\mathbb{R}^d$. In this framework, the role of polynomials is played by iterated Stratonovich integrals and the role of Taylor expansions is played by Stratonovich-Taylor expansions.

3.1 One step cubature measure

Let $X_{t,x}$ be the unique solution of equation (2.1). We choose a version of $X_{t,x}$ that coincides on $C_{0,bv}([0,T], \mathbb{R}^p)$, the subspace of $C_0([0,T], \mathbb{R}^p)$ of functions of bounded variation, with the pathwise solution.

Definition 3.1 *A measure $\mathbb{Q}_1^m$ assigning positive weights $\lambda_1, ..., \lambda_{c_d}$ to paths $\omega_1, ..., \omega_{c_d} \in C_{0,bv}([0,1], \mathbb{R}^{p+1})$ is a cubature measure of degree $m \in \mathbb{N}$ for the Wiener measure, if for all $\alpha = (\alpha_1, ..., \alpha_k) \in \mathcal{A}(m)$,*

$$\mathbb{E}_{\mathbb{P}}[J_\alpha(\mathbf{1})_{0,1}] = \mathbb{E}_{\mathbb{Q}_1^m}[J_\alpha(\mathbf{1})_{0,1}] = \sum_{j=1}^{c_d} \lambda_j \int_{0<t_1<\cdots<t_k<1} \circ d\omega_j^{\alpha_1}(t_1) \cdots \circ d\omega_j^{\alpha_k}(t_k).$$

The constant $c_d = c_d(m,p)$ only depends on the degree and the dimension of the Brownian motion.

Lyons and Victoir [18] proved that one can always find a cubature measure supported on at most $card(\mathcal{A}(m))$ continuous paths of bounded variation. They also gave an explicit expression of degree-five cubature measure. In [11], the authors have constructed cubature formulas of higher degrees and for various dimensions of the driving Brownian motion.

Remark 3.2 *Assume $\mathbb{Q}_1^m = \sum_{j=1}^{c_d} \lambda_j \delta_{\omega_j}$ is a cubature measure on $[0,1]$. Then, for any $T > 0$, as $J_\alpha(\mathbf{1})_{0,T} \overset{\mathcal{L}}{=} T^{\|\alpha\|/2} J_\alpha(\mathbf{1})_{0,1}$, we have that $\mathbb{Q}_T^m \triangleq \sum_{j=1}^{c_d} \lambda_j \delta_{\langle T,\omega\rangle_j}$ is a cubature measure for the Wiener measure restricted to in $C_0([0,T], \mathbb{R}^p)$, where*

$$\langle T,\omega\rangle_j^i(s) \triangleq \begin{cases} T\omega_j^i(s/T) & if \quad i = 0 \\ T^{1/2}\omega_j^i(s/T) & if \quad i = 1, ..., p \end{cases}, s \in [0,T], j = 1, ..., c_d.$$

Definition 3.3 *We define the cubature approximation of $P_T f(x) \triangleq \mathbb{E}_{\mathbb{P}}[f(X_{T,x})]$ by $\bar{P}_T f(x) \triangleq \mathbb{E}_{\mathbb{Q}_T^m}[f(X_{T,x})]$.*

Remark 3.4 *Let $u(t,x)$ be the solution at time t of $\frac{\partial u}{\partial t}(t,x) = Lu(t,x), u(0,x) = f(x)$, where the operator L is defined by $Lf = V_0 f + \frac{1}{2}\sum_{i=1}^{p} V_i^2 f$. Then, $u(T,x) = \mathbb{E}_{\mathbb{P}}[f(X_{T,x})]$. Hence, $\bar{P}_T f(x)$ is an approximation of the semigroup $P_T f(x)$, which has infinitesimal generator L. In other words, the cubature on Wiener space can be used to produce a high order approximation method for solving second order parabolic differential equations.*

The main tool to bound the approximation error relies on the Stratonovich-Taylor expansion and is stated in the following lemma, which is easy to prove.

Lemma 3.5 *Let $f \in C_b^\infty\left(\mathbb{R}^d;\mathbb{R}\right), m \in \mathbb{N}$. Then,*

$$f\left(X_{T,x}\right) = \sum_{\alpha \in \mathcal{A}(m)} V_\alpha f\left(x\right) J_\alpha\left(1\right)_{0,T} + R_m\left(T,x,f\right).$$

The remainder process $R_m\left(T,x,f\right)$ satisfies

$$\sup_{x \in \mathbb{R}^d} \mathbb{E}_{\mathbb{Q}_T^m}\left[\left|R_m\left(T,x,f\right)\right|\right] \le C \sum_{j=m+1}^{m+2} T^{j/2} \sup_{\alpha \in \mathcal{A}(j)\setminus\mathcal{A}(j-1)} \left\|V_\alpha f\right\|_\infty,$$

where $C = C(p,m,\mathbb{Q}_1)$ is a positive constant that only depends on p, m and $\mathbb{Q}_1$.

Using the previous lemma, Lemma 2.1 and a triangular inequality argument and one obtains a bound for the error of the one step cubature approximation.

Proposition 3.6 *Let $\mathbb{Q}_T^m$ be a degree m cubature measure then*

$$\sup_{x \in \mathbb{R}^d} \left|P_T f\left(x\right) - \mathbb{E}_{\mathbb{Q}_T^m}\left[f\left(X_{T,x}\right)\right]\right| \le C \sum_{j=m+1}^{m+2} T^{j/2} \sup_{\alpha \in \mathcal{A}(j)\setminus\mathcal{A}(j-1)} \left\|V_\alpha f\right\|_\infty,$$

where $C = C(p,m,\mathbb{Q}_1)$ is a positive constant that only depends on p, m and $\mathbb{Q}_1$.

3.2 Iterated cubature measure

In general, the bound obtained in the previous proposition do not allow to directly get a good approximation of $P_T f\left(x\right)$ when T is large. To overcome this difficulty one iterates the cubature measure along a partition $\Pi^{n,T} \triangleq \{0 = t_0 < t_1 < \cdots < t_n = T\}$ of $[0,T]$. We will denote by $\Pi^{j,T} \triangleq \{0 = t_0 < t_1 < \cdots < t_j\}, j = 1,...n$, the subpartitions of $\Pi^{n,T}$ and $s_j \triangleq t_j - t_{j-1}, j = 1,...,n$.

Definition 3.7 *Let the measure $\mathbb{Q}_1^m = \sum_{j=1}^{c_d} \lambda_j \delta_{\omega_j}$ define a cubature formula on $[0,1]$ and $\Pi^{n,T}$ be a partition of $[0,T]$. The global cubature measure $\mathbb{Q}_{\Pi^{n,T}}^m$ is defined by*

$$\mathbb{Q}_{\Pi^{n,T}}^m = \sum_{(i_1,\ldots,i_n)} \lambda_{i_1} \cdots \lambda_{i_n} \delta_{\langle s_1,\omega\rangle_{i_1} \otimes \cdots \otimes \langle s_k,\omega\rangle_{i_n}},$$

where $\omega \circ \hat{\omega}$ denotes the concatenation of the paths ω and $\hat{\omega}$.

It is also useful to view the cubature formulas on Wiener space as Markov operators acting on discrete measures on $\mathbb{R}^d$. This interpretation justifies the following definition introduced by Litterer and Lyons [17].

Definition 3.8 *Given a positive measure $\mu = \sum_{i=1}^{l} \mu_i \delta_{x_i}$ on $\mathbb{R}^d$ and a cubature measure $\mathbb{Q}_1^m = \sum_{j=1}^{c_d} \lambda_j \delta_{\omega_j}$, we define the KLV_m operation with respect to μ over a time step s by*

$$KLV_m(\mu, s) \triangleq \sum_{i=1}^{l} \sum_{j=1}^{c_d} \mu_i \lambda_j \delta_{X_{s,x_i}(\langle s,\omega \rangle_j)},$$

where $X_{s,x_i}(\langle s,\omega\rangle_j)$ is the solution at time s of the following ODE

$$\begin{aligned} dX_{u,x_i}(\langle s,\omega\rangle_j) &= \sum_{k=0}^{p} V_k X_{u,x_i}(\langle s,\omega\rangle_j) d \langle s,\omega\rangle_j^k (u), \quad 0 \le u \le s \\ X_{0,x_i}(\langle s,\omega\rangle_j) &= x_i. \end{aligned}$$

One can also iterate the KLV_m operation along a partition $\Pi^{n,T}$.

Definition 3.9 *Let $\mathbb{Q}_1^m = \sum_{i=1}^{c_d} \lambda_i \delta_{\omega_i}$ be a cubature measure of degree m and let $\Pi^{n,T}$=$\{0 = t_0 < t_1 < \cdots < t_n = T\}$ be a partition of $[0,T]$. The KLV_m operation along $\Pi^{n,T}$, is defined recursively by*

$$KLV_m\left(\Pi^{j+1,T}, x\right) \triangleq KLV_m\left(KLV_m(\Pi^{j,T}, x), s_{j+1}\right), j = 1, ..., n-1,$$

and $KLV_m\left(\Pi^1, x\right) = KLV_m(\delta_x, s_1)$.

The following remark makes the connection between the two point of views.

Remark 3.10 *Let $\mathcal{B}$ denote the set of multi-indices $\{\varnothing\} \cup \bigcup_{k=1}^{\infty} \{1, ..., c_d\}^k$. For any $\beta = (\beta_1, ..., \beta_k) \in \mathcal{B}$ define $\lambda_\beta = \lambda_{\beta_1} \cdots \lambda_{\beta_k}$ and points $x_\beta \in \mathbb{R}^d$ by setting $x_\beta = X_{s_1,x}(\langle s_1,\omega\rangle_{\beta_1}), \beta \in \{(1), ..., (c_d)\}$, and $x_\beta = X_{s_k, x_{\beta-}}(\langle s_k,\omega\rangle_{\beta_k}), \quad \beta \in \mathcal{B}, |\beta| > 1$. Then, the global cubature measure along $\Pi^{n,T}$ can be written as the following discrete measure on paths $\mathbb{Q}^m_{\Pi^{n,T}} = \sum_{\beta \in \mathcal{B}, |\beta|=n} \lambda_\beta \delta_{\langle s_1,\omega\rangle_{\beta_1} \otimes \cdots \otimes \langle s_n,\omega\rangle_{\beta_n}}$, while the KLV_m operation along $\Pi^{n,T}$ can be written as the following discrete measure on $\mathbb{R}^d$, $KLV_m(\Pi^{n,T}, x) = \sum_{\beta \in \mathcal{B}, |\beta|=n} \lambda_\beta \delta_{x_\beta}$. Moreover, $\mathbb{E}_{\mathbb{Q}^m_{\Pi^{n,T}}}[f(X_{t,x})]$ $= KLV_m(\Pi^{n,T}, x)(f)$.*

Remark 3.11 *The iterative procedure to generate $\mathbb{Q}^m_{\Pi^{n,T}}$ can be viewed as an c_d-ary tree, which we will call the cubature tree. Hence, the support of the measure $\mathbb{Q}^m_{\Pi^{n,T}}$ (and of $KLV_m(\Pi^{n,T}, x)$) grows exponentially with the number of subintervals of the partition. In particular, we have to solve $\frac{c_d^{n+1}-1}{c_d-1}$ ODEs to obtain the points in the support of $KLV_m(\Pi^{n,T}, x)$. When n is large the computational cost associated to solving these ODEs can not be ignored and some mechanism to control the size of the support of $\mathbb{Q}^m_{\Pi^{n,T}}$ is needed. The basic approach is to allow the size of the tree to grow only up to a constant decided by the use and then to keep it constant by culling the branches with small weights. The procedure can be random. For example, Ninomiya [20] proposed to use the TBBA algorithm of Crisan and Lyons. Litterer and Lyons [17] have recently introduced a deterministic recombination procedure that essentially allows to change the original cubature measure with a measure with smaller support, without increasing the error.*

4 Cubature applied to filtering

In this section we introduce the setup for the filtering problem. We also present the approach by Crisan and Ghazali [5] to the application of cubature on Wiener space to filtering.

