UNIFORM APPROXIMATIONS
OF DISCRETE-TIME FILTERS

KARI HEINE,* *Tampere University of Technology*

DAN CRISAN,** *Imperial College London*

Abstract

Throughout recent years, various sequential Monte Carlo methods, i.e. particle filters, have been widely applied to various applications involving the evaluation of the generally intractable stochastic discrete-time filter. Although convergence results exist for finite-time intervals, a stronger form of convergence, namely, uniform convergence, is required for bounding the error on an infinite-time interval. In this paper we prove easily verifiable conditions for the filter applications that are sufficient for the uniform convergence of certain particle filters. Essentially, the conditions require the observations to be accurate enough. No mixing or ergodicity conditions are imposed on the signal process.

Keywords: Discrete-time filter; particle filter; Monte Carlo filter; uniform convergence

2000 Mathematics Subject Classification: Primary 93E11

Secondary 93E15; 60E35; 62E20

1. Introduction

In the stochastic discrete-time filtering problem, the goal is to recover an unknown realisation of a discrete-time Markov chain $X := (X_i)_{i \geq 0}$ at any given time i based on the partial information obtained by observing the realisation of an associated discrete-time process $Y := (Y_i)_{i > 0}$. The solution to this problem is the probability measure-valued discrete-time stochastic process $\pi := (\pi_i)_{i \geq 0}$, where π_i is the conditional distribution of X_i given the random variables $Y_1, Y_2, \ldots, Y_i$. In the literature, the process π is often referred to as the *optimal* or *Bayesian filter*. In more detail, the filtering framework under consideration is as follows. We assume an underlying probability space $(\Omega, \mathcal{F}, P)$ and define the *signal process* X for all $i > 0$ as

$$X_i := f_i(X_{i-1}) + W_i, \tag{1.1}$$

where X_0 is an independent random variable with a distribution P_0 on the Borel σ-field $\mathcal{B}(\mathbb{R}^{d_s})$ in the Euclidean space $\mathbb{R}^{d_s}$, W_i is an independent random variable with a distribution P_{W_i} on $\mathcal{B}(\mathbb{R}^{d_s})$, and $f_i \colon \mathbb{R}^{d_s} \to \mathbb{R}^{d_s}$ is measurable. The *observation process* Y is defined for all $i > 0$ as

$$Y_i := h_i(X_i) + V_i,$$

where $h_i \colon \mathbb{R}^{d_s} \to \mathbb{R}^{d_m}$ is measurable and V_i is an independent random variable with a distribution P_{V_i} on $\mathcal{B}(\mathbb{R}^{d_m})$. Then, by definition, the *filter process* π satisfies, for all bounded and measurable functions $\varphi \colon \mathbb{R}^{d_s} \to \mathbb{R}$ and all $i \geq 0$,

$$\int \varphi \, d\pi_i = E[\varphi(X_i) \mid \mathcal{Y}_i],$$

where $\mathcal{Y}_0 := \{\varnothing, \Omega\}$ and $\mathcal{Y}_i := \sigma(Y_j, 1 \leq j \leq i)$ is the σ-field generated by the observations $Y_1, Y_2, \ldots, Y_i$. Roughly speaking, the *filtering distribution* π_i can be considered as a representation of the knowledge of the unknown realisation of X_i given the information $\mathcal{Y}_i$. The filter process can also be expressed as a recursion

$$\pi_i = Q_i(\pi_{i-1}),$$

where $\pi_0 = \mathrm{P}_0$ and the probability measure-valued (random) mapping Q_i is specified later in Section 2. In practice, the evaluation of the mapping Q_i is intractable and, therefore, it is approximated by a mapping Q_i^θ, where $\theta \in (0, \infty)$ is the parameter of the approximation. That is, a greater value of θ implies a smaller approximation error, but also, in practice, a higher computational cost. Throughout recent years, *sequential Monte Carlo (SMC) methods*, or, *particle filters* have been widely applied to the approximation of this mapping in various practical applications. In SMC methods, the parameter θ is the size of the Monte Carlo sample, i.e. the number of particles. It has been shown for a certain class of SMC methods that, under feasible conditions, the approximation converges to the exact filter as the number of particles tends to ∞ (see, e.g. [2] and the references therein). These convergence results can be used for bounding the approximation error for a fixed parameter value on a finite-time interval. In order to bound the error on an infinite-time interval, a stronger form of convergence, namely, *uniform convergence* is required. The approximation is said to be uniformly convergent (in mean) with respect to a metric d if

$$\lim_{\theta \to \infty} \sup_{i > 0} \mathrm{E}[d(Q_i \circ \cdots \circ Q_1(\pi_0), Q_i^\theta \circ \cdots \circ Q_1^\theta(\pi_0))] = 0. \qquad (1.2)$$

One of the earliest contributions regarding the uniform convergence of SMC methods is given in [4], where it has been shown that if the filter itself is stable, the so-called *Monte Carlo particle filter* is uniformly convergent. The Monte Carlo particle filter is an SMC method that does not have a branching mechanism. For the required filter stability, it suffices to have mixing or ergodic signal kernels [1], [5]. In [14], the uniform convergence of the Monte Carlo filter has been extended to the estimation of parameters whose kernels are not mixing. It has also been shown in [5] that the mixing property or the ergodicity of the signal kernel are sufficient conditions for the uniform convergence of the *interacting particle system*, which in contrast to the Monte Carlo particle filter has interacting particles because of the incorporated branching mechanism. A somewhat different approach can be found in [10], where the uniform convergence of a *rejection-sampling*-based particle filter has been considered in the case of a mixing signal. In [12], it has been shown that the mixing condition can be relaxed by imposing the mixing condition on the kernels of the unnormalised filter process. This is a slightly weaker condition than the mixing property of the signal. Except for [10], the uniform convergence in the references above is considered in the sense of (1.2).

The main results of this paper, Theorem 3.1 and Theorem 4.1, provide easily verifiable sufficient conditions for the uniform convergence of two different filter approximations. The first approximation is called the *truncated filter* and the motivation for the analysis of this approximation is purely theoretical. In practice, the truncated filter is intractable. Nevertheless, it plays an important role in the proof of the uniform convergence of the second filter approximation which represents a certain class of feasible particle filters. The principle of the proof is to show that the given particle filter approximates uniformly the truncated filter which, in turn, is a uniform approximation of the exact filter and, hence, the particle filter approximates the exact filter uniformly. The idea of using the truncated filter in the analysis

was originally introduced in [11] and [13], and here we generalise the results reported in [13]. Other types of truncation appear frequently in the literature. For example, in [9] and [15], the underlying Markov process is truncated instead of the filtering equations. Moreover, in [13], the uniform convergence is proved for a rejection-sampling-based particle filter, while the approximation in this work employs importance sampling. There is a fundamental difference between these two approaches if the motivation for the analysis of the uniform convergence is to study the computational cost that ensures a certain level of accuracy. In the importance sampling approach, the computational cost is determined by the number of particles, while in the rejection sampling approach we must take into account the total number of samples, including the random number of rejected samples.

It is obvious from the references above that there is a close connection between the stability of the filter and the uniform convergence of its (particle) approximations. In many cases (see, e.g. [4], [5], [10], and [13]), the stability of the filter has been shown to be a sufficient condition for the uniform convergence of certain particle approximations. In this work we do not specifically discuss the filter stability, but it should be noted that the conditions that we impose on the filter applications are similar to those used in [3] to prove the stability of the filter. The following assumptions are made about the filter framework.

(A1) There exist $\alpha, \delta > 0$ such that, for all $x, y \in \mathbb{R}^{d_s}$ and $i > 0$,

$$\|f_i(x) - f_i(y)\| \leq \alpha \|x - y\| + \delta.$$

(A2) For all $i > 0$, $h_i = \tilde{h}_i + \bar{h}_i$, where $\tilde{h}_i$ is a bijection, and there exist $\beta_0 > 0$, $\beta > 0$, and $\gamma \geq 0$ such that $\tilde{h}_i$ and $\tilde{h}_i^{-1}$ are Lipschitz with coefficients β_0 and β, respectively, and $\bar{h}_i$ satisfies, for all $x \in \mathbb{R}^{d_s}$, $\|\bar{h}_i(x)\| \leq \gamma$.

(A3) There exist $m_1, M_1, a_1, A_1, b_1, B_1 > 0$ such that, for all $i > 0$,

$$m_1 \exp(-a_1 \|x\|^{b_1}) \leq \rho_{W_i}(x) \leq M_1 \exp(-A_1 \|x\|^{B_1}),$$

where ρ_{W_i} is the density of P_{W_i} with respect to the Lebesgue measure λ_{d_s} on $\mathscr{B}(\mathbb{R}^{d_s})$.

(A4) There exist $m_2, M_2, a_2, A_2, b_2, B_2 > 0$ such that, for all $i > 0$,

$$m_2 \exp(-a_2 \|x\|^{b_2}) \leq \rho_{V_i}(x) \leq M_2 \exp(-A_2 \|x\|^{B_2})$$

and

$$\rho_{X_0/\beta}(x) \leq M_2 \exp(-A_2 \|x^{B_2}\|),$$

where ρ_{V_i} is the density of P_{V_i} with respect to the Lebesgue measure λ_{d_m} on $\mathscr{B}(\mathbb{R}^{d_m})$ and $\rho_{X_0/\beta}$ is the probability density function of the random variable X_0/β with respect to the Lebesgue measure λ_{d_s} on $\mathscr{B}(\mathbb{R}^{d_s})$.

The assumption on the upper bound for the density $\rho_{X_0/\beta}$ can be omitted. In this case, however, no rates for the uniform convergence results are available and, therefore, we include it in this work (see [7] for the corresponding results without this assumption). Also, note that the function $\tilde{h}_i$ is a (topological) homeomorphism between the signal space $\mathbb{R}^{d_s}$ and the observation space $\mathbb{R}^{d_m}$. In particular, this implies that the two spaces must have the same dimension, i.e. $d_s = d_m$. We will assume that this is the case from now on. However, we will still keep the notation $\mathbb{R}^{d_s}$ for the signal space and $\mathbb{R}^{d_m}$ for the observation space to distinguish between them.

This paper is organised as follows. In Section 2 we introduce some general purpose notation and specify the truncated filter approximation. Also, some important intermediate results are given. In Section 3, the uniform convergence of the truncated filter is proved. In Section 4 we describe a general class of particle filters and provide the proof of uniform convergence. In Section 5 we describe a feasible Monte Carlo-based particle filter approximation which exemplifies the uniform convergence result. Finally, in Section 6, the uniform convergence results are illustrated by some numerical experiments.

2. Exact and truncated filter

In this section we describe in detail the discrete-time filter π and its truncated filter approximation π^{Δ}. Let us start by introducing some general notation and terminology.

For an arbitrary measurable space $(S, \mathcal{S})$, let $\mathcal{P}(\mathcal{S})$ denote the set of all probability measures on the σ-field $\mathcal{S}$. The set of bounded and measurable mappings $\varphi \colon S \to \mathbb{R}$ is denoted by $\mathcal{B}_{\mathrm{b}}(\mathcal{S})$. For all $\varphi \in \mathcal{B}_{\mathrm{b}}(\mathcal{S})$ and $\mu \in \mathcal{P}(\mathcal{S})$, we let $\mu(\varphi)$ denote the Lebesgue integral of φ with respect to the measure μ. If, in particular, $\varphi = \mathbf{1}_A$, where $\mathbf{1}_A$ denotes the indicator function of a set $A \in \mathcal{S}$, then we use the shorthand notation $\mu(A) := \mu(\mathbf{1}_A)$.

Let $(S_1, \mathcal{S}_1)$ and $(S_2, \mathcal{S}_2)$ be arbitrary measurable spaces. A mapping $K \colon S_1 \times \mathcal{S}_2 \to [0, 1]$ is called a *transition probability* if it satisfies

(a) for all $x \in S_1$, $K(x, \cdot) \in \mathcal{P}(\mathcal{S}_2)$;

(b) for all $A \in \mathcal{S}_2$, $K(\cdot, A)$ is measurable.