4.1 Stochastic filtering setup

Let $(\Omega, \mathcal{F}, P)$ be the probability space defined on Section 2, assumed to accommodate a k-dimensional Wiener process W independent of B. Let $\mathbb{F} = \{\mathcal{F}_t\}_{0\leq t\leq T}$ be a filtration satisfying the usual conditions of completeness and right continuity. In this probability space we consider a partially observed system $(X, Y) = \{(X_t, Y_t)\}_{0\leq t\leq T}$. The unobserved process $X = \{X_t\}_{0\leq t\leq T}$, called *the signal*, is the solution of the d-dimensional Stratonovich SDE (2.1) , that is,

$$dX_{t,x} = V_0(X_{t,x})dt + \sum_{j=1}^{p} V_j\left(X_{t,x}\right) \circ dB_t^j, X_{0,x} = x, 0 \leq t \leq T,$$

where $x \in \mathbb{R}^d$, $V_i \in C_b^\infty(\mathbb{R}^d, \mathbb{R}^d)$ and $B = (B^j)_{j=1}^p = \{(B_t^j)_{j=1}^p\}_{0\leq t\leq T}$ is a p-dimensional $\mathbb{F}$-Brownian motion. To simplify the notation, we will suppress the dependence of $X_{t,x}$ on x and write X_t. The observed component $Y = \{Y_t\}_{0\leq t\leq T}$, called *the observation process*, is given by the following k-dimensional process

$$Y_t = \int_0^t h\left(X_s\right) ds + dW_s, 0 \leq t \leq T,$$

where $h : \mathbb{R}^d \to \mathbb{R}^k$ is a bounded measurable function and $W = (W^j)_{j=1}^k = \{(W_t^j)_{j=1}^k\}_{0\leq t\leq T}$ is a k-dimensional $\mathbb{F}$-Brownian motion independent of B. Let $\{\mathcal{Y}_t\}_{0\leq t\leq T}$ be the usual augmentation of the filtration generated by the process Y, that is, $\mathcal{Y}_t = \sigma(\{Y_s\}_{s\leq t}) \vee \mathcal{N}, \quad 0 \leq t \leq T$ where $\mathcal{N}$ are the P-null sets of $(\Omega, \mathcal{F}, P)$. The stochastic filtering problem consists of determining the conditional distribution π_T of the signal X at time T given the information accumulated from observing Y in the time interval $[0, T]$; that is, for φ bounded Borel measurable, it consists of computing $\pi_T(\varphi) = \mathbb{E}_{\mathbb{P}}[\varphi(X_T)|\mathcal{Y}_T]$. The process

$$\tilde{Z}_t \triangleq \exp\left(-\sum_{i=1}^{k}\int_0^t h^i\left(X_s\right) dW_s^i - \frac{1}{2}\sum_{i=1}^{k}\int_0^t h^i(X_s)^2 ds\right), \quad 0 \leq t \leq T,$$

is an $\mathbb{F}$-martingale. For a fixed $0 \leq t \leq T$, we can define a new probability measure $\tilde{\mathbb{P}}^t$ on $\mathcal{F}_t$ via $\frac{d\tilde{\mathbb{P}}^t}{d\mathbb{P}}|_{\mathcal{F}_t} \triangleq \tilde{Z}_t$. By the martingale properties of $\tilde{Z}_t$, the family of probability measures $\{\tilde{\mathbb{P}}^t\}_{0\leq t\leq T}$ is consistent and this property allows us to define a new probability $\tilde{\mathbb{P}}$ which is equivalent to $\mathbb{P}$ on $\bigcup_{0\leq t<\infty} \mathcal{F}_t$. By means of Girsanov's theorem, Y becomes, under $\tilde{\mathbb{P}}$, a Brownian motion independent of the signal X. Note also that the law of X is invariant under this change of probability measure. In order to construct numerical algorithms to approximate π_T one relies, crucially, in the Kallianpur-Striebel formula, see [12],

$$\pi_T\left(\varphi\right) = \frac{\rho_T\left(\varphi\right)}{\rho_T\left(\mathbf{1}\right)},$$

where ρ_t, called the unnormalised conditional distribution, is given by

$$\rho_T(\varphi) \triangleq \mathbb{E}_{\tilde{\mathbb{P}}}\left[\varphi(X_T)\exp\left(\sum_{i=1}^{k}\int_0^T h^i(X_t)\,dY_t^i - \frac{1}{2}\sum_{i=1}^{k}\int_0^T h^i(X_t)^2 dt\right)\middle| \mathcal{Y}_T\right].$$

Thanks to the Kallianpur-Striebel formula the problem is reduced to find an approximation of the above functional.

Remark 4.1 *$\rho_T(\varphi)$ is the expected value of a functional of the signal X, which is parametrized by the observation process Y. This representation shows the fact that the signal X enters the problem only through the evolution of its law, whilst its path properties are not relevant. On the other hand, the observed path of Y determines the functional to be integrated and the distribution of Y only plays a secondary role. In practice, we will only know the values of Y along the points in a partition of $[0,T]$ and we may not know the law of X. Hence, we will need to approximate the Y-dependent functional of X as well as the law of X. Therefore, the filtering problem can be viewed as a particular case within the theory of weak approximations of SDEs.*

It follows from the previous remark that the design of an approximating scheme for $\rho_T(\varphi)$ should contain the following three components:

1. The discretization the Y-dependent functional of X.
2. The approximation of the law of the signal X.
3. The control of the computational effort.

An important ingredient to establish a weak approximation result is the space of test functions for which the result holds. As pointed out before, this space of test functions depends on Y. Moreover, its elements have to integrate not only with respect to the law of X but also with respect to the law of the approximating processes under consideration. The suitable candidate when using cubature formulas is the following (see Crisan and Ghazali [5]). In the following $\|\cdot\|_p$ denotes the $L(\tilde{\mathbb{P}})$ norm.

Definition 4.2 *Let $C_{\mathbb{R}^k}[0,T]$ be the space of continuous functions $y:[0,T]\to\mathbb{R}^k$ and $C_b^{Y,\infty}(\mathbb{R}^d)$ the set of measurable functions $f:\mathbb{R}^d\times C_{\mathbb{R}^k}[0,T]\to\mathbb{R}$ satisfying the following properties:*

1. *For any $y\in C_{\mathbb{R}^k}[0,T]$ the function $x\to f(x,y)$ belongs to $C_b^\infty(\mathbb{R}^d)$.*
2. *For any multi-index $\alpha\in\mathcal{D}\triangleq\cup_{k=1}^\infty\{1,...,d\}^k\cup\{\varnothing\}$, any $x\in\mathbb{R}^d$ and $p\geq 1$, the partial derivative $D_\alpha f(x,Y)$ in the first variable satisfies $\|D_\alpha f(x,Y)\|_p<\infty$.*
3. *For any multi-index $\alpha\in\mathcal{D}$ and $p\geq 1$, we have $|||D_\alpha f(x,Y)|||_{p,\infty}\triangleq\sup_{x\in\mathbb{R}^d}\|D_\alpha f(x,Y)\|_p<\infty$.*

Definition 4.3 *For any function $f\in C_b^{Y,\infty}(\mathbb{R}^d)$, $j\in\mathbb{N}, p\geq 1$ we define the norms $|||D_\alpha f(x,Y)|||_{p,j}=\sum_{\alpha\in\mathcal{D}(j)}|||D_\alpha f(x,Y)|||_{p,\infty}$, where $\mathcal{D}(j)\triangleq\{\alpha\in D:\|\alpha\|\leq j\}$.*

4.2 Picard's filter

In this section we introduce the discretization of the Y-dependent functional of X to be integrated. This discretization was first introduced by Picard [22] and we shall call it Picard's filter, see also [3] and [23]. Assume that we have an uniform partition $\Pi^{n,T} \triangleq \{t_i = \frac{iT}{n}\}_{i=0,...,n}$ of the interval $[0,T]$ and that we know $\{Y_{t_i}\}_{i=0,...,n}$, the values of the observation process Y on $\Pi^{n,T}$. For any $\varphi \in C_b^\infty$, we can define

$$\begin{array}{llll} \Theta^{n,\varphi}: & \left(\mathbb{R}^d\right)^{n+1} & \longrightarrow & \mathbb{R} \\ & (z_0,...,z_n) & \mapsto & \varphi(z_n)\exp\left(\sum_{r=0}^{n} h_r(z_r)\right) \end{array}, \tag{4.1}$$

where $h_r : \mathbb{R}^d \longrightarrow \mathbb{R}, r = 0,...,n$, are the following functions

$$h_r(z) \triangleq \sum_{i=1}^{k}\{h^i(z)\Delta Y_r^i - \frac{T}{2n}\left(h^i(z)\right)^2\},$$

$h_n(z) \triangleq 0$, and $\Delta Y_r^i \triangleq (Y_{t_{r+1}}^i - Y_{t_r}^i), r = 0,...,n-1$. Next, define $\rho_T^n(\varphi) \triangleq \mathbb{E}_{\tilde{\mathbb{P}}}[\Theta^{n,\varphi}(X_{t_0,x},...,X_{t_n,x})|\mathcal{Y}_T]$. Note that φ and $h_r, r = 0,...,n$ belong to $C_b^{Y,\infty}(\mathbb{R}^d)$ and as $\Theta^{n,\varphi}$ is a product of these functions it also belongs to $C_b^{Y,\infty}(\mathbb{R}^d)$. The following result was proved by Picard [22] :

Theorem 4.4 *Let φ be a bounded and Lipschitz continuous function.Then, there exists a constant $C = C(T, \|\varphi\|_\infty)$ independent of n such that*

$$\|\rho_T(\varphi) - \rho_T^n(\varphi)\|_2 \leq \frac{C}{n}.$$

See [4] for an updated account on the discretization of the continuous time filtering problem.

Remark 4.5 *The previous theorem shows that, for uniform partitions, ρ_T^n is a first-order approximation of ρ_T. As the algorithms we are going to develop will be based on the Picard discretization, the error of these algorithms when approximating ρ_T will not be better than C/n.*

4.3 The cubature approximation

The second step is to approximate the law of the signal X. In this paper we will use the cubature on Wiener space to do this. We define the cubature approximation to $\rho_T^n$of Picard's filter by

$$\bar{\rho}_T^n(\varphi) \triangleq \mathbb{E}_{\mathbb{Q}^m_{\Pi^{n,T}}}[\Theta^{n,\varphi}(X_{t_0,x},...,X_{t_n,x})|\mathcal{Y}_T].$$

In order to analyse the error when approximating $\rho_T^n(\varphi)$ by $\bar{\rho}_T^n(\varphi)$ it is convenient to introduce an alternative representations for Picard's filter and its approximation. We define operators $\{R_t^i\}_{i=1}^n$ and $\{\bar{R}_t^i\}_{i=1}^n$ for $\varphi \in C_b^\infty\left(\mathbb{R}^d\right)$, $x \in \mathbb{R}^d$ and $t \in (0,T]$ by

$$R_t^i\varphi(x) \triangleq \mathbb{E}_{\tilde{\mathbb{P}}}[\varphi(X_{t,x})\exp(h_i(X_{t,x}))|\mathcal{Y}_t], \quad \bar{R}_t^i\varphi(x) \triangleq \mathbb{E}_{\mathbb{Q}_t^m}[\varphi(X_{t,x})\exp(h_i(X_{t,x}))|\mathcal{Y}_t].$$

To simplify the notation we also define $R_t^{i,j}\varphi(x) \triangleq R_t^i \cdots R_t^j \varphi(x)$ and $\bar{R}_t^{i,j}\varphi(x) \triangleq \bar{R}_t^i \cdots \bar{R}_t^j \varphi(x)$ for $1 \leq i < j \leq n$. Then, we have that

$$\rho_T^n(\varphi) = \exp(h_0(x)) R^1_{\frac{T}{n}} \cdots R^n_{\frac{T}{n}} \varphi(x) = \exp(h_0(x)) R^{1,n}_{\frac{T}{n}} \varphi(x),$$

and

$$\bar{\rho}_T^n(\varphi) = \exp(h_0(x)) \bar{R}^1_{\frac{T}{n}} \cdots \bar{R}^n_{\frac{T}{n}} \varphi(x) = \exp(h_0(x)) \bar{R}^{1,n}_{\frac{T}{n}} \varphi(x).$$

The main result concerning the cubature approximation of Picard's filter is the following theorem proved by Crisan and Ghazali in [5]. The result basically says that $\bar{\rho}_T^n$ is an approximation of order $(m-1)/2$ of ρ_T^n, where m is the degree of the cubature measure.

Theorem 4.6 *There is a positive constant $C = C(T, m, p)$ such that for all $\varphi \in C_b^{m+2}(\mathbb{R}^d; \mathbb{R})$, $p \geq 1$, we have $\|\bar{\rho}_T^n(\varphi) - \rho_T^n(\varphi)\|_p \leq C n^{-(m-1)/2} \|\varphi\|_{\infty,m+2}$ where*

$$\|\varphi\|_{\infty,m+2} \triangleq \|\varphi\|_\infty + \sum_{k=1}^{m+2} \max_{j_1,\ldots,j_k \in \{1,\ldots,d\}} \left\| \frac{\partial^k \varphi}{\partial x_{j_1} \cdots \partial x_{j_k}} \right\|_\infty .$$

A sketch of the proof of Theorem 4.6 is as follows. From a variation of Lemmas 2.1 and 3.5 applied to functions $f \in C_b^{Y,\infty}(\mathbb{R}^d)$ one obtains that

$$\sup_{x \in \mathbb{R}^d} \left| \mathbb{E}_{\mathbb{Q}_t^m}[f(X_{t,x})] - \mathbb{E}_{\tilde{\mathbb{P}}}[f(X_{t,x})] \right| \leq C \sum_{i=m+1}^{m+2} t^{i/2} |||f|||_{p,i}.$$

This error bound is used to prove that

$$|||R^{j-1}_{T/n} R^{j,n}_{T/n} \varphi(x) - \bar{R}^{j-1}_{T/n} R^{j,n}_{T/n} \varphi(x)|||_{p,\infty} \leq C n^{-(m+1)/2} ||\varphi||_{\infty,m+2}.$$

The previous bound is combined with a telescopic expansion of

$$\exp(h_0(x)) R^{1,n}_{\frac{T}{n}} \varphi(x) - \exp(h_0(x)) \bar{R}^{1,n}_{\frac{T}{n}} \varphi(x)$$

to prove that

$$||| \exp(h_0(x)) R^{1,n}_{\frac{T}{n}} \varphi(x) - \exp(h_0(x)) \bar{R}^{1,n}_{\frac{T}{n}} \varphi(x) |||_{p,\infty} \leq C n^{-(m-1)/2} \|\varphi\|_{\infty,m+2}$$

from which the result follows easily.

Corollary 4.7 *There is a positive constant $C = C(T, m, \|\varphi\|_{\infty,m+2})$ such that for all $\varphi \in C_b^{m+2}(\mathbb{R}^d; \mathbb{R})$, we have $\mathbb{E}_{\tilde{\mathbb{P}}}[|\bar{\pi}_T^n(\varphi) - \pi_T(\varphi)|] \leq \frac{C}{n}$.*

Proof. Using the triangle inequality and the estimates in Theorems 4.4 and 4.6 we have that $\|\rho_T(\varphi) - \bar{\rho}_T^n(\varphi)\|_2 \leq \frac{C}{n}$. The result follows from using the Cauchy-Schwarz inequality to the following inequality

$$|\bar{\pi}_T^n(\varphi) - \pi_T(\varphi)| \leq \frac{\|\varphi\|_\infty}{\rho_T(\mathbf{1})} |\bar{\rho}_T^n(\mathbf{1}) - \rho_T(\mathbf{1})| + \frac{1}{\rho_T(\mathbf{1})} |\bar{\rho}_T^n(\varphi) - \rho_T(\varphi)|,$$

and the fact that $\left\|\rho_T(\mathbf{1})^{-1}\right\|_p$ is finite for any $p \geq 1$. ■

We will also need the following lemmas regarding the cubature approximation of Picard's filter.