A transition probability $K \colon S \times \mathcal{S} \to [0, 1]$ can be used for defining a mapping $\varphi \in \mathcal{B}_{\mathrm{b}}(\mathcal{S}) \mapsto K(\varphi) \in \mathcal{B}_{\mathrm{b}}(\mathcal{S})$ such that

$$K(\varphi)(x) := \int \varphi(y) K(x, \, \mathrm{d}y)$$

and a mapping $\mu \in \mathcal{P}(\mathcal{S}) \mapsto \mu K \in \mathcal{P}(\mathcal{S})$ such that

$$\mu K(\varphi) := \int \left(\int \varphi(y) K(x, \, \mathrm{d}y) \right) \mu(\mathrm{d}x) \tag{2.1}$$

for all $\varphi \in \mathcal{B}_{\mathrm{b}}(\mathcal{S})$. The *Dobrushin ergodic coefficient* [6] (see also [5]) of a transition probability $K \colon S \times \mathcal{S} \to [0, 1]$ is defined as

$$\alpha_S(K) := 1 - \sup_{\substack{x, y \in S \\ A \in \mathcal{S}}} |K(x, A) - K(y, A)|.$$

For all finite measures μ on $\mathcal{S}$ and all μ-integrable functions $\psi \colon S \to \mathbb{R}_+$ such that $\mu\psi > 0$, we define a mapping $\mu \mapsto \psi \cdot \mu \in \mathcal{P}(\mathcal{S})$ such that

$$(\psi \cdot \mu)(\varphi) = \frac{\mu(\psi\varphi)}{\mu(\psi)} \quad \text{for all } \varphi \in \mathcal{B}_{\mathrm{b}}(\mathcal{S}).$$

Let $K_i \colon \mathbb{R}^{d_\mathrm{s}} \times \mathcal{B}(\mathbb{R}^{d_\mathrm{s}}) \to [0, 1]$ denote the transition probabilities of the signal X, i.e. for all $A \in \mathcal{B}(\mathbb{R}^{d_\mathrm{s}})$,

$$K_i(X_{i-1}, A) = \mathrm{P}(X_i \in A \mid X_{i-1}).$$

For all $x_{i-1}, x_i \in \mathbb{R}^{d_\mathrm{s}}$, we define

$$k_i(x_{i-1}, x_i) := \rho_{W_i}(x_i - f_i(x_{i-1})),$$

which, according to (1.1), is the density of $K_i(x_{i-1}, \cdot) \in \mathcal{P}(\mathcal{B}(\mathbb{R}^{d_s}))$ with respect to λ_{d_s} for all $x_{i-1} \in \mathbb{R}^{d_s}$. Let us also define

$$g_{i,y}(x) := \rho_{V_i}(y - h_i(x)) \quad \text{and} \quad g_i(x) := g_{i, Y_i}(x).$$

The function g_i is often referred to as the *likelihood function* in the literature. In this case, it can be shown that the filter process π is given by the recursion

$$\pi_i := g_i \cdot \pi_{i-1} K_i,$$

where $\pi_0 = P_0$. For all $i > 0$, we let Q_i denote the mapping $\pi_{i-1} \mapsto \pi_i$.

The truncated approximation $\pi^\Delta := (\pi_i^\Delta)_{i \geq 0}$ of the exact filter π is defined by the recursion

$$\pi_i^\Delta := g_i^\Delta \cdot \pi_{i-1}^\Delta K_i,$$

where $\pi_0^\Delta = \pi_0$, $g_i^\Delta(x) := g_i(x) \mathbf{1}_{D_i(\Delta)}(x, Y_i)$, and

$$D_i(\Delta) := \{(x, y) \in \mathbb{R}^{d_s} \times \mathbb{R}^{d_m} \mid \|y - \tilde{h}_i(x)\| \leq \Delta\}.$$

For all $i > 0$, we let Q_i^Δ denote the mapping $\pi_{i-1}^\Delta \mapsto \pi_i^\Delta$. Moreover, we define, for all $i \geq j > 0$,

$$Q_{j,i}^\Delta := Q_i^\Delta \circ \cdots \circ Q_j^\Delta$$

and, if $i < j$, the mapping $Q_{j,i}^\Delta$ is defined to be the identity. Then we can define, for all $i \geq j \geq 0$,

$$\pi_{i,j}^\Delta := Q_{j+1,i}^\Delta(\pi_j). \tag{2.2}$$

Accordingly, $\pi_{i,0}^\Delta = \pi_i^\Delta$ and $\pi_{i,i}^\Delta = \pi_i$.

According to the definitions above, the filtering distribution π_i can be represented as a composition of nonlinear operations $Q_1, Q_2, \ldots, Q_i$ applied to the initial distribution π_0. It has been pointed out in [5] and [13] that π_i admits another representation which is especially useful in the analysis of the filter stability and the uniform convergence of filter approximations. In this alternative representation, given in Lemma 2.2(i), below, π_i is obtained by applying only one nonlinear operation and a composition of linear operations of the form (2.1) to the initial distribution. The usefulness of this representation is due to the fact that these linear mappings are contractions in the metric induced by the total variation norm, which we denote by $\| \cdot \|_{\text{TV}}$. The contraction coefficient is determined by the Dobrushin ergodic coefficient $\alpha_S(K)$ and, because of the contractivity, the errors cannot increase over these linear operations. In order to specify this alternative representation of π_i, we define, for all $\Delta > 0$ and $i \geq j > 0$, the mapping $S_{j,i}^\Delta \colon \mathbb{R}^{d_s} \times \mathcal{B}(\mathbb{R}^{d_s}) \to [0, 1]$ as

$$S_{j,i}^\Delta(x, A) := \frac{K_j(g_j^\Delta K_{j+1}(g_{j+1}^\Delta \cdots K_i(g_i^\Delta)) \mathbf{1}_A)(x)}{K_j(g_j^\Delta K_{j+1}(g_{j+1}^\Delta \cdots K_i(g_i^\Delta)))(x)}$$

and the mapping $\psi_{j,i}^\Delta \colon \mathbb{R}^{d_s} \to \mathbb{R}$ as

$$\psi_{j,i}^\Delta(x) := \begin{cases} K_j(g_j^\Delta K_{j+1}(g_{j+1}^\Delta \cdots K_i(g_i^\Delta)))(x), & i \geq j, \\ 1, & i < j. \end{cases}$$

Because of the truncation, it suffices to consider the Dobrushin ergodic coefficients of the kernels $S_{j,i}^{\Delta}$ that are restricted to the truncated sets, i.e. $\alpha_{C_{j-1}(\Delta)}(S_{j,i}^{\Delta})$, where

$$C_i(\Delta) := \begin{cases} \{x \in \mathbb{R}^{d_s} \mid \|Y_i - \tilde{h}_i(x)\| \leq \Delta\}, & i > 0, \\ \{x \in \mathbb{R}^{d_s} \mid \|x\| \leq \beta\Delta + \beta\gamma\}, & i = 0, \end{cases} \tag{2.3}$$

for all $i, \Delta > 0$. Under assumptions (A1)–(A4), the generally intractable coefficient $\alpha_{C_{j-1}(S_{j,i}^{\Delta})}$ admits an explicit lower bound, given in Lemma 2.1, below. For this purpose, we define, for $i, \Delta > 0$, $\varepsilon_i(\Delta) := \varepsilon(\Delta, \xi_i)$, where $\varepsilon \colon \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ is defined as

$$\varepsilon(x, y) := \frac{m_1}{M_1} \exp(-a_1((\alpha\beta + \beta)(x + \gamma) + \delta + y)^{b_1})$$

and

$$\xi_i := \begin{cases} \beta\|V_i\| + \|W_i\| + \alpha\beta\|V_{i-1}\|, & i > 1, \\ \beta\|V_i\| + \|W_i\| + \alpha\|X_0\|, & i = 1. \end{cases} \tag{2.4}$$

Lemma 2.1. *For all $i \geq j > 0$ and $\Delta > 0$, we have $\alpha_{C_{j-1}(\Delta)}(S_{j,i}^{\Delta}) \geq \varepsilon_j(\Delta)$.*

Proof. According to (2.3) and (A2), for all $i > 0$ and $x_i \in C_i(\Delta)$,

$$\|x_i - X_i\| \leq \beta\|\tilde{h}_i(x_i) - Y_i + \bar{h}_i(X_i) + V_i\| \leq \beta\Delta + \beta\gamma + \beta\|V_i\|.$$

If $i = 0$ then, for all $x_0 \in C_0(\Delta)$,

$$\|x_0 - X_0\| \leq \|x_0\| + \|X_0\| \leq \beta\Delta + \beta\gamma + \|X_0\|.$$

Therefore, according to (A1) and (2.4), for all $i > 0$,

$$\begin{aligned} \|x_i - f_i(x_{i-1})\| &\leq \|x_i - X_i\| + \|f_i(X_{i-1}) + W_i - f_i(x_{i-1})\| \\ &\leq \|x_i - X_i\| + \alpha\|x_{i-1} - X_{i-1}\| + \delta + \|W_i\| \\ &\leq (\alpha\beta + \beta)(\Delta + \gamma) + \delta + \xi_i. \end{aligned}$$

The claim then follows by substituting the last form into

$$\begin{aligned} \alpha_{C_{j-1}(\Delta)}(S_{j,i}^{\Delta}) &\geq \|k_i\|_{\infty}^{-1} \inf_{\substack{x_i \in C_i(\Delta) \\ x_{i-1} \in C_{i-1}(\Delta)}} k_i(x_{i-1}, x_i) \\ &\geq \frac{m_1}{M_1} \inf_{\substack{x_i \in C_i(\Delta) \\ x_{i-1} \in C_{i-1}(\Delta)}} \exp(-a_1\|x_i - f_i(x_{i-1})\|^{b_1}), \end{aligned}$$

where the first inequality can be shown to hold (see, e.g. [3] and [13]). $\quad\square$

The following lemma summarises some properties that are used throughout the remainder of this work. Excluding minor differences, the proofs can be found in [5] and [13] (see also [3]).

Lemma 2.2. *For all $i \geq j > 0$, $\ell > 0$, $\Delta > 0$, and $\pi_0 \in \mathcal{P}(\mathcal{B}(\mathbb{R}^{d_s}))$,*

(i) $Q_{j,i}^{\Delta}(\pi_0) = (\psi_{j,i}^{\Delta} \cdot \pi_0) S_{j,i}^{\Delta} S_{j+1,i}^{\Delta} \cdots S_{i,i}^{\Delta}$;

(ii)

$$\|\pi_j - \pi_j^{\Delta}\|_{\text{TV}} \leq \sum_{m=1}^{j} \prod_{n=m+2}^{j} (1 - \varepsilon_n(\Delta)) \min\left(1, \frac{\|\pi_{m,m} - \pi_{m,m-1}\|_{\text{TV}}}{\varepsilon_{m+1}(\Delta)}\right);$$

(iii) *there exists $r > 0$ such that*

$$E\left[\prod_{n=1}^{i}(1 - \varepsilon_{\ell+n}(\Delta)) \,\bigg|\, \mathcal{F}_\ell\right] \leq (1 - \tilde{\varepsilon}(\Delta))^{i-1},$$

where $\mathcal{F}_\ell := \sigma(X_0, W_1, \ldots, W_\ell, V_1, \ldots, V_\ell)$ and $\tilde{\varepsilon}(\Delta) := \varepsilon(\Delta, r)/2$.

3. Uniformly convergent approximation by truncation

In this section we prove sufficient conditions for the uniform convergence of the truncated filter π^Δ to the exact filter π, i.e.

$$\lim_{\Delta \to \infty} \sup_{i>0} E[\|\pi_i - \pi_i^\Delta\|_{\mathrm{TV}}] = 0.$$

We start by establishing bounds for the tails of the distribution of the random variable ξ_i. At this point, the analysis differs from the approach used in [3] (see also [7]), where the stochastic process $(\xi_i)_{i>0}$ was bounded by showing that, for all $\epsilon > 0$, there exist positive random variables c_1 and c_2 such that, for all $i > 0$, we have, P-almost surely (P-a.s.),

$$\xi_i \leq (c_1 + (a_1 - \epsilon)^{-1} \ln i)^{1/B_1} + (\alpha\beta + \beta)(c_2 + (a_2 - \epsilon)^{-1} \ln i)^{1/B_2}.$$

This almost-sure upper bound for the stochastic process $(\xi_i)_{i>0}$ is obviously increasing in i and, therefore, it does not enable uniform bounds for the random variables $\varepsilon_i(\Delta)$. Here, instead of bounding the random variables $(\xi_i)_{i>0}$ directly, we use assumptions (A1)–(A4) to derive a uniform, i.e. independent of i, upper bound for the densities of the random variables $(\xi_i)_{i>0}$. This further enables the derivation of uniform bounds for integrals that are involved with the random variables $\varepsilon_i(\Delta)$.

Proposition 3.1. *For all $\epsilon > 0$, there exists $c_3(\epsilon) > 0$ such that, for all $x > 0$ and $i > 0$,*

$$\rho_{\xi_i}(x) \leq c_3(\epsilon) \exp((-A_3 + \epsilon)x^{B_1}),$$

where

$$A_3 := \begin{cases} A_1 A_2(A_1^{1/B_1}(\alpha\beta + \beta) + A_2^{1/B_1})^{-B_1} & \text{if } B_1 = B_2, \\ A_1 & \text{if } B_1 < B_2. \end{cases}$$

Proof. The proof follows from Lemma A.1 in Appendix A.