Lemma 4.8 *Assuming the notation in Remark 3.10, we have that*

$$\bar{\rho}_t^n(\varphi) = \sum_{\beta \in \mathcal{B}, |\beta| = n} \lambda_\beta w_0(x_{\beta[0]}) \cdots w_{n-1}(x_{\beta[n-1]}) \varphi(x_\beta),$$

where $w_r(x_{\beta[r]}) \triangleq \exp\left(h_r\left(x_{\beta[r]}\right)\right), r = 0, ..., n-1$, *are called the filtering weights and by convention* $x_{\beta[0]} = x$.

Proof. Note that, $\bar{R}_t^n \varphi(x) = \sum_{\beta_n=1}^{c_d} \lambda_{\beta_n} \varphi(X_{t,x}(\langle t, \omega \rangle_{\beta_n}))$ and

$$\bar{R}_t^i \varphi(x) = \sum_{\beta_i=1}^{c_d} \lambda_{\beta_i} \varphi(X_{t,x}(\langle t, \omega \rangle_{\beta i})) w_i(h_i(X_{t,x}(\langle t, \omega \rangle_{\beta_i}))), \quad i = 0, ..., n-1.$$

Set $\delta = T/n$. We have that

$$\bar{R}_\delta^n \varphi(x) = \sum_{\beta_n=1}^{c_d} \lambda_{\beta_n} \varphi(X_{\delta,x}(\langle \delta, \omega \rangle_{\beta_n})) = \sum_{\beta_n=1}^{c_d} \lambda_{\beta_n} \varphi(x_{(\beta_n)}(x)),$$

where we have used the notation $x_{(\beta_n)}(x) \triangleq X_{\delta,x} \langle \delta, \omega \rangle_{\beta n}$. Applying $\bar{R}_\delta^{n-1}$ to $\bar{R}_\delta^n \varphi(x)$ we get

$$\begin{aligned}
&\bar{R}_\delta^{n-1}\left(\bar{R}_\delta^n \varphi(x)\right) = \bar{R}_\delta^{n-1}\left(\sum_{\beta_n=1}^{c_d} \lambda_{\beta_n} \varphi(x_{(\beta_n)}(x))\right) \\
&= \sum_{\beta_{n-1}, \beta_n=1}^{c_d} \lambda_{\beta_{n-1}} \lambda_{\beta_n} \varphi(X_{\delta, x_{(\beta_n)}(x)}(\langle \delta, \omega \rangle_{\beta_n})) w_{n-1}(h_{n-1}(X_{\delta,x}(\langle \delta, \omega \rangle_{\beta_{n-1}}))) \\
&= \sum_{\beta_{n-1}, \beta_n=1}^{c_d} \lambda_{\beta_{n-1}} \lambda_{\beta_n} \varphi(x_{(\beta_{n-1}, \beta_n)}(x)) w_{n-1}(h_{n-1}(x_{(\beta_{n-1})}(x))),
\end{aligned}$$

where we have used the notation

$$x_{(\beta_{n-1}, \beta_n)}(x) \triangleq X_{\delta, x_{(\beta_n)}(x)}(\langle \delta, \omega \rangle_{\beta_n}) = X_{2\delta, x}(\langle \delta, \omega \rangle_{\beta_{n-1}} \otimes \langle \delta, \omega \rangle_{\beta_n}).$$

Iterating this procedure it is clear that we get the result. ■

Remark 4.9 *From the previous lemma it follows that the computation of the cubature approximation of Picard's filter requires knowledge of all intermediate nodes in the cubature tree, contrasting to the typical use of cubature methods where the knowledge of the leafs is sufficient to compute the approximation. Obviously, this is due to the particular form of the functional to be integrated that depends explicitly on the values of* X_t *along the points of the partition and not just on the terminal value.*

Lemma 4.10 *For any* $p \geq 1$, *we have that* $\left\|\bar{\rho}_T^n(\mathbf{1})^{-1}\right\|_p < \infty$.

Proof. Lemma 4.8 and Jensen inequality yield that

$$\bar{\rho}_T^n(\mathbf{1})^{-p} \leq \sum_{\beta \in \mathcal{B}, |\beta| = n} \lambda_\beta \left(w_0(x_{\beta[0]}) \cdots w_{n-1}(x_{\beta[n-1]})\right)^{-p}.$$

By the definition of the exponential weights we can write

$$
\begin{aligned}
&\big(w_0(x_{\beta[0]})\cdots w_{n-1}(x_{\beta[n-1]})\big)^{-p}\\
&=\exp\left\{-p\sum_{r=0}^{n-1}h_r\left(x_{\beta[r]}\right)\right\}\\
&=\exp\left\{\sum_{r=0}^{n-1}\sum_{i=1}^{k}\{-ph^i\left(x_{\beta[r]}\right)\Delta Y_r^i+\frac{pT}{2n}\left(h^i\left(x_{\beta[r]}\right)\right)^2\}\right\}
\end{aligned}
$$

As Y is a k-dimensional standard Brownian motion under $\tilde{\mathbb{P}}$ we have that

$$
\begin{aligned}
\mathbb{E}[\big(w_0(x_{\beta[0]})\cdots w_{n-1}(x_{\beta[n-1]})\big)^{-p}]&=\exp\left\{\sum_{r=0}^{n-1}\sum_{i=1}^{k}\frac{(p^2+p)T}{2n}\left(h^i\left(x_{\beta[r]}\right)\right)^2\right\}\\
&\leq\exp\left\{\frac{(p^2+p)kT}{2}\left\|h\right\|_\infty^2\right\}.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\left\|\bar{\rho}_T^n(\mathbf{1})^{-1}\right\|_p^p=\mathbb{E}_{\tilde{\mathbb{P}}}[\bar{\rho}_T^n(\mathbf{1})^{-p}]&\leq\sum_{\beta\in\mathcal{B},|\beta|=n}\lambda_\beta\mathbb{E}_{\tilde{\mathbb{P}}}[\big(w_0(x_{\beta[0]})\cdots w_{n-1}(x_{\beta[n-1]})\big)^{-p}]\\
&\leq\exp\left\{\frac{(p^2+p)kT}{2}\left\|h\right\|_\infty^2\right\}\underbrace{\sum_{\beta\in\mathcal{B},|\beta|=n}\lambda_\beta}_{=1}<\infty
\end{aligned}
$$

■

5 The control of the computational effort

The tree based branching algorithm is a method that assigns a number of particles to different sites, according to a probability distribution with finite support on the sites. The computational effort is controlled as it is proportional to the number of particles. This is equivalent to generate rational valued random distributions which are unbiased estimators of the original probability distribution. The interesting feature of the method is that the assignment is done satisfying a certain minimum variance property. The results presented here can be extended to probability distributions with infinite support.

Let $\mathcal{X}=\{x_i\}_{i=1}^k$ be a given set and $\Gamma=\{\gamma_i\}_{i=1}^k$ a probability distribution with support $\mathcal{X}$. The problem is to generate a family of random variables $\hat{\Gamma}=\{\hat{\gamma}_i\}_{i=1}^k$, defined on some probability space $(\Omega^*,\mathcal{F}^*,\mathbb{P}^*)$, with values in $\{0,...,N\}$ and such that

$$\mathbb{E}[\hat{\gamma}_i]=N\gamma_i,\quad i=1,...,k, \tag{5.1}$$

$$\sum_{i=1}^k\gamma_i=N, \tag{5.2}$$

$$\mathrm{Var}[\hat{\gamma}_i]=\min_{\delta\in\mathcal{P}(\gamma_i)}\mathrm{Var}[\delta],\quad i=1,...,k, \tag{5.3}$$

where $\mathcal{P}(\gamma_i), i = 1, ..., k$, denote the set of all random variables with values in $\{0, ..., N\}$ and satisfying (5.1) . Let $[x]$ denote the integer part of $x \in \mathbb{R}$ and $\{x\} = x - [x]$ denote the fractional part of $x \in \mathbb{R}$. It is immediate that any family $\hat{\Gamma}$ of random variable with marginal distributions given by

$$\gamma_i \triangleq \begin{cases} [N\gamma_i] & \text{with probability} \quad 1 - \{N\gamma_i\} \\ [N\gamma_i] + 1 & \text{with probability} \quad \{N\gamma_i\} \end{cases} , \quad i = 1, ..., k$$

satisfies the minimal variance property (5.3) and the unbiasedness condition (5.1). It can be helpful to use a "particle" picture to describe the random variables in the set $\hat{\Gamma}$. Essentially, one can think that Γ is the empirical measure associated to a set of N particles that are allocated to the sites $\mathcal{X}$. Hence, $\hat{\gamma}_i$ represents the number of particles allocated to site x_i. This number is random and its mean is given by $N\gamma_i$ (which is not necessarily integer) However, its generation is not straightforward as condition (5.2) makes the random variables corresponding to different sites x_i to be correlated. The TBBA precisely allows to construct a family of random variables satisfying (5.1),(5.3) and the additional condition (5.2). The name of the algorithm comes from the fact that it can be described using a binary tree structure. The description is as follows.

1. We start with a k-ary tree. This tree has a root node initially storing N particles and k leaves that represent the sites where the particles have to be allocated.

2. We embed the k-ary tree into a binary tree satisfying the following rules.

   (a) The set of all leaves of the tree is $\mathcal{X}$.

   (b) Each node z of the tree has a positive weight γ_z.

   (c) If two different nodes share the same parent their weights add up to the weight of the parent.

   (d) The weights of all leaves which are descendants of a particular node add up to the weight of that node.

3. We move the N particles down along the tree until they get to the leaves using the following TBBA rules:

   (a) We start by allocating all N particles to the root node (the corresponding weight of the root is $\sum_{i=1}^{k} \gamma_i = 1$).

   (b) We then proceed recursively as follows: let z be a node with $\hat{\gamma}_z$ particles and weight γ_z. If z has two child nodes z_1 and z_2, then $\gamma_z = \gamma_{z_1} + \gamma_{z_2}$ and we will split the $\hat{\gamma}_z$ particles associated to z into $\hat{\gamma}_{z_1}$ particles associated to z_1 and $\hat{\gamma}_{z_2}$ particles associated to z_2, i.e., $\hat{\gamma}_z = \hat{\gamma}_{z_1} + \hat{\gamma}_{z_2}$, according to the following two possible cases.

      - Case 1: $[N\gamma_z] = [N\gamma_{z_1}] + [N\gamma_{z_2}]$
        - $\hat{\gamma}_{z_1} \triangleq [N\gamma_{z_1}] + (\hat{\gamma}_z - [N\gamma_z])u_m$,
        - $\hat{\gamma}_{z_2} \triangleq [N\gamma_{z_2}] + (\hat{\gamma}_z - [N\gamma_z])(1 - u_m)$, where

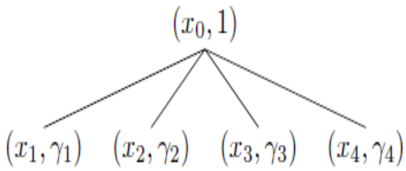
$$u_m \triangleq \begin{cases} 0 & \text{with prob} \quad \{N\gamma_{z_2}\}/\{N\gamma_z\} \\ 1 & \text{with prob} \quad \{N\gamma_{z1}\}/\{N\gamma_z\} \end{cases} .$$

- Case 2: $[N\gamma_z] = [N\gamma_{z_1}] + [N\gamma_{z_2}] + 1$
  - $\hat{\gamma}_{z_1} \triangleq [N\gamma_{z_1}] + 1 + (\hat{\gamma}_z - ([N\gamma_z] + 1))u_m$,
  - $\hat{\gamma}_{z_2} \triangleq [N\gamma_{z_2}] + 1 + (\hat{\gamma}_z - ([N\gamma_z] + 1))(1 - u_m)$, where

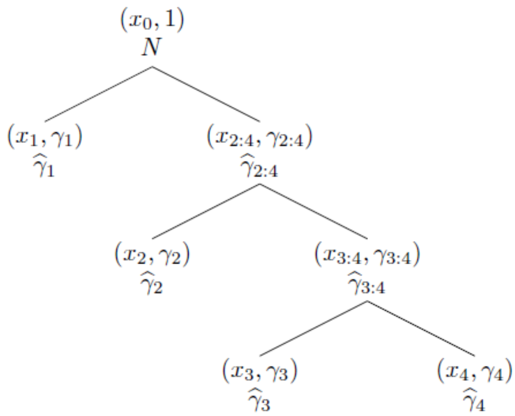
$$u_m \triangleq \begin{cases} 0 & \text{with prob} \quad (1 - \{N\gamma_{z_2}\})/(1 - \{N\gamma_z\}) \\ 1 & \text{with prob} \quad (1 - \{N\gamma_{z1}\})/(1 - \{N\gamma_z\}) \end{cases} .$$

Note that for each intermediate node in the tree we need to generate a random variable u_m. These random variables are independent of each other. The best way to understand how the algorithm works is to see some examples.