By defining a constant $A_2^* = A_2^*(\alpha, a_1, A_1, A_2, \beta, b_1, B_1)$ as

$$A_2^* := \left(2 + \left(1 + \frac{a_1^{1/b_1}}{A_3^{1/b_1}}\right)^{b_1}\right) a_1(\alpha\beta + \beta)^{b_1},$$

we have the following result, which is a refinement of Proposition 3.4 of [13].

Theorem 3.1. *If one of the following two conditions holds:*

(i) $B_2 > b_1 = B_1$;

(ii) $B_2 = b_1 = B_1$ and $A_2 > A_2^*$;

then there exist $c_4, c_5 > 0$ such that

$$\sup_{i>0} \mathrm{E}[\|\pi_i - \pi_i^\Delta\|_{\mathrm{TV}}] \le c_4 \exp(-c_5 \Delta^{B_1}). \tag{3.1}$$

Proof. Let

$$J_i(\Delta) := \min\left(1, \frac{\|\pi_{i,i}^\Delta - \pi_{i,i-1}^\Delta\|_{\mathrm{TV}}}{\varepsilon_{i+1}(\Delta)}\right).$$

Then, according to (ii) and (iii) of Lemma 2.2 and the fact that $J_j(\Delta)$ is $\mathcal{F}_{j+1}$-measurable,

$$\mathrm{E}[\|\pi_i - \pi_i^\Delta\|_{\mathrm{TV}}] \le \sum_{j=1}^{i} \mathrm{E}\left[\mathrm{E}\left[\prod_{n=j+2}^{i}(1 - \varepsilon_n(\Delta)) \,\Big|\, \mathcal{F}_{j+1}\right] J_j(\Delta)\right]$$

$$= \sum_{j=1}^{i} \mathrm{E}\left[\mathrm{E}\left[\prod_{n=1}^{i-j-1}(1 - \varepsilon_{n+j+1}(\Delta)) \,\Big|\, \mathcal{F}_{j+1}\right] J_j(\Delta)\right]$$

$$\le \sum_{j=1}^{i}(1 - \tilde{\varepsilon}(\Delta))^{(i-j-2)^+} \mathrm{E}[J_j(\Delta)], \tag{3.2}$$

where $(\cdot)^+ := \max(0, \cdot)$. In the following we derive an upper bound for $\mathrm{E}[J_j(\Delta)]$ which is independent of j and, therefore, can be brought outside the summation. For the remaining sum, it follows by the convergence of the geometric series that

$$\sum_{j=1}^{i}(1 - \tilde{\varepsilon}(\Delta))^{(i-j-2)^+} \le 2 + \frac{1}{\tilde{\varepsilon}(\Delta)}. \tag{3.3}$$

It can be shown that, for all $\epsilon > 0$, there exists $c_6(\epsilon) > 0$ such that, for all $\Delta \ge \Delta_0 > \gamma$ and $i > 0$,

$$\tau(\Delta) := c_6(\epsilon) \exp((-A_2 + \epsilon)(\Delta - \gamma)^{B_2}) \ge \int_{\|y\|>\Delta-\gamma} \rho_{V_i}(y) \lambda_{d_\mathrm{m}}(\mathrm{d}y).$$

For all $u, v \ge 0$, we have

$$\min(1, uv) \le \min(1, u) + \min(1, v). \tag{3.4}$$

By applying this inequality to $\mathrm{E}[J_j(\Delta)]$ we have, for all $q \in \mathbb{R}$,

$$\mathrm{E}[J_j(\Delta)] \le \mathrm{E}\left[\min\left(1, \frac{\|\pi_{i,i}^\Delta - \pi_{i,i-1}^\Delta\|_{\mathrm{TV}}}{\tau(\Delta)^q}\right)\right] + \mathrm{E}\left[\min\left(1, \frac{\tau(\Delta)^q}{\varepsilon_{i+1}(\Delta)}\right)\right]. \tag{3.5}$$

Because $\|\pi_{i,i}^\Delta - \pi_{i,i-1}^\Delta\|_{\mathrm{TV}} = \pi_i(\overline{C_i(\Delta)})$, for the first expectation in (3.5), we have

$$\mathrm{E}\left[\min\left(1, \frac{\|\pi_{i,i}^\Delta - \pi_{i,i-1}^\Delta\|_{\mathrm{TV}}}{\tau(\Delta)^q}\right)\right] \le \mathrm{E}\left[\frac{\|\pi_{i,i}^\Delta - \pi_{i,i-1}^\Delta\|_{\mathrm{TV}}}{\tau(\Delta)^q}\right] = \frac{\mathrm{E}[\pi_i(\overline{C_i(\Delta)})]}{\tau(\Delta)^q}. \tag{3.6}$$

Moreover,

$$
\begin{aligned}
\mathrm{E}[\pi_i(\overline{C_i(\Delta)})] &= \mathrm{E}\left[\mathrm{E}\left[\frac{\int_{\overline{C_i(\Delta)}} g_i \, \mathrm{d}\pi_{i-1} K_i}{\int g_i \, \mathrm{d}\pi_{i-1} K_i} \,\bigg|\, \mathcal{F}_{i-1}\right]\right] \\
&= \mathrm{E}\left[\int \left(\frac{\int_{\overline{C_i(\Delta)}} g_{i,y}(x_i)\pi_{i-1}K_i(\mathrm{d}x_i)}{\int g_{i,y}(x_i)\pi_{i-1}K_i(\mathrm{d}x_i)} \int g_{i,y}(x_i)\pi_{i-1}K_i(\mathrm{d}x_i)\right)\lambda_{d_{\mathrm{m}}}(\mathrm{d}y)\right] \\
&= \mathrm{E}\left[\int \left(\int g_{i,y}(x_i)\,\mathbf{1}_{\overline{D_i(\Delta)}}(x_i,y)\lambda_{d_{\mathrm{m}}}(\mathrm{d}y)\right)\pi_{i-1}K_i(\mathrm{d}x_i)\right] \\
&= \mathrm{E}\left[\int \left(\int_{\|y\|\geq\Delta-\gamma} \rho_{V_i}(y)\lambda_{d_{\mathrm{m}}}(\mathrm{d}y)\right)\pi_{i-1}K_i(\mathrm{d}x_i)\right] \\
&\leq \tau(\Delta).
\end{aligned}
$$