Example 5.1 *Assume that we have* $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$ *and* $\Gamma = \{\gamma_1, \gamma_2, \gamma_3, \gamma_4\}$. *We start with the following 4-ary tree.*

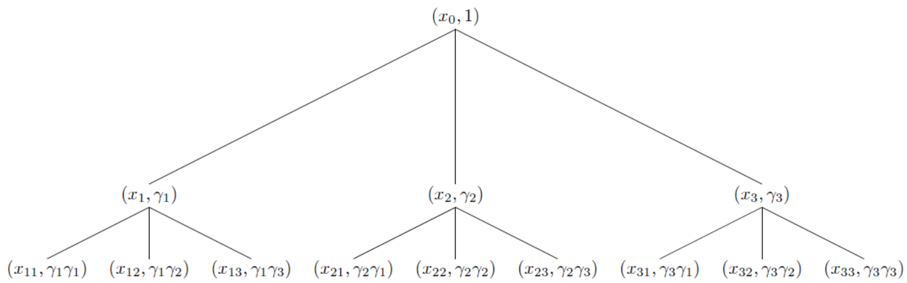


In order to construct the embedded binary tree we start by adding N particles to the root node. Then we assign the site x_1 *and the probability* γ_1 *to the left child node of the root. On the right child node we assign the auxiliary site* $x_{2:4} \triangleq \{x_i\}_{i=2}^4$ *with weight* $\gamma_{2:4} \triangleq \sum_{i=2}^4 \gamma_i$. *Now we apply the TBBA rules and get* $\hat{\gamma}_1$ *particles for the site* x_1 *and* $\hat{\gamma}_{2:4}$ *particles for the site* $x_{2:4}$. *Next we take the site* $x_{2:4}$ *as it were the root node and repeat the procedure. That is, on the left child node of the node* $x_{2:4}$ *we assign the site* x_2 *with probability* γ_2 *and on the right child node we assign the auxiliary node* $x_{3:4} \triangleq \{x_i\}_{i=3}^4$ *with weight* $\gamma_{3:4} \triangleq \sum_{i=3}^4 \gamma_i$. *We apply the TBBA rules to the nodes* $x_{2:4}, x_3$ *and* $x_{3:4}$ *and obtain* $\hat{\gamma}_2$ *particles for* x_2 *and* $\hat{\gamma}_{3:4}$ *particles for* $x_{3:4}$. *Iterating this procedure until the set of leaves coincides with* $\mathcal{X}$ *(in this case one more time) we end up with a set of random variables* $\hat{\Gamma} = \{\hat{\gamma}_i\}_{i=1}^4$ *satisfying the desired properties. The embedded binary tree is the following:*

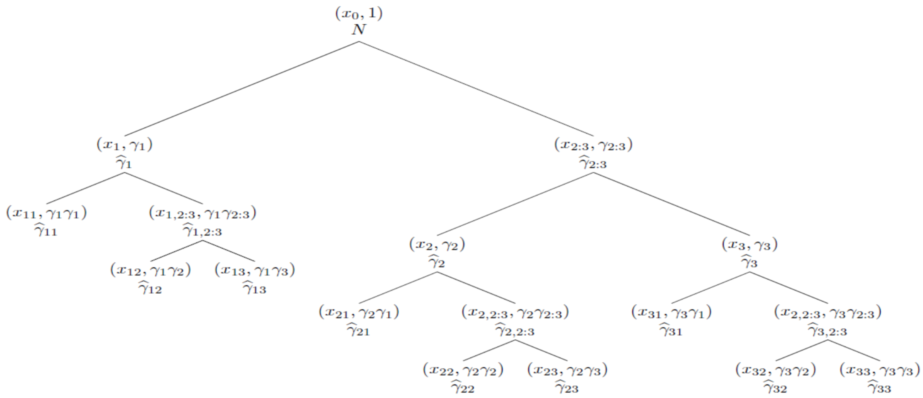


Note, that the way to embed the 4-ary tree into a binary one is by no means unique, as we well may have chosen another way of grouping the sites. This degree of freedom can be exploited in practice.

Example 5.2 *Assume that we have the following ternary tree of depth 2*



where the $\Gamma^1 \triangleq \{\gamma_i\}_{i=1}^3$ *is a probability distribution on* $\mathcal{X}^1 \triangleq \{x_i\}_{i=1}^3$ *and, obviously,* $\Gamma^2 \triangleq \{\gamma_i\gamma_j\}_{i,j=1}^3$ *is a probability distribution on* $\mathcal{X}^2 \triangleq \{x_{ij}\}_{i,j=1}^3$*. If we were just interested in sampling from* Γ^2*, we could repeat the procedure of the previous example with* $\Gamma = \Gamma^2$ *and* $\mathcal{X} = \mathcal{X}^2$*. However, we usually also need to sample from* Γ^1*. Moreover, it is more efficient to first apply the* TBBA *algorithm to* Γ^1 *and then apply the* TBBA *algorithm again to each of the sites in* $\mathcal{X}^1$ *(taking into account that now the weight of root node is not* 1*). This method is more efficient because for the sites in* $\mathcal{X}^1$ *that are assigned zero particles we do not need to apply the* TBBA *again, we just set zero particles to its descendants. The generated tree is as follows.*



Assume we have an n-times iterated k-ary tree such that at the first level of the tree we have a probability distribution $\Gamma^1 = \{\gamma_i\}_{i=1}^k$. Moreover, assume that the probability distributions in the next levels are generated by iterating the distribution in the first level, that is $\Gamma^l = \{\lambda_{i_1} \cdots \lambda_{i_l}\}_{i_1,\dots i_l=1}^k$. The previous example shows, that the TBBA will provide an approximation of the probability distribution not just at the final level, but also at all intermediate levels. Let z be a node in the iterated k-ary tree with $\hat{\gamma}_z$ particles assigned and γ_z weight. The algorithm that allocates the $\hat{\gamma}_z$ particles in z to its k direct descendants according to the probability law $\{\gamma_i\}_{i=1}^k$ is as follows:

Algorithm TBBA$(N, \hat{\gamma}_z, \gamma_z, \{\gamma_i\}_{i=1}^k)$

$\kappa_1 := N\gamma_z, \quad \kappa_2 := \hat{\gamma}_z$
for $i = 1$ **to** $k-1$
  draw $u_i \sim Unif[0,1]$
  if $\{N\gamma_z\gamma_i\} + \{\kappa_1 - N\gamma_z\gamma_i\} < 1$ **then**
    if $u_i < 1 - (\{N\gamma_z\gamma_i\}/\{\kappa_1\})$ **then**
      $\hat{\gamma}_i := [N\gamma_z\gamma_i]$
    else
      $\hat{\gamma}_i := [N\gamma_z\gamma_i] + (\kappa_2 - [\kappa_1])$
    end if
  else
    if $u_i < 1 - (1 - \{N\gamma_z\gamma_i\})/(1 - \{\kappa_1\})$ **then**
      $\hat{\gamma}_i := [N\gamma_z\gamma_i] + 1$
    else
      $\hat{\gamma}_i := [N\gamma_z\gamma_i] + (\kappa_2 - [\kappa_1])$
    end if
  end if
  $\kappa_1 := \kappa_1 - N\gamma_z\gamma_i$
  $\kappa_2 := \kappa_2 - \hat{\gamma}_i$
end for
$\hat{\gamma}_k := \kappa_2$
return $\{\hat{\gamma}_i\}_{i=1}^k$

Using this notation, the approximation to the probability measure Γ with support $\mathcal{X}$ is given by TBBA$(N, N, 1, \{\gamma_i\}_{i=1}^k)$. The algorithm generates a (random) measure with a support that is an at most N sites of the original k as it is the empirical distribution of N particles. Some of the properties satisfied by the random variables $\{\hat{\gamma}_i\}_{i=1}^k$ generated by TBBA$(N, N, 1, \{\gamma_i\}_{i=1}^k)$ are stated in the following proposition:

Proposition 5.3 *The random variables* $\{\hat{\gamma}_i\}_{i=1}^k = \text{TBBA}(N, N, 1, \{\gamma_i\}_{i=1}^k)$ *have the following properties.*

1. $\sum_{i=1}^k \hat{\gamma}_i = N$.
2. *For any* $i = 1, ..., k$, *we have* $\mathbb{E}_{P^*}[\hat{\gamma}_i] = N\gamma_i$.
3. *For any* $i = 1, ..., k$,γ_i *has minimal variance, specifically* $\mathbb{E}_{\mathbb{P}^*}[(\hat{\gamma}_i - N\gamma_i)^2] = \{N\gamma_i\}(1 - \{N\gamma_i\})$.
4. *For any* $1 \leq i < j \leq k$, *the random variables* γ_i *and* γ_j *are negatively correlated. That is,* $\mathbb{E}_{\mathbb{P}^*}[(\hat{\gamma}_i - N\gamma_i)(\hat{\gamma}_j - N\gamma_j)] \leq 0$.

Proof. See for example Proposition 9.3. in [1]. ∎

Note that, for any bounded function $\varphi : \mathcal{X} \to \mathbb{R}$, we have

$$\mathbb{E}_{\mathbb{P}^*}[\left(\sum_{i=1}^k \varphi(x_i)\frac{\hat{\gamma}_i}{N} - \sum_{i=1}^k \varphi(x_i)\gamma_i\right)^2]$$

$$\leq \frac{\|\varphi\|_\infty^2}{N^2}\mathbb{E}_{\mathbb{P}^*}[\left(\sum_{i=1}^{k}(\hat{\gamma}_i - N\gamma_i)\right)^2] \underset{4.}{\leq} \frac{\|\varphi\|_\infty^2}{N^2}\sum_{i=1}^{k}\mathbb{E}_{\mathbb{P}^*}[(\hat{\gamma}_i - N\gamma_i)^2]$$

$$\underset{3.}{=} \frac{\|\varphi\|_\infty^2}{N^2}\sum_{i=1}^{k}\{N\gamma_i\}(1-\{N\gamma_i\} \leq \frac{\|\varphi\|_\infty^2}{N^2}\sum_{i=1}^{k}N\gamma_i = \frac{\|\varphi\|_\infty^2}{N},$$

where we have used properties 3. and 4. in Proposition 5.3. This entails that $\hat{\gamma}_i/N$ is an unbiased approximation, in $L^2(\mathbb{P}^*)$, to γ_i.

6 Convergence results for the KLV particle filter

In this section we present the main results of the paper. Let $\mathbb{Q}_1^m = \sum_{i=1}^{c_d}\lambda_i\delta_{\omega_i}$ be a cubature measure of degree $m \in \mathbb{N}$. Assume that $\Pi^{n,T}$ is the uniform partition of the interval $[0,T]$ with $n+1$ points, i.e., $\Pi^{n,T} = \{t_j = \frac{jT}{n}\}_{j=0}^n$ and that we know the values of the observation process Y in $\Pi^{n,T}$. Let $\mathcal{B}$ be the total set of multi-indices with values in $\{1,...,c_d\}$, where c_d is the number of paths in the support of $\mathbb{Q}_1^m$. For $j = 1,...n$, define $\mathcal{B}^j = \{\beta \in \mathcal{B} : |b| = j\}$. Let $\mathcal{X}^j = \{x_\beta : \beta \in \mathcal{B}^j\}, j = 1,..,n$ be the support of $KLV_m(\Pi^{j,T}, x)$, the discrete measure obtained by the j-iteration of the KLV operation or cubature measure along the partition $\Pi^{n,T}$. According to Remark 3.10, we can see the global cubature measure along the partition $\Pi^{n,T}$ as discrete measure on $\mathbb{R}^d$ with points indexed by $\mathcal{B}^n$. By Lemma 4.8 we have the following expression for $\bar{\rho}_T^n$, the cubature approximation of Picard's filter,

$$\bar{\rho}_T^n = \sum_{\beta\in\mathcal{B}^n}\lambda_\beta w_0(x_{\beta[0]})\cdots w_{n-1}(x_{\beta[n-1]})\delta_{x_\beta}. \tag{6.1}$$

In the following two sections we are going to introduce two different approximation procedures for (6.1). Both procedures will be based on the TBBA. The first approximation, denoted by $\tilde{\rho}_T^{n,N}$will only use the cubature weights to allocate the particles along the cubature tree while the second one will combine both the cubature and the filtering weights. Hence, the second approximation denoted by $\hat{\rho}_T^{n,N}$ incorporates a correction mechanism similar to the one in the classical particle filters where law of the signal is approximated using the Euler scheme. We will assume that the probability space $(\Omega, \mathcal{F}, \tilde{\mathbb{P}})$ is rich enough to carry the auxiliary random variables needed to apply the TBBA. As Y is also defined on $(\Omega, \mathcal{F}, \tilde{\mathbb{P}})$, in what follows $\mathbb{E}_{\tilde{\mathbb{P}}}[\cdot|\mathcal{Y}_T]$ will denote the expectation with respect to the TBBA random variables.

6.1 Computational control of the KLV particle filter based on the cubature weights only

We define a collection $\{\tilde{\Gamma}_j\}_{j=1}^n$ of random variables according to the following recursion. Let $\Gamma^1 \triangleq \{\gamma_\beta\}_{\beta\in\mathcal{B}^1} = \{\lambda_i\}_{i=1}^{c_d}$ be the cubature weights. Define $\tilde{\Gamma}_1 \triangleq \{\tilde{\gamma}_\beta\}_{\beta\in\mathcal{B}^1} = \text{TBBA}(N, N, 1, \Gamma^1)$. For any $j \in \{2,...,n\}$, define $\tilde{\Gamma}_j \triangleq \bigcup_{\beta\in\tilde{\mathcal{B}}^j}\text{TBBA}(N, \tilde{\gamma}_{\beta-}, \lambda_{\beta-}, \Gamma^1)$, where, for $\beta \in \mathcal{B}^j, \lambda_{\beta-} \triangleq \lambda_{\beta_1}\cdots\lambda_{\beta_{j-1}}$. Note that $\text{TBBA}(N, 0, \lambda_{\beta-}, \Gamma^1)$ returns all the random variables equal to zero. Moreover, $\tilde{\Gamma}_j$ has the same distribution of $\text{TBBA}(N, N, 1, \Gamma^j)$, where $\Gamma^j \triangleq \{\lambda_\beta\}_{\beta\in\mathcal{B}^j}$.