Substitution into (3.6) yields

$$
\mathrm{E}\left[\min\left(1, \frac{\|\pi_{i,i}^{\Delta} - \pi_{i,i-1}^{\Delta}\|_{\mathrm{TV}}}{\tau(\Delta)^q}\right)\right] \leq c_6(\epsilon)^{1-q}\exp((1-q)(-A_2+\epsilon)(\Delta-\gamma)^{B_2}). \quad (3.7)
$$

Let us then consider the second integral in (3.5). Because, for all $\Delta > 0$ and $q > 0$, $\tau(\Delta)^q/\varepsilon(\Delta,\xi)$ is a strictly increasing function of ξ, we have, according to Proposition 3.1, for all $\theta > 0$,

$$
\begin{aligned}
\mathrm{E}\left[\min\left(1, \frac{\tau(\Delta)^q}{\varepsilon_{i+1}(\Delta)}\right)\right] &\leq c_3(\epsilon)\int_0^{\theta\Delta} \frac{\tau(\Delta)^q}{\varepsilon(\Delta,\xi)}\exp((-A_3+\epsilon)\xi^{B_1})\,\mathrm{d}\xi \\
&\quad + c_3(\epsilon)\int_{\theta\Delta}^{\infty} \exp((-A_3+\epsilon)\xi^{B_1})\,\mathrm{d}\xi \\
&\leq \frac{c_7(\epsilon)\tau(\Delta)^q}{\varepsilon(\Delta,\theta\Delta)} + c_8(\epsilon,\theta)\exp((-A_3+2\epsilon)(\theta\Delta)^{B_1}). \quad (3.8)
\end{aligned}
$$

By combining (3.2), (3.3), (3.5), (3.7), and (3.8), we have

$$
\begin{aligned}
\mathrm{E}[\|\pi_i - \pi_i^{\Delta}\|_{\mathrm{TV}}] \leq \left(2 + \frac{1}{\tilde{\varepsilon}(\Delta)}\right)\Bigg(&c_6(\epsilon)^{1-q}\exp((1-q)(-A_2+\epsilon)(\Delta-\gamma)^{B_2}) \\
&+ \frac{c_7(\epsilon)\tau(\Delta)^q}{\varepsilon(\Delta,\theta\Delta)} + c_8(\epsilon,\theta)\exp((-A_3+2\epsilon)(\theta\Delta)^{B_1})\Bigg).
\end{aligned}
$$

By using the definitions of $\tilde{\varepsilon}$, ε, and τ, it follows that (3.1) holds if (i) is satisfied. In case (ii), (3.1) holds if (q,θ) is a solution of

$$
\begin{aligned}
(1-q)A_2 &> a_1(\alpha\beta+\beta)^{b_1}, \\
qA_2 &> a_1(\alpha\beta+\beta+\theta)^{b_1} + a_1(\alpha\beta+\beta)^{b_1}, \\
A_3\theta^{B_1} &> a_1(\alpha\beta+\beta)^{b_1}.
\end{aligned}
$$

We can check that a solution to this system of inequalities exists if $A_2 > A_2^*$.

4. Uniform convergence of a particle filter approximation

Let us define $\mathcal{Y} := \bigcup_{i\geq 0} \mathcal{Y}_i$ and $\mathcal{H} := (\mathcal{H}_i)_{i\geq 0}$, such that $\mathcal{H}$ is a nondecreasing sequence of sub-σ-fields $\mathcal{H}_i \subset \mathcal{F}$ satisfying $\mathcal{H}_0 = \mathcal{Y}$. Also, for all $\Delta > 0$ and $i > 0$, we let K_i^{Δ} denote the restriction of K_i to the set $C_i(\Delta)$, that is, for all $x \in \mathbb{R}^{d_{\mathrm{s}}}$ and $A \in \mathcal{B}(\mathbb{R}^{d_{\mathrm{s}}})$, we have
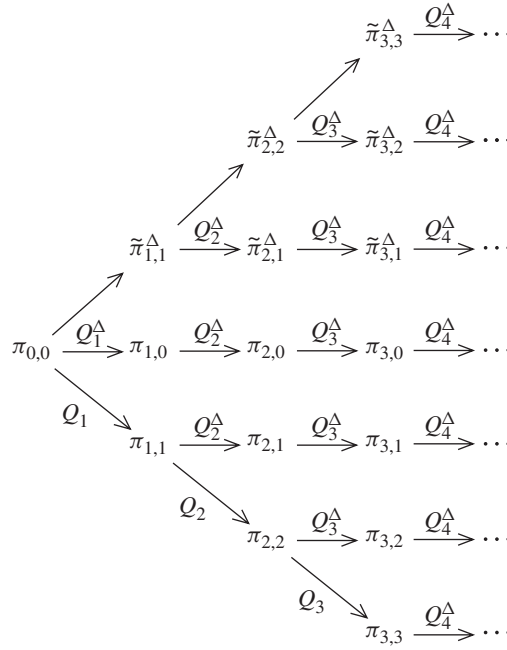
FIGURE 1: The exact filter, its truncated approximation, and the particle approximation.

$K_i^\Delta(x, A) = K_i(x, C_i(\Delta) \cap A)$. The class of particle filters under consideration is then defined as the set of all $\mathcal{P}(\mathcal{B}(\mathbb{R}^{d_s}))$-valued $\mathcal{H}$-adapted stochastic processes $\tilde{\pi}^\Delta = (\tilde{\pi}_i^\Delta)_{i \geq 0}$ such that $\tilde{\pi}_0^\Delta = \pi_0$ and, for all $i > 0$,

$$\tilde{\pi}_i^\Delta = g_i^\Delta \cdot v_i^\Delta, \tag{4.1}$$

where v_i^Δ is a $\mathcal{H}_i$-measurable random probability measure taking values in $\mathcal{P}(\mathcal{B}(\mathbb{R}^{d_s}))$ such that

(A5) $0 < v_i^\Delta(g_i^\Delta) < \infty$, P-a.s.;

(A6) there exist $a_4, M_4 > 0$ such that, for all $\Delta \geq \Delta_0$,

$$\sup_{i > 0} \sup_{\|\varphi\|_\infty \leq 1} \mathrm{E}[|\tilde{\pi}_{i-1}^\Delta K_i^\Delta(\varphi) - v_i^\Delta(\varphi)| \mid \mathcal{H}_{i-1}] \leq M_4 \exp(-a_4 \Delta^{b_2}).$$

Analogously to the definition of $\pi_{i,j}^\Delta$ in (2.2), we define $\tilde{\pi}_{i,j}^\Delta := Q_{i,j+1}^\Delta(\tilde{\pi}_j^\Delta)$. Accordingly, $\tilde{\pi}_{i,i}^\Delta = \tilde{\pi}_i^\Delta$. See Figure 1 for an illustration. In the proof of the uniform convergence theorem, we need the following lemma.