In particular, if $\tilde{\gamma}_\beta, \tilde{\gamma}_\alpha \in \tilde{\Gamma}_n$, $\alpha \neq \beta$, then $\tilde{\gamma}_\beta$ and $\tilde{\gamma}_\alpha$ satisfy

$$\mathbb{E}_{\tilde{\mathbb{P}}}[\tilde{\gamma}_\beta|\mathcal{Y}_T] = N\lambda_\beta,$$
$$\mathbb{E}_{\tilde{\mathbb{P}}}[(\tilde{\gamma}_\beta - N\lambda_\beta)^2 |\mathcal{Y}_T] = \{N\lambda_\beta\}(1 - \{N\lambda_\beta\}) \leq N\lambda_\beta,$$
$$\mathbb{E}_{\tilde{\mathbb{P}}}[(\tilde{\gamma}_\beta - N\lambda_\beta)(\tilde{\gamma}_\alpha - N\lambda_\alpha)|\mathcal{Y}_T] \leq 0$$

and $\sum_{\beta\in\mathcal{B}^n} \tilde{\gamma}_\beta = N$. We define $\tilde{\rho}_T^{n,N} \triangleq \frac{1}{N} \sum\limits_{\beta\in\tilde{\mathcal{B}}^n} \tilde{\gamma}_\beta w_0(x)\cdots w_{n-1}(x_{\beta[n-1]})\delta_{x_\beta}$.

Theorem 6.1 *There exists a positive random variable $\tilde{C} = \tilde{C}(m,T,\varphi,Y)$ such that*

$$\mathbb{E}_{\tilde{\mathbb{P}}}[(\bar{\rho}_T^n(\varphi) - \tilde{\rho}_T^{n,N}(\varphi))^2|\mathcal{Y}_T] \leq \frac{\tilde{C}}{N},$$

for all $\varphi \in C_b^{m+2}(\mathbb{R}^d)$. Moreover, $\tilde{C}(m,T,\varphi,Y)$ is integrable with respect to $\tilde{\mathbb{P}}$.

Proof. We have that

$$\begin{aligned}
&\mathbb{E}_{\tilde{\mathbb{P}}}[(\bar{\rho}_T^n(\varphi) - \tilde{\rho}_T^{n,N}(\varphi))^2|\mathcal{Y}_T] \\
&= \frac{1}{N^2}\mathbb{E}_{\tilde{\mathbb{P}}}[\left(\sum_{\beta\in\mathcal{B}^n}(\tilde{\gamma}_\beta - N\lambda_\beta)w_0(x)\cdots w_{n-1}(x_{\beta[n-1]})\varphi(x_\beta)\right)^2|\mathcal{Y}_T] \\
&\leq \frac{1}{N^2}\sum_{\beta\in\mathcal{B}^n}\mathbb{E}_{\tilde{\mathbb{P}}}[(\tilde{\gamma}_\beta - N\lambda_\beta)^2|\mathcal{Y}_T]w_0^2(x)\cdots w_{n-1}^2(x_{\beta[n-1]})\varphi^2(x_\beta) \\
&= \frac{1}{N^2}\sum_{\beta\in\mathcal{B}^n}\{N\lambda_\beta\}(1-\{N\lambda_\beta\})w_0^2(x)\cdots w_{n-1}^2(x_{\beta[n-1]})\varphi^2(x_\beta) \\
&\leq \frac{\|\varphi\|_\infty^2}{N}\sum_{\beta\in\mathcal{B}^n}\lambda_\beta w_0^2(x)\cdots w_{n-1}^2(x_{\beta[n-1]}) \\
&= \frac{\|\varphi\|_\infty^2}{N}\mathbb{E}_{\mathbb{Q}^m_{\Pi^n,T}}[\left(\Theta^{n,\mathbf{1}}(X_{t_0,x},...,X_{t_n,x})\right)^2|\mathcal{Y}_T]
\end{aligned}$$

where we have used the TBBA properties of $\tilde{\Gamma}^n$. Then,

$$\tilde{C}(m,T,\varphi,Y) = \|\varphi\|_\infty^2\sum_{\beta\in\mathcal{B}^n}\lambda_\beta w_0^2(x)\cdots w_{n-1}^2(x_{\beta[n-1]}).$$

The integrability of $\tilde{C}(m,T,\varphi,Y)$ is deduced using similar arguments as in Lemma 4.10. ■

Definition 6.2 *We define $\tilde{\pi}_T^{n,N}$ by*

$$\tilde{\pi}_T^{n,N} \triangleq \frac{\tilde{\rho}_T^{n,N}}{\tilde{\rho}_T^{n,N}(\mathbf{1})} = \frac{\sum\limits_{\beta\in\tilde{\mathcal{B}}^n}\tilde{\gamma}_\beta w_0(x)\cdots w_{n-1}(x_{\beta[n-1]})\delta_{x_\beta}}{\sum\limits_{\alpha\in\tilde{\mathcal{B}}^n}\tilde{\gamma}_\alpha w_0(x)\cdots w_{n-1}(x_{\alpha[n-1]})}. \tag{6.2}$$

Corollary 6.3 *There exists a positive random variable $\tilde{K} = \tilde{K}(m,T,\varphi,Y)$ such that*

$$\mathbb{E}_{\tilde{\mathbb{P}}}[|\bar{\pi}_T^n(\varphi) - \tilde{\pi}_T^{n,N}(\varphi)||\mathcal{Y}_T] \leq \frac{\tilde{K}}{N^{1/2}},$$

for all $\varphi \in C_b^{m+2}(\mathbb{R}^d)$. Moreover, $\tilde{K}(m,T,\varphi,Y)$ is integrable with respect to $\tilde{\mathbb{P}}$.

Proof. We have that

$$
\begin{aligned}
\left|\bar{\pi}_T^n(\varphi)-\tilde{\pi}_T^{n,N}(\varphi)\right| &\le \left|\frac{\tilde{\pi}_T^{n,N}(\varphi)}{\bar{\rho}_T^n(\mathbf{1})}(\bar{\rho}_T^n(\mathbf{1})-\tilde{\rho}_T^{n,N}(\mathbf{1}))\right|+\left|\frac{1}{\bar{\rho}_T^n(\mathbf{1})}(\tilde{\rho}_T^{n,N}(\varphi)-\bar{\rho}_T^n(\varphi))\right| \\
&\le \frac{\|\varphi\|_\infty}{\bar{\rho}_T^n(\mathbf{1})}\left|\bar{\rho}_T^n(\mathbf{1})-\tilde{\rho}_T^n(\mathbf{1})\right|+\frac{1}{\bar{\rho}_T^n(\mathbf{1})}\left|\bar{\rho}_T^n(\varphi)-\tilde{\rho}_T^n(\varphi)\right|.
\end{aligned}
$$

By Theorem 6.1, we get

$$
\begin{aligned}
&\mathbb{E}_{\tilde{\mathbb{P}}}[|\bar{\pi}_T^n(\varphi)-\tilde{\pi}_T^{n,N}(\varphi)||\mathcal{Y}_T] \\
&\le \frac{1}{N^{1/2}}\frac{1}{\bar{\rho}_T^n(\mathbf{1})}\left(\|\varphi\|_\infty \tilde{C}(m,T,\mathbf{1},Y)^{1/2}+\tilde{C}(m,T,\varphi,Y)^{1/2}\right).
\end{aligned}
$$

Hence,

$$
\tilde{K}(m,T,\varphi,Y)=\frac{1}{\bar{\rho}_T^n(\mathbf{1})}\left(\|\varphi\|_\infty \tilde{C}(m,T,\mathbf{1},Y)^{1/2}+\tilde{C}(m,T,\varphi,Y)^{1/2}\right).
$$

Finally, that $\mathbb{E}_{\tilde{\mathbb{P}}}[|\tilde{K}(m,T,\varphi,Y)|]<\infty$ follows, using Cauchy-Schwarz inequality, from the integrability of $\tilde{C}(m,T,\varphi,Y)$ and Lemma 4.10. ■

6.2 Computational control of the KLV particle filter based on the cubature and filtering weights

We define a collection $\{\hat{\Gamma}_j\}_{j=1}^n$ of random variables according to the following recursion. Let $\bar{\mathcal{B}}^1=\mathcal{B}^1$ and $\{\gamma_\beta\}_{\beta\in\bar{\mathcal{B}}^1}$ be defined by

$$
\{\gamma_\beta\}_{\beta\in\bar{\mathcal{B}}^1}\triangleq\{\frac{\lambda_\beta w_0(x_{\beta[0]})}{\sum\limits_{\alpha\in B^1}\lambda_\alpha w_0(x_{\alpha[0]})}\}_{\beta\in\bar{\mathcal{B}}^1}=\{\frac{\lambda_\beta w_0(x)}{\sum\limits_{\alpha\in B^1}\lambda_\alpha w_0(x)}\}_{\beta\in\bar{\mathcal{B}}^1}=\{\lambda_\beta\}_{\beta\in\bar{\mathcal{B}}^1}.
$$

Next, define $\hat{\Gamma}_1\triangleq \mathrm{TBBA}(N,N,1,\{\gamma_\beta\}_{\beta\in\bar{\mathcal{B}}^1}),\hat{\mathcal{B}}^1\triangleq\{\beta\in\bar{\mathcal{B}}^1:\hat{\gamma}_\beta>0\}$. For any $j\in\{2,...,n\}$, let $\bar{\mathcal{B}}^j\triangleq\{\beta\in\mathcal{B}^j:\beta-\in\hat{\mathcal{B}}^{j-1}\}$ and

$$
\{\gamma_\beta\}_{\beta\in\bar{\mathcal{B}}^j}\triangleq\{\frac{\hat{\gamma}_{\beta-}\lambda_{\beta_j}w_{j-1}(x_{\beta[j-1]})}{\sum\limits_{\alpha\in\bar{\mathcal{B}}^j}\hat{\gamma}_{\alpha-}\lambda_{\alpha_j}w_{j-1}(x_{\alpha[j-1]})}\}_{\beta\in\bar{\mathcal{B}}^j}.
$$

Finally, define $\hat{\Gamma}_j\triangleq\mathrm{TBBA}(N,N,1,\{\gamma_\beta\}_{\beta\in\bar{\mathcal{B}}^j}),\hat{\mathcal{B}}^j\triangleq\{\beta\in\bar{\mathcal{B}}^j:\hat{\gamma}_\beta>0\}$.

Note that, by construction, $\sum_{\beta\in\hat{\mathcal{B}}^j}\hat{\gamma}_\beta=N$ and the TBBA weights are recursively defined and random. Set $\mathcal{G}_j\triangleq\sigma(\{\hat{\Gamma}_r\}_{r=1}^j),j=1,...,n$. Conditionally to $\mathcal{Y}_T\vee\mathcal{G}_{j-1}$, the family of random variables $\hat{\Gamma}_j$ satisfies the minimal variance properties of the TBBA generated random variables. In particular $\hat{\gamma}_\beta\in\hat{\Gamma}_j$ satisfies $\mathbb{E}_{\tilde{\mathbb{P}}}[\hat{\gamma}_\beta|\mathcal{Y}_T\vee\mathcal{G}_{j-1}]=N\gamma_\beta$, and

$$
\mathbb{E}_{\tilde{\mathbb{P}}}[(\hat{\gamma}_\beta-N\gamma_\beta)^2|\mathcal{Y}_T\vee\mathcal{G}_{j-1}]=\{N\gamma_\beta\}(1-\{N\gamma_\beta\})\le N\gamma_\beta,
$$

where

$$
\gamma_\beta=\frac{\hat{\gamma}_{\beta-}\lambda_{\beta_j}w_{j-1}(x_{\beta[j-1]})}{\sum\limits_{a\in\bar{\mathcal{B}}^j}\hat{\gamma}_{\alpha-}\lambda_{\alpha_j}w_{j-1}(x_{\alpha[j-1]})}.
$$

We define

$$\hat{\rho}_T^{n,N} \triangleq \left(\prod_{l=1}^{n}\sigma_l\right)\sum_{\beta\in\bar{\mathcal{B}}^n}\hat{\gamma}_\beta\varphi\left(x_\beta\right)\delta_{x_\beta},$$

where $\sigma_1 \triangleq \frac{1}{N}\sum_{\beta\in\mathcal{B}^1}\lambda_\beta w_0(x)$ and $\sigma_l \triangleq \frac{1}{N}\sum\limits_{\beta\in\bar{\mathcal{B}}^l}\hat{\gamma}_{\beta-}\lambda_{\beta_l}w_{l-1}(x_{\beta[l-1]}), l = 2, ..., n.$

Theorem 6.4 *There exists a positive random variable $\hat{C} = \hat{C}\left(m, T, \varphi, Y\right)$ such that*

$$\mathbb{E}_{\tilde{\mathbb{P}}}[(\bar{\rho}_T^n\left(\varphi\right) - \hat{\rho}_T^{n,N}\left(\varphi\right))^2|\mathcal{Y}_T] \le \frac{\hat{C}}{N},$$

for all $\varphi \in C_b^{m+2}\left(\mathbb{R}^d\right)$. Moreover, $\hat{C}\left(m, T, \varphi, Y\right)$ is integrable with respect to $\tilde{\mathbb{P}}$.

Proof. Define

$$\bar{\Psi}_T^j(\varphi) \triangleq \sum_{\beta\in\mathcal{B}^j}\lambda_\beta w_0(x_{\beta[0]})\cdots w_{j-1}(x_{\beta[j-1]})\varphi\left(x_\beta\right), \quad j = 1, ..., n,$$

$$\hat{\Psi}_T^j(\varphi) \triangleq \left(\prod_{l=1}^{j}\sigma_l\right)\sum_{\beta\in\bar{\mathcal{B}}^j}\hat{\gamma}_\beta\varphi\left(x_\beta\right), \quad j = 1, ..., n,$$

and note that $\bar{\Psi}_T^n(\varphi) = \bar{\rho}_T^n\left(\varphi\right)$ and $\hat{\Psi}_T^n(\varphi) = \hat{\rho}_T^{n,N}\left(\varphi\right)$. We shall proceed by induction.