Lemma 4.1. *Suppose that μ and v are random probability measures taking values in $\mathcal{P}(\mathcal{S})$ and that the random mappings $\psi, \theta \colon S \to \mathbb{R}_+$ satisfy $\mu\psi, v\theta > 0$, P-a.s. Then*

(i)
$$\frac{1}{2} \sup_{\|\varphi\|_\infty \leq 1} \mathrm{E}[|\mu K(\varphi) - v K(\varphi)|] \leq (1 - \alpha_S(K)) \sup_{\|\varphi\|_\infty \leq 1} \mathrm{E}[|\mu\varphi - v\varphi|];$$

(ii)
$$\frac{1}{2} \sup_{\|\varphi\|_\infty \leq 1} \mathrm{E}[|(\psi \cdot \mu)(\varphi) - (\theta \cdot v)(\varphi)|] \leq \sup_{\|\varphi\|_\infty \leq 1} \mathrm{E}\left[\frac{|\mu(\psi\varphi) - v(\theta\varphi)|}{\mu\psi}\right].$$

Proof. To prove (i), it is observed that, for all $x \in S$,

$$\sup_{\|\varphi\|_\infty \leq 1} \mathrm{E}[|\mu K(\varphi) - \nu K(\varphi)|] = \sup_{\|\varphi\|_\infty \leq 1} \mathrm{E}[|\mu(K\varphi - K\varphi(x)) - \nu(K\varphi - K\varphi(x))|]$$

$$\leq \sup_{\|\varphi\|_\infty \leq 1} \|K\varphi - K\varphi(x)\|_\infty \sup_{\|\psi\|_\infty \leq 1} \mathrm{E}[|\mu\psi - \nu\psi|]. \quad (4.2)$$

Moreover,

$$\sup_{\|\varphi\|_\infty \leq 1} \|K\varphi - K\varphi(x)\|_\infty = \sup_{\substack{\|\varphi\|_\infty \leq 1 \\ z \in S}} |K\varphi(z) - K\varphi(x)|$$

$$\leq \sup_{\substack{\|\varphi\|_\infty \leq 1 \\ x, z \in S}} |K\varphi(z) - K\varphi(x)|$$

$$= 2 \sup_{\substack{A \in \mathscr{S} \\ x, z \in S}} |K(z, A) - K(x, A)|$$

$$= 2(1 - \alpha_S(K)). \quad (4.3)$$

By combining (4.2) and (4.3), we have (i). To prove (ii), it is observed that

$$|(\psi \cdot \mu)(\varphi) - (\theta \cdot \nu)(\varphi)| \leq \frac{|\mu(\psi\varphi) - \nu(\theta\varphi)|}{\mu\psi} + \|\varphi\|_\infty \frac{|\mu\psi - \nu\theta|}{\mu\psi}. \quad (4.4)$$

Because

$$\mathrm{E}\left[\frac{|\mu\psi - \nu\theta|}{\mu\psi}\right] \leq \sup_{\|\varphi\|_\infty \leq 1} \mathrm{E}\left[\frac{|\mu(\psi\varphi) - \nu(\theta\varphi)|}{\mu\psi}\right],$$

the claim follows by taking expectations in (4.4) and using the fact that $\|\varphi\|_\infty \leq 1$.

By defining a constant $a_4^* = a_4^*(\alpha, a_1, A_1, a_2, A_2, \beta, b_1, B_1)$ as

$$a_4^* := 2\left(1 + \left(1 + \frac{a_1^{1/b_1}}{A_3^{1/b_1}}\right)^{b_1}\right)a_1(\alpha\beta + \beta)^{b_1} + a_2,$$

we have the following result.

Theorem 4.1. *If one of the following conditions holds:*

(i) $b_1 = B_1 < B_2 \leq b_2$ *and* $a_2 < a_4$;

(ii) $b_1 = B_1 = B_2 < b_2$, $a_2 < a_4$, *and* $A_2^* < A_2$;

(iii) $b_1 = B_1 = B_2 = b_2$, $a_4^* < a_4$, *and* $A_2^* < A_2$;

then there exist $c_9, c_{10} > 0$ *such that*

$$\sup_{i > 0} \mathrm{E}\left[\sup_{\|\varphi\|_\infty \leq 1} \mathrm{E}[|\pi_i\varphi - \tilde{\pi}_i^\Delta\varphi| \mid \mathcal{Y}_i]\right] \leq c_9 \exp(-c_{10}\Delta^{B_1}).$$

Proof. By the triangle inequality (see Figure 1),

$$
\mathrm{E}\left[\sup_{\|\varphi\|_\infty \leq 1} \mathrm{E}[|\pi_i\varphi - \tilde{\pi}_i^\Delta \varphi| \mid \mathcal{Y}_i]\right]
$$

$$
\leq \mathrm{E}\left[\sup_{\|\varphi\|_\infty \leq 1} \mathrm{E}[|\pi_i\varphi - \pi_i^\Delta \varphi| \mid \mathcal{Y}_i]\right] + \sum_{j=1}^{i} \mathrm{E}\left[\sup_{\|\varphi\|_\infty \leq 1} \mathrm{E}[|\tilde{\pi}_{i,j}^\Delta \varphi - \tilde{\pi}_{i,j-1}^\Delta \varphi| \mid \mathcal{Y}_i]\right]
$$

$$
\leq \mathrm{E}\left[\sup_{\|\varphi\|_\infty \leq 1} |\pi_i\varphi - \pi_i^\Delta \varphi|\right] + \sum_{j=1}^{i} \mathrm{E}\left[\sup_{\|\varphi\|_\infty \leq 1} \mathrm{E}[\mathrm{E}[|\tilde{\pi}_{i,j}^\Delta \varphi - \tilde{\pi}_{i,j-1}^\Delta \varphi| \mid \mathcal{H}_{j-1}] \mid \mathcal{Y}_i]\right]
$$

$$
\leq 2\,\mathrm{E}[\|\pi_i - \pi_i^\Delta\|_{\mathrm{TV}}] + \sum_{j=1}^{i} \mathrm{E}\left[\sup_{\|\varphi\|_\infty \leq 1} \mathrm{E}[|\tilde{\pi}_{i,j}^\Delta \varphi - \tilde{\pi}_{i,j-1}^\Delta \varphi| \mid \mathcal{H}_{j-1}]\right]. \tag{4.5}
$$