- Case $j = 1$. By the TBBA properties of $\{\hat{\gamma}_\beta\}_{\beta\in\bar{\mathcal{B}}^1}$, we get

$$\begin{aligned}
&\mathbb{E}_{\tilde{\mathbb{P}}}[(\hat{\Psi}_T^1(\varphi) - \bar{\Psi}_T^1(\varphi))^2|\mathcal{Y}_T] \\
&= \mathbb{E}_{\tilde{\mathbb{P}}}[\sigma_1^2\left(\sum_{\beta\in\bar{\mathcal{B}}^1}(\frac{\lambda_\beta w_0\left(x_{\beta[0]}\right)}{\sigma_1} - \hat{\gamma}_\beta)\varphi\left(x_\beta\right)\right)^2|\mathcal{Y}_T] \\
&\le \sigma_1^2\sum_{\beta\in\bar{\mathcal{B}}^1}\mathbb{E}_{\tilde{\mathbb{P}}}[(N\gamma_\beta - \hat{\gamma}_\beta)^2\varphi^2\left(x_\beta\right)|\mathcal{Y}_T] \\
&= \sigma_1^2\sum_{\beta\in\bar{\mathcal{B}}^1}\{N\gamma_\beta\}(1 - \{N\gamma_\beta\})\varphi^2\left(x_\beta\right) \\
&\le \frac{\left(\sum_{\beta\in\mathcal{B}^1}\lambda_\beta w_0(x_{\beta[0]})\right)^2}{N}\sum_{\beta\in\bar{\mathcal{B}}^1}\gamma_\beta\varphi^2\left(x_\beta\right) \\
&\le \frac{\bar{\Psi}_T^1\left(\mathbf{1}\right)\bar{\Psi}_T^1\left(\varphi^2\right)}{N}.
\end{aligned}$$

  Note that, $\bar{\Psi}_T^1\left(\varphi^2\right) \le \|\varphi\|_\infty^2\left(\bar{\Psi}_T^1\left(\mathbf{1}\right)\right)$. Moreover,

$$\left(\bar{\Psi}_T^1\left(\mathbf{1}\right)\right)^2 = w_0(x)^2 = \exp(\sum_{j=1}^{k}2\{h^j(x)(Y_{t_1}^j - Y_{t_0}^j) - \frac{T}{2n}(h^j(x))^2\}).$$

  Hence,

$$\mathbb{E}_{\tilde{\mathbb{P}}}[\left(\bar{\Psi}_T^1\left(\mathbf{1}\right)\right)^2] = \mathbb{E}_{\tilde{\mathbb{P}}}[\exp(\sum_{j=1}^{k}2\{h^j(x)(Y_{t_1}^j - Y_{t_0}^j) - \frac{T}{2n}(h^j(x))^2\})]$$

$$= \prod_{j=1}^{k} \mathbb{E}_{\tilde{\mathbb{P}}}[\exp\{2h^j(x)(Y_{t_1}^j - Y_{t_0}^j) - \frac{T}{n}(h^j(x))^2\}]$$

$$= \exp\left(\frac{T}{n}\sum_{j=1}^{k}(h^j(x))^2\right) \leq \exp(kT\left\|h\right\|_\infty^2)$$

- Case $j = n$. We can write

$$\bar{\Psi}_T^n(\varphi)$$
$$= \sum_{\beta\in\mathcal{B}^n} \lambda_\beta w_0(x_{\beta[0]})\cdots w_{n-1}(x_{\beta[n-1]})\varphi\left(x_\beta\right)$$
$$= w_0\left(x_{\beta[0]}\right)\bar{R}_{T/n}^{1,n}\varphi = w\left(x_{\beta[0]}\right)\bar{R}_{T/n}^{1,n-1}(w_{n-1}\bar{R}_{T/n}^n\varphi) = \bar{\Psi}_T^{n-1}(w_{n-1}\bar{R}_{T/n}^n\varphi)$$

and $\hat{\Psi}_T^n(\varphi) = \left(\prod_{l=1}^n \sigma_l\right)\sum_{\beta\in\bar{\mathcal{B}}^n}\hat{\gamma}_\beta\varphi(x_\beta)$. Consider the auxiliary random measure

$$\alpha_T^n\left(\varphi\right) \triangleq \mathbb{E}_{\tilde{\mathbb{P}}}[\hat{\Psi}_T^n(\varphi)|\mathcal{Y}_T\vee\mathcal{G}_{n-1}]$$
$$= \left(\prod_{l=1}^{n-1}\sigma_l\right)\sum_{\beta\in\bar{\mathcal{B}}^n}\hat{\gamma}_{\beta-}\lambda_{\beta_n}w_{n-1}(x_{\beta[n-1]})\varphi(x_\beta)$$
$$= \hat{\Psi}_T^{n-1}(w_{n-1}\sum_{\alpha\in\bar{B}^1}\lambda_\alpha\varphi\left(x_{\beta[n-1]*\alpha}\right)) = \hat{\Psi}_T^{n-1}(w_{n-1}\bar{R}_{T/n}^n\varphi).$$

Then,

$$\mathbb{E}_{\tilde{\mathbb{P}}}[(\bar{\Psi}_T^n(\varphi) - \hat{\Psi}_T^n(\varphi))^2|\mathcal{Y}_T]$$
$$= \mathbb{E}_{\tilde{\mathbb{P}}}[(\bar{\Psi}_T^n(\varphi) - \alpha_T^n\left(\varphi\right))^2|\mathcal{Y}_T] + \mathbb{E}_{\tilde{\mathbb{P}}}[(\alpha_T^n\left(\varphi\right) - \hat{\Psi}_T^n(\varphi))^2|\mathcal{Y}_T]$$
$$\triangleq A_1 + A_2.$$

because

$$\mathbb{E}_{\tilde{\mathbb{P}}}[(\bar{\Psi}_T^n(\varphi) - \alpha_T^n\left(\varphi\right))(\alpha_T^n\left(\varphi\right) - \hat{\Psi}_T^n(\varphi))|\mathcal{Y}_T]$$
$$= \mathbb{E}_{\tilde{\mathbb{P}}}[(\bar{\Psi}_T^n(\varphi) - \alpha_T^n\left(\varphi\right))\mathbb{E}_{\tilde{\mathbb{P}}}[(\alpha_T^n\left(\varphi\right) - \hat{\Psi}_T^n(\varphi))|\mathcal{Y}_T\vee\mathcal{G}_{n-1}]]$$
$$= 0,$$

by the definition of $\alpha_T^n\left(\varphi\right)$. Note that

$$A_1 = \mathbb{E}_{\tilde{\mathbb{P}}}[(\bar{\Psi}_T^{n-1}(w_{n-1}\bar{R}_{T/n}^n\varphi) - \hat{\Psi}_T^{n-1}(w_{n-1}\bar{R}_{T/n}^n\varphi))^2|\mathcal{Y}_T],$$

and by the induction hypothesis we get

$$\mathbb{E}_{\tilde{\mathbb{P}}}[(\bar{\Psi}_T^{n-1}(\varphi) - \hat{\Psi}_T^{n-1}(\varphi))^2|\mathcal{Y}_T] \leq \frac{\bar{\Psi}_T^{n-1}\left(\mathbf{1}\right)\bar{\Psi}_T^{n-1}\left(\varphi^2\right)}{N}.$$

Hence,

$$A_1 \leq \frac{\bar{\Psi}_T^{n-1}\left(\mathbf{1}\right)\bar{\Psi}_T^{n-1}((w_{n-1}\bar{R}_{T/n}^n\varphi)^2)}{N}$$

$$
\begin{aligned}
&\le \frac{\|\varphi\|_\infty^2 \, \bar{\Psi}_T^{n-1}(\mathbf{1}) \, \bar{\Psi}_T^{n-1}((w_{n-1}\bar{R}_{T/n}^n \mathbf{1})^2)}{N} \\
&= \frac{\|\varphi\|_\infty^2 \, \bar{\Psi}_T^{n-1}(\mathbf{1}) \, \bar{\Psi}_T^{n-1}(w_{n-1}^2)}{N},
\end{aligned}
$$

because $\bar{R}_{T/n}^n \mathbf{1} = \mathbf{1}.$ For the term A_2, we can write

$$
\begin{aligned}
A_2 &= \mathbb{E}_{\tilde{\mathbb{P}}}[\left(\prod_{l=1}^n \sigma_l^2\right) \\
&\quad \times \mathbb{E}_{\tilde{\mathbb{P}}}[\left(\sum_{\beta\in\bar{\mathcal{B}}^n} (\frac{\hat{\gamma}_{\beta-}\lambda_{\beta_n} w_{n-1}(x_{\beta[n-1]})}{\sigma_n} - \hat{\gamma}_\beta)\varphi(x_\beta)\right)^2 |\mathcal{Y}_T \vee \mathcal{G}_{n-1}]|\mathcal{Y}_T] \\
&\le \mathbb{E}_{\tilde{\mathbb{P}}}[\left(\prod_{l=1}^n \sigma_l^2\right) \sum_{\beta\in\bar{\mathcal{B}}^n} \mathbb{E}_{\tilde{\mathbb{P}}}[(N\gamma_\beta - \hat{\gamma}_\beta)^2 |\mathcal{Y}_T \vee \mathcal{G}_{n-1}]\varphi^2(x_\beta)],
\end{aligned}
$$

where we have used the $\mathcal{G}_{n-1}$ measurability of $\{\sigma_l\}_{l=1}^n$ and that the random variables $\hat{\Gamma}_n$ conditionally to $\mathcal{G}_{n-1}$ are negatively correlated. Using now the TBBA properties of $\hat{\Gamma}_n$ we have that

$$
\begin{aligned}
A_2 &\le \mathbb{E}_{\tilde{\mathbb{P}}}[\left(\prod_{l=1}^n \sigma_l^2\right) \sum_{\beta\in\bar{\mathcal{B}}^n} \{N\gamma_\beta\}(1-\{N\gamma_\beta\})\varphi^2(x_\beta)\,|\mathcal{Y}_T] \\
&\le \mathbb{E}_{\tilde{\mathbb{P}}}[\left(\prod_{l=1}^{n-1} \sigma_l^2\right) \sigma_n^2 N \sum_{\beta\in\bar{\mathcal{B}}^n} \gamma_\beta \varphi^2(x_\beta)\,|\mathcal{Y}_T] \\
&\le \frac{\|\varphi\|_\infty^2}{N}\mathbb{E}_{\tilde{\mathbb{P}}}[\left(\prod_{l=1}^{n-1} \sigma_l^2\right)\left(\sum_{\beta\in\bar{\mathcal{B}}^n} \hat{\gamma}_{\beta-}\lambda_{\beta n} w_{n-1}(x_{\beta[n-1]})\right)^2 |\mathcal{Y}_T].
\end{aligned}
$$

Note that

$$
\left(\prod_{l=1}^{n-1} \sigma_l\right) \sum_{\beta\in\bar{\mathcal{B}}^n} \hat{\gamma}_{\beta-}\lambda_{\beta_n} w(x_{\beta[n-1]}) = \hat{\Psi}_T^{n-1}(w_{n-1}\bar{R}_{T/n}^n(\mathbf{1})) = \hat{\Psi}_T^{n-1}(w_{n-1}),
$$

because $\bar{R}_{T/n}^n(\mathbf{1}) = \mathbf{1}$. By taking iteratively conditional expectations with respect to $\mathcal{Y}_T \vee \mathcal{G}_j, j = 1, n-2$, it is easy to see that $\mathbb{E}_{\tilde{\mathbb{P}}}[\hat{\Psi}_T^{n-1}(w_{n-1})] = \bar{\Psi}_T^{n-1}(w_{n-1})$. Hence, using again the induction hypothesis we get that

$$
\mathbb{E}_{\tilde{\mathbb{P}}}[(\hat{\Psi}_T^{n-1}(w_{n-1}))^2|\mathcal{Y}_T] \le (\bar{\Psi}_T^{n-1}(w_{n-1}))^2 + \frac{\bar{\Psi}_T^{n-1}(\mathbf{1})\,\bar{\Psi}_T^{n-1}\left(w_{n-1}^2\right)}{N},
$$

which yields

$$
\begin{aligned}
A_2 &\le \frac{\|\varphi\|_\infty^2 (\bar{\Psi}_T^{n-1}(w_{n-1}))^2}{N} + \frac{\|\varphi\|_\infty^2 \, \bar{\Psi}_T^{n-1}(\mathbf{1})\,\bar{\Psi}_T^{n-1}\left(w_{n-1}^2\right)}{N^2} \\
&\le \|\varphi\|_\infty^2 \frac{(\bar{\Psi}_T^n(\mathbf{1}))^2 + \bar{\Psi}_T^{n-1}(\mathbf{1})\,\bar{\Psi}_T^{n-1}\left(w_{n-1}^2\right)}{N}.
\end{aligned}
$$

Combining the bounds for A_1 and A_2 we get that

$$\mathbb{E}_{\tilde{\mathbb{P}}}[(\bar{\Psi}_T^n(\varphi)-\hat{\Psi}_T^n(\varphi))^2|\mathcal{Y}_T] \leq \frac{\|\varphi\|_\infty^2}{N}\left\{2\bar{\Psi}_T^{n-1}(\mathbf{1})\,\bar{\Psi}_T^{n-1}(w_{n-1}^2)+\left(\bar{\Psi}_T^n(\mathbf{1})\right)^2\right\},$$

and, therefore,

$$\hat{C}\left(m,T,\varphi,Y\right)=\|\varphi\|_\infty^2\left\{2\bar{\Psi}_T^{n-1}(\mathbf{1})\,\bar{\Psi}_T^{n-1}(w_{n-1}^2)+\left(\bar{\Psi}_T^n(\mathbf{1})\right)^2\right\}$$

The integrability of $\hat{C}\left(m,T,\varphi,Y\right)$ follows by using similar arguments as in the proof of Lemma 4.10.

■

Definition 6.5 *We define* $\hat{\pi}_T^{n,N}$ *by*

$$\hat{\pi}_T^{n,N} \triangleq \frac{\hat{\rho}_T^{n,N}(\varphi)}{\hat{\rho}_T^{n,N}(\mathbf{1})}=\frac{1}{N}\sum_{\beta\in\bar{\mathcal{B}}^n}\hat{\gamma}_\beta\varphi\left(x_\beta\right). \tag{6.3}$$

Theorem 6.4 yields the following result.

Corollary 6.6 *There exists a positive random variable* $\hat{K}=\hat{K}\left(m,T,\varphi\right)$ *such that* $\mathbb{E}_{\tilde{\mathbb{P}}}[|\bar{\pi}_T^n\left(\varphi\right)-\hat{\pi}_T^{n,N}\left(\varphi\right)||\mathcal{Y}_T]\leq\frac{\hat{K}(m,T,\varphi,Y)}{N^{1/2}}$, *for all* $\varphi\in C_b^{m+2}\left(\mathbb{R}^d\right)$. *Moreover,* $\hat{K}\left(m,T,\varphi,Y\right)$ *is integrable with respect to* $\tilde{\mathbb{P}}$.

Proof. The proof is the same as the one for Corollary 6.3, making the obvious changes and using Theorem 6.4. ■

6.3 Main result for the KLV particle filter

The main result of the paper is the following:

Theorem 6.7 *There exists a positive constants* $\tilde{c}=\tilde{c}\left(m,T,\varphi\right)$ *and* $\hat{c}=\hat{c}\left(m,T,\varphi\right)$ *such that, for all* $\varphi\in C_b^{m+2}\left(\mathbb{R}^d\right)$, *we have*

$$\mathbb{E}_{\tilde{\mathbb{P}}}[|\pi_T\left(\varphi\right)-\tilde{\pi}_T^{n,N}\left(\varphi\right)|]\leq\tilde{c}(\frac{1}{N^{1/2}}+\frac{1}{n}),$$

and

$$\mathbb{E}_{\tilde{\mathbb{P}}}[|\pi_T\left(\varphi\right)-\hat{\pi}_T^{n,N}\left(\varphi\right)|]\leq\hat{c}(\frac{1}{N^{1/2}}+\frac{1}{n}),$$

where $\tilde{\pi}_T^{n,N}$ *and* $\hat{\pi}_T^{n,N}$ *are defined in* (6.2) *and* (6.3), *respectively.*

Proof. Using the triangular inequality we have that

$$\mathbb{E}_{\tilde{\mathbb{P}}}[|\pi_T\left(\varphi\right)-\tilde{\pi}_T^{n,N}\left(\varphi\right)|]\leq\mathbb{E}_{\tilde{\mathbb{P}}}[|\pi_T\left(\varphi\right)-\bar{\pi}_T^n\left(\varphi\right)|]+\mathbb{E}_{\tilde{\mathbb{P}}}[|\bar{\pi}_T^n\left(\varphi\right)-\tilde{\pi}_T^{n,N}\left(\varphi\right)|].$$

By Corollary 4.7, it follows that

$$\mathbb{E}_{\tilde{\mathbb{P}}}[|\pi_T\left(\varphi\right)-\bar{\pi}_T^n\left(\varphi\right)|]\leq\frac{C}{n}$$

and, by Corollary 6.3, we have that

$$\begin{aligned}\mathbb{E}_{\tilde{\mathbb{P}}}[|\bar{\pi}_T^n\left(\varphi\right)-\tilde{\pi}_T^{n,N}\left(\varphi\right)|]&\leq\mathbb{E}_{\tilde{\mathbb{P}}}[\mathbb{E}_{\tilde{\mathbb{P}}}[|\bar{\pi}_T^n\left(\varphi\right)-\tilde{\pi}_T^{n,N}\left(\varphi\right)||\mathcal{Y}_T]]\\&\leq\frac{1}{N^{1/2}}\mathbb{E}_{\tilde{\mathbb{P}}}[\tilde{K}\left(m,T,\varphi,Y\right)]=\frac{\tilde{c}(m,T,\varphi)}{N^{1/2}},\end{aligned}$$

which yields the result. The proof for $\hat{\pi}_T^{n,N}$ is analogous. ■

7 Numerical simulations

In this section we present some numerical experiments to test the performance of the algorithms we have introduced. The model chosen is a particular case of the Beneš problem. This is a stochastic filtering problem with a nonlinear dynamics for the signal and a linear dynamics for the observation process. This problem has an analytical finite dimensional solution, known as the Beneš filter. In particular the conditional distribution of the signal, given the observation process, is equal to a Gaussian mixture. Although the model do not satisfy some of the conditions for which our theory holds (the function h and the derivatives of f are not bounded) we believe that is good benchmark to test the new algorithms as it allows for sufficient rich (non-linear) behaviour while still having a closed form expression for its solution.

7.1 The model and its exact solution

The dynamics of the signal is given by the following one-dimensional SDE

$$X_t = x_0 + \int_0^t f(X_s)ds + \sigma V_t, \quad 0 \le t \le T, \tag{7.1}$$

where $f(x) = \mu\sigma \tanh(\frac{\mu}{\sigma}x)$, and the observation process is given by the one-dimensional process $Y_t = \int_0^t h(X_s)ds + W_t$, where $h(x) = h_1 x + h_2, V_t$ and W_t are two independent, one-dimensional Brownian motions, x_0, μ, h_1 and $h_2 \in \mathbb{R}$ and $\sigma > 0$. The conditional law of X_t given $\mathcal{Y}_t$ in the previous problemhas an exact expression. It is a weighted mixture of two normal distributions, see Chapter 6 in [1]. That is, given a realization $Y = \{Y_s\}_{0 \le s \le t}$ of the observation process, we have the following equality in law for X_t given $\mathcal{Y}_t$,

$$\pi_t = w_t^+ \mathcal{N}(A_t^+/(2B_t), 1/(2B_t)) + w_t^- \mathcal{N}(A_t^-/(2B_t), 1/(2B_t)),$$

where

$$\begin{aligned}
w_t^{\pm} &\triangleq \exp\left((A_t^{\pm})^2/(4B_t)\right)/(\exp\left((A_t^+)^2/(4B_t)\right) + \exp\left((A_t^-)^2/(4B_t)\right)) \\
A_t^{\pm} &\triangleq \pm\frac{\mu}{\sigma} + h_1\Psi_t + \frac{h_2 + h_1 x_0}{\sigma\sinh(h_1\sigma t)} - \frac{h_2}{\sigma}\coth(h_1\sigma t), \\
B_t &\triangleq \frac{h_1}{2\sigma}\coth(h_1\sigma t), \quad \Psi_t \triangleq \int_0^t \frac{\sinh(h_1\sigma s)}{\sinh(h_1\sigma t)} dY_s,
\end{aligned}$$

and $\mathcal{N}(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2. Recall that in practice we only observe Y at a finite partition $\Pi^{n,T} = \{0 = t_0 < t_1 < \cdots < t_{n-2} < t_{n-1} = T\}$ of the interval $[0, T]$. Hence, we have to approximate the integral in the definition of Ψ_t using a Riemann-Stieltjes sum. In particular, we will use $\Psi_{t_i} \simeq \sum_{k=0}^{i-1} \frac{\sinh(h_1\sigma t_{k+1})}{\sinh(h_1\sigma t_i)}(Y_{t_{k+1}} - Y_{t_k})$, for all $t_i \in \Pi^{n,T}$.

From the expression for the density of the posterior distribution π_t, one can compute h_{Xt} the density of the signal X_t, which we will call the prior. Setting $h_2 = 0$ and taking limits when h_1 tends to zero one has that

$$\mathrm{Law}(X_t) = w_t^+ \mathcal{N}(\bar{A}_t^+/(2\bar{B}_t), 1/(2\bar{B}_t)) + w_t^- \mathcal{N}(\bar{A}_t^-/(2\bar{B}_t), 1/(2\bar{B}_t)),$$

where

$$\bar{w}_t^{\pm} \triangleq \exp\left((\bar{A}_t^{\pm})^2/(4\bar{B}_t)\right)/(\exp\left((\bar{A}_t^+)^2/(4\bar{B}_t)\right) + \exp\left((\bar{A}_t^-)^2/(4\bar{B}_t)\right))$$

$$\bar{A}_t^{\pm} \triangleq \frac{x_0 \pm \mu\sigma t}{\sigma^2 t}, \quad \bar{B}_t \triangleq \frac{1}{2\sigma^2 t}.$$

7.2 Numerical experiments

In the numerical experiments we set certain values for the parameters $\mu, \sigma, h_1, h_2, x_0$ and T, and we compute a realisation of X_t and Y_t using the Euler scheme and an equidistant partition $\Pi^{m,T} = \{t_i = \frac{i}{m}T\}_{i=0,\ldots,m}$ with $m = 10^6$. We will call the realisation of X_t the seed particle. All the other simulations will be done assuming that we are given the fixed path Y_t, computed from the seed particle path. Usually we will consider partitions $\Pi^{n,T}$, with $n < m$ and the values of Y_t in these partitions will be obtained using linear interpolation of the values of Y_t in $\Pi^{m,T}$.

Using the previously simulated discrete path of Y, we can approximate the integral Ψ_t and compute the values of $w_t^{\pm}, A_t^{\pm}, B_t\ \bar{w}_t^{\pm}, \bar{A}_t^{\pm}$ and $\bar{B}_t$ for $t \in \Pi^{m,T}$, obtaining the parameters of the normal mixtures for the prior and the posterior at each $t \in \Pi^{m,T}$. Hence, we can plot the exact densities of the prior and the posterior and compute any expectation with respect to them, at each $t \in \Pi^{m,T}$.

7.2.1 Visualising the approximation

A first experiment to actually visualize the approximation mechanism is the following. We use the following set of parameter values

$$\mu = 2.2, \quad h_1 = 0.15, \quad h_2 = 0.0, \quad \sigma = 2.2, x_0 = 0.0 \quad T = 10.0.$$

In Figure 1, we plot the density of the prior at time $t = 8.3$. We also plot an histogram of the position of the particles as well as its actual positions, represented by small vertical bars below the horizontal axis. In addition, we draw a triangle which represents the position of the seed particle. To compute the law of the signal is equivalent to solve the stochastic filtering problem with the sensor function h being equal to zero, that is, making all the filtering weights equal to 1. Hence, the prior and posterior densities coincide in this case. Figure 1 points out the main issue that any particle filter without resampling is going to face when solving this problem. Note that the law of the signal is a symmetric bimodal distribution with the distance between the two modes increasing as time increases. Eventually, the distribution of the signal consists in two non-overlapping peaks. This means that the seed particle will be in any of these two peaks with equal probability and will remain there forever. As no observation data is available, the prior distribution itself offers a bad approximation to the signal as half of the mass and, hence, half of the particles will always be placed in the wrong peak, that is, in the one that does not contain the signal.

In Figure 2, we plot the prior and posterior densities at time $t = 8.3$. We also plot the histogram of the particle approximation of the posterior density. We use the first version of the KLV-filter where the TBBA is only applied to the cubature weights (the data is not taken into account). Here half of the particles are still situated in the wrong place and are assigned small weights. Hence, half of the computational effort is wasted in maintaining alive particles that are contributing zero to the approximation.

In Figure 3, we plot the same elements of Figure 2 but using the second version of the KLV-filter with resampling. As a result, now all the particles

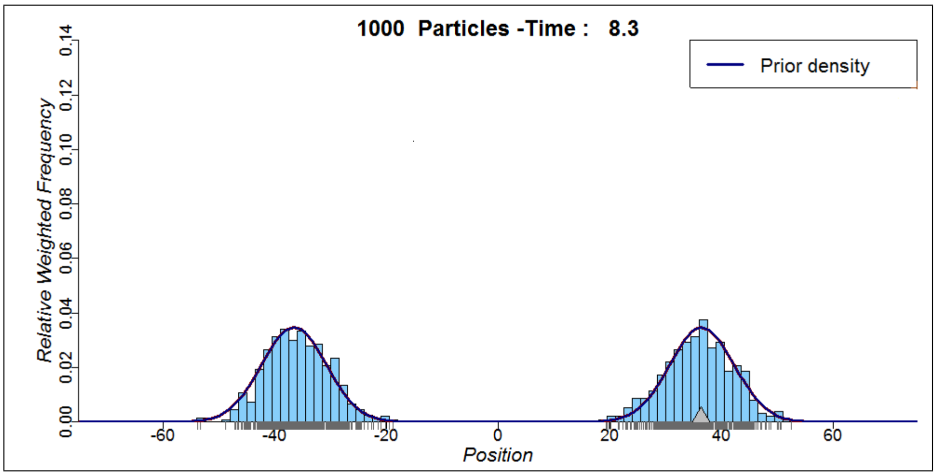


Figure 1: Plot of the exact density of the signal X_t and the histogram of its particle approximation, both at $t = 8.3$.

are placed on the correct peak, the one with the signal, and contribute with a non-null weights to the approximation.

7.2.2 Convergence of the KLV particle filter

In this section we investigate numerically the convergence of the KLV filter in terms of the number or time steps in the partition and the number of particles. We use the following set of parameter values

$$\mu = 0.5, \quad h_1 = 0.4, \quad h_2 = 0.0, \quad \sigma = 0.8, x_0 = 1.0 \quad T = 10.0.$$

We will estimate $\pi_T(\varphi)$ where $\varphi(x) = x$, that is, the conditional expectation of the signal at time T given the observation process up to time T. We take as the exact value for $\pi_T(\varphi)$ the value given by the Beneš filter $\mathbb{E}[X_t|\mathcal{Y}_t] = \pi_T(\varphi) = (w_T^+ A_T^+ + w_T^- A_T^-)/(2B_T)$. First, we estimate $\pi_T(\varphi)$ using $\hat{\pi}_T^{n,N}(\varphi)$, where the number of particles $N = 10^5$ and we choose various values for n, the number of steps in the discretization partition. We compute $\hat{\pi}_T^{n,N}(\varphi)$ using cubature formulas of degree 3 and 5. In Figure 4, we plot the logarithm of the absolute error in the estimation of $\pi_T(\varphi)$ by $\hat{\pi}_T^{n,N}(\varphi)$ against the logarithm of the number of time steps. Both, the cubature formula of degree 5 and degree 3 give a rate of convergence of one. Hence, it is clear that the discretisation error of Picard's filter is a lower bound for the discretisation error of the method, even when one uses a degree 5 (order 2 for approximating the law of the signal) cubature formula.

In Fig 5, we plot the absolute error obtained using $\hat{\pi}_T^{n,N}(\varphi)$ with $n = 150$ fixed and varying the number of particles N.

The number of particles used in the simulation ranges from 2 to $2^{20} \simeq 1048576$. Apparently, with $2^{15} = 32768$ particles we already obtain a good partial sampling of the cubature tree (we get close to the discretization error) and there is no significant improvement in using a larger number of particles.

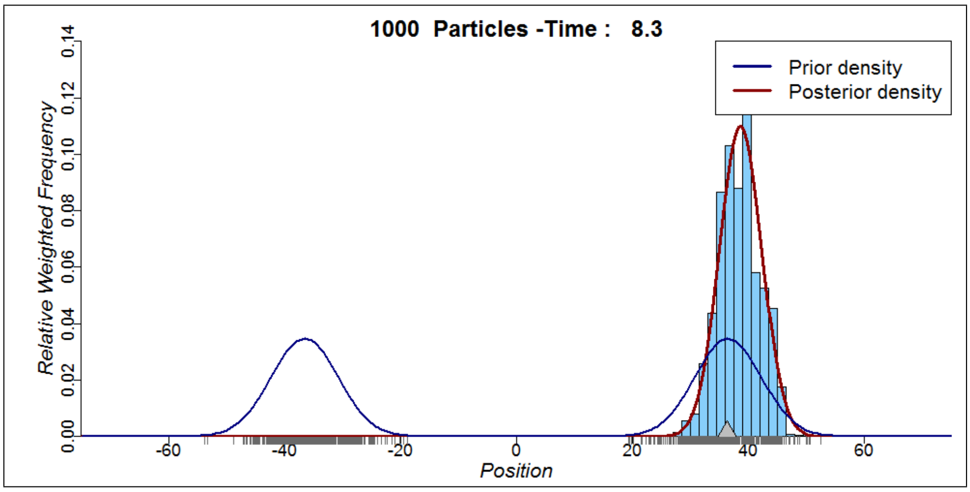


Figure 2: Plot of the exact density of X_t, the exact density of $X_t|Y_t$ and the histogram of its particle approximation, both at $t = 8.3$.
First version of the KLV particle filter.

7.2.3 Comparison with the classical particle filter based on Euler approximation

In this section we compare the performance of our algorithm against the classical particle filter implemented using the Euler scheme to approximate the signal and the TBBA to perform the resampling at each time step. We will denote by $\mathring{\pi}_T^{n,N}(\varphi)$ the classical particle filter. In what follows, we add a subindex r indicating the result of the r-th independent launch of both algorithms, $r = 1, ..., M$. Hence, our approximations will be

$$\hat{\pi}_T^{n,N,M}(\varphi) = \frac{1}{M}\sum_{r=1}^{M}\hat{\pi}_{T,r}^{n,N}(\varphi), \quad \mathring{\pi}_T^{n,N,M}(\varphi) = \frac{1}{M}\sum_{r=1}^{M}\mathring{\pi}_{T,r}^{n,N}(\varphi).$$

We set the same number of launches $M = 10$ and the same number of steps in the partition for both estimators. However, the number of particles used in our algorithm will be just $N = 100$ while we set $N = 10000$ for the classical particle filter. We use the following set of parameter values

$$\mu = 0.05, \quad h_1 = 0.8, \quad h_2 = 0.0, \quad \sigma = 1.0, x_0 = 0.0 \quad T = 20.0.$$

The following graph, Figure 6, plots the absolute error of the estimates obtained using the classical particle filter (Euler) and the new algorithm with cubature formulas of degree 3 and 5 against the CPU time used in each computation.

One can see from Figure 6, that the KLV particle filter obtains better errors with less computational time. In addition, it seems that the KLV particle filter is more robust to the resampling procedure than the classical particle filter, in the sense that the additional randomness added by resampling does not increase the error when using a large number of time steps.

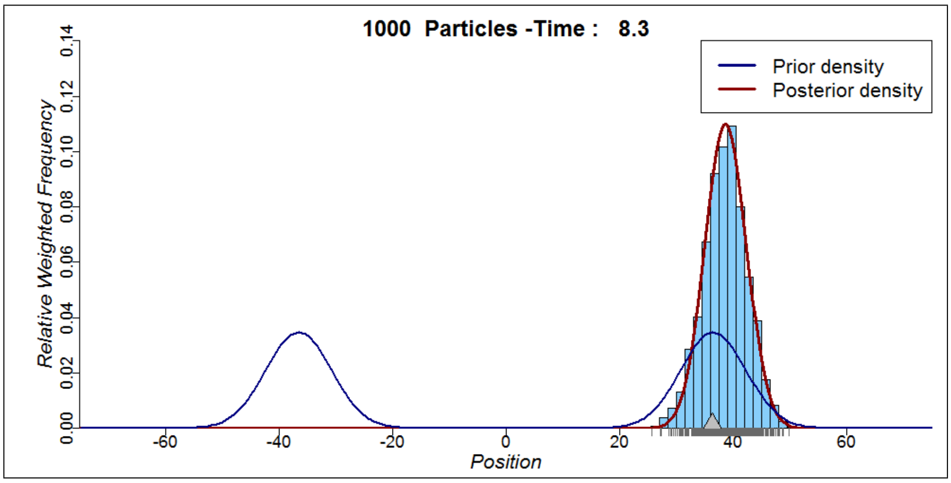


Figure 3: Plot of the exact density of X_t, the exact density of $X_t|Y_t$ and the histogram of its particle approximation, both at $t = 8.3$.
Second version of the KLV particle filter.

Acknowledgements

The second author acknowledges the support of the Commission for Universities and Research of the Department of Innovation, Universities and Enterprise of the Government of Catalonia, grant BP-DGR 2009.

References

[1] Bain, A. & Crisan, D. 2009 *Fundamentals of Stochastic Filtering.* Stochastic Modelling and Applied Probability. Springer, Berlin.

[2] Cass, T. & Litterer, C. 2011 On the error estimate for cubature on Wiener space. arXiv:1105.3996v1.

[3] Clark, J. M. C. 1978 The design of robust approximations to the stochastic differential equations of nonlinear filtering. *Communication systems and random process theory* (Proc. 2nd NATO Advanced Study Inst., Darlington, 1977), pp. 721–734. NATO Advanced Study Inst. Ser., Ser. E: Appl. Sci., No. **25**.

[4] Crisan, D. 2011 Discretizing the Continuous Time Filtering Problem. Order of Convergence. *OUP Handbook on Nonlinear Filtering*, Eds Dan Crisan and Boris Rozovskii, pp. 572-597, Oxford University Press, Oxford.

[5] Crisan, D. & Ghazali, S. 2007 On the convergence rates of a general class of weak approximations of SDEs. *Stochastic differential equations: theory and applications*, pp. 221–248. Interdiscip. Math. Sci., 2, World Sci. Publ., Hackensack, NJ, 2007.

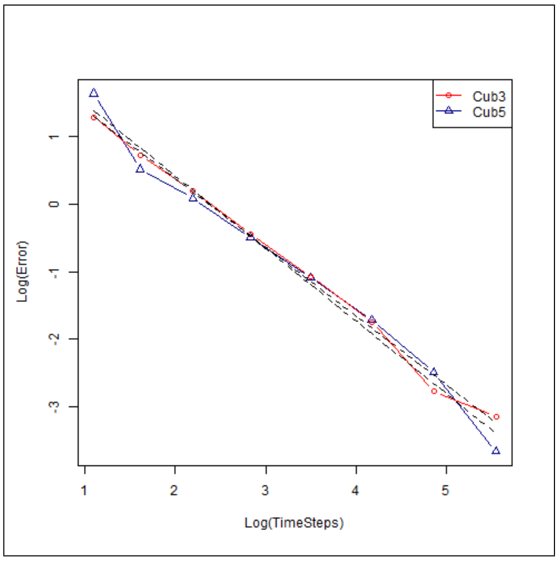


Figure 4: Absolute Error vs Number of Time Steps

[6] Crisan, D. & Lyons, T. 2002 Minimal entropy approximations and optimal algorithms. *Monte Carlo Methods Appl.* **8**, no. 4, pp. 343–355.

[7] Crisan, D. & Manolarakis, K. 2010 Solving Backward Stochastic Differential Equations using the Cubature Method. Application to Nonlinear Pricing. *Progress in analysis and its applications*, pp. 389–397, World Sci. Publ., Hackensack, NJ.

[8] Crisan, D. & Obanubi, O.2012 Particle Filters with Random Resampling Times. *Stochastic Processes and their Applications* **122**, no.4, pp. 11332–1368

[9] Del Moral, P., Doucet, A. & Jasra, A. 2012 On adaptive resampling strategies for sequential Monte Carlo methods. *Bernoulli* **18**, no.1, pp. 252–278.

[10] Gordon, N. J., Salmond, D. J. & Smith, A. F. 1993 Novel approach to nonlinear/non-gaussian Bayesian state estimator. *Radar and Signal Processing. IEE Proceedings F*, **140**, pp. 107-113.

[11] Gyurkó, L. & Lyons, T. 2011 Efficient and Practical Implementations of Cubature on Wiener Space. *Stochastic Analysis 2010*. Springer-Verlag Berlin Heidelberg.

[12] Kallianpur, G. and Striebel, C., 1968, Estimation of stochastic systems: Arbitrary system process with aditive white noise observation errors. *Ann. Math. Statist.*, 39 (3), pp. 785-801.

[13] Kusuoka, S. 2001 Approximation of expectations of diffusion processes and mathematical finance. *Advanced studies in Pure Mathematics* **31**, pp. 147-165.

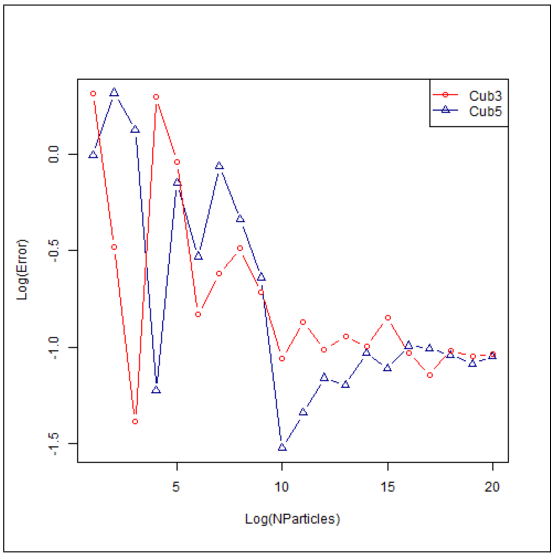


Figure 5: Absolut Error vs Number of Particles

[14] Kusuoka, S. 2003 Approximation of expectations of diffusion processes based on Lie algebra and Malliavin calculus. *UTMS* **34**

[15] Kusuoka, S. 2003 Malliavin Calculus Revisited. *J. Math. Sci. Univ. Tokyo* **10**, pp. 261-277.

[16] Kusuoka, S. & Stroock, D. 1987 Application of the Malliavin calculus III. *J. Fac. Sci. Univ. Tokyo 1A* **34**, pp. 391-442.

[17] Litterer, C. & Lyons,T. 2011 Introducing Cubature to Filtering. *The Oxford Handbook of Nonlinear Filtering*. Oxford University Press, Oxford.

[18] Lyons, T. & Victoir, N. 2003 Cubature on Wiener space. *Proc. R. Soc. Lond. A* **460**, pp. 169-198.

[19] Nee C. 2011 *Sharp gradient bounds for heat-kernels and applications.* DPhil Thesis, Imperial College London.

[20] Ninomiya, S. 2003 A partial sampling method applied to the Kusuoka approximation. *Monte Carlo Methods and Appl.* **9**, no. 1, pp. 27-38.

[21] Ninomiya, M. 2010 Application of the Kusuoka approximation with a tree-based branching algorithm to the pricing of interest-rate derivatives under the HJM model. *LMS J. Comput. Math.* **13**, pp. 208-221.

[22] Picard, J. 1984 Approximation of nonlinear filtering problems and order of convergence. Filtering and control of random processes (Paris, 1983), 219–236, *Lecture Notes in Control and Inform. Sci.*, **61**, Springer, Berlin.

[23] Talay, D. 1984 Efficient numerical schemes for the approximation of expectations of functionals of the solution of a SDE and applications. Filtering

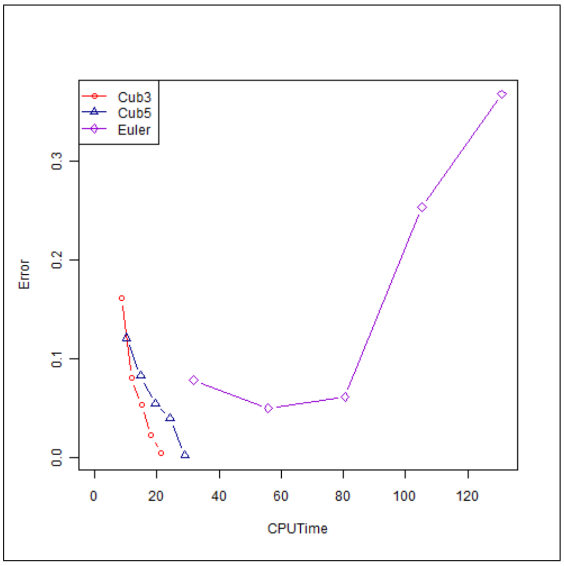


Figure 6: Comparison of the new algorithm and the classical particle filter: Absolut error vs CPU time.

and control of random processes (Paris, 1983), 294–313, *Lecture Notes in Control and Inform. Sci.*, **61**, Springer, Berlin.