Because $\mathrm{E}[\|\pi_i - \pi_i^\Delta\|_{\mathrm{TV}}]$, according to Theorem 3.1, is uniformly convergent to 0 as $\Delta \to \infty$, we only have to consider the sum in (4.5). It is first observed that, according to Lemma 2.2(i),

$$
\tilde{\pi}_{i,j-1}^\Delta = Q_{j+1,i}^\Delta(\tilde{\pi}_{j,j-1}^\Delta) = (\psi_{j+1,i}^\Delta \cdot \tilde{\pi}_{j,j-1}^\Delta) S_{j+1,i}^\Delta \cdots S_{i,i}^\Delta,
$$
$$
\tilde{\pi}_{i,j}^\Delta = Q_{j+1,i}^\Delta(\tilde{\pi}_{j,j}^\Delta) = (\psi_{j+1,i}^\Delta \cdot \tilde{\pi}_{j,j}^\Delta) S_{j+1,i}^\Delta \cdots S_{i,i}^\Delta,
$$

and, therefore, by Lemma 4.1(i) we have

$$
\sup_{\|\varphi\|_\infty \leq 1} \mathrm{E}[|\tilde{\pi}_{i,j}^\Delta \varphi - \tilde{\pi}_{i,j-1}^\Delta \varphi| \mid \mathcal{H}_{j-1}] \leq 4(1 - \alpha_{C_j(\Delta)}(S_{j+1,i}^\Delta \cdots S_{i,i}^\Delta)) J_{j,i}(\Delta), \tag{4.6}
$$

where

$$
J_{j,i}(\Delta) := \tfrac{1}{2} \sup_{\|\varphi\|_\infty \leq 1} \mathrm{E}[|(\psi_{j+1,i}^\Delta \cdot \tilde{\pi}_{j,j}^\Delta)(\varphi) - (\psi_{j+1,i}^\Delta \cdot \tilde{\pi}_{j,j-1}^\Delta)(\varphi)| \mid \mathcal{H}_{j-1}].
$$

It is then observed that

$$
\psi_{j+1,i}^\Delta \cdot \tilde{\pi}_{j,j}^\Delta = \psi_{j+1,i}^\Delta \cdot (g_j^\Delta \cdot \nu_j^\Delta) = g_j^\Delta \psi_{j+1,i}^\Delta \cdot \nu_j^\Delta,
$$
$$
\psi_{j+1,i}^\Delta \cdot \tilde{\pi}_{j,j-1}^\Delta = \psi_{j+1,i}^\Delta \cdot (g_j^\Delta \cdot \tilde{\pi}_{j-1}^\Delta K_j) = g_j^\Delta \psi_{j+1,i}^\Delta \cdot \tilde{\pi}_{j-1}^\Delta K_j.
$$

Thus, by Lemma 4.1(ii) and the fact that $\tilde{\pi}_{j-1}^\Delta K_j(g_j^\Delta \psi_{j+1,i}^\Delta)$ is $\mathcal{H}_{j-1}$-measurable, we have

$$
J_{j,i}(\Delta) = \tfrac{1}{2} \sup_{\|\varphi\|_\infty \leq 1} \mathrm{E}[|(g_j^\Delta \psi_{j+1,i}^\Delta \cdot \nu_j^\Delta)(\varphi) - (g_j^\Delta \psi_{j+1,i}^\Delta \cdot \tilde{\pi}_{j-1}^\Delta K_j)(\varphi)| \mid \mathcal{H}_{j-1}]
$$

$$
\leq \frac{\sup_{\|\varphi\|_\infty \leq 1} \mathrm{E}[|\nu_j^\Delta(g_j^\Delta \psi_{j+1,i}^\Delta \varphi) - \tilde{\pi}_{j-1}^\Delta K_j(g_j^\Delta \psi_{j+1,i}^\Delta \varphi)| \mid \mathcal{H}_{j-1}]}{\tilde{\pi}_{j-1}^\Delta K_j(g_j^\Delta \psi_{j+1,i}^\Delta)}
$$

$$
\leq \frac{\|g_j^\Delta \psi_{j+1,i}^\Delta\|_\infty}{\tilde{\pi}_{j-1}^\Delta K_j(g_j^\Delta \psi_{j+1,i}^\Delta)} M_4 \exp(-a_4 \Delta^{b_2}), \tag{4.7}
$$

where the last inequality follows from assumption (A6). Moreover,

$$
\frac{\|g_j^\Delta \psi_{j+1,i}^\Delta\|_\infty}{\tilde{\pi}_{j-1}^\Delta K_j(g_j^\Delta \psi_{j+1,i}^\Delta)} = \frac{\|g_j^\Delta \psi_{j+1,i}^\Delta\|_\infty}{\tilde{\pi}_{j-1}^\Delta K_j(g_j^\Delta)} \frac{\tilde{\pi}_{j-1}^\Delta K_j(g_j^\Delta)}{\tilde{\pi}_{j-1}^\Delta K_j(g_j^\Delta \psi_{j+1,i}^\Delta)} \leq \frac{\|g_j^\Delta\|_\infty}{\tilde{\pi}_{j-1}^\Delta K_j(g_j^\Delta)} \frac{\|\psi_{j+1,i}^\Delta\|_\infty}{\tilde{\pi}_{j,j-1}^\Delta(\psi_{j+1,i}^\Delta)}.
$$
$$\tag{4.8}$$

It is observed that, for all $j > 0$,

$$
\frac{\psi_{j+1,i}^{\Delta}(x)}{\tilde{\pi}_{j,j-1}^{\Delta}(\psi_{j+1,i}^{\Delta})} = \frac{K_{j+1}(g_{j+1}^{\Delta}\psi_{j+2,i}^{\Delta})(x)}{\tilde{\pi}_{j,j-1}^{\Delta}(\mathbf{1}_{C_j(\Delta)}K_{j+1}(g_{j+1}^{\Delta}\psi_{j+2,i}^{\Delta}))}
$$

$$
\leq \frac{\|k_{j+1}\|_{\infty}}{\inf_{x\in C_j(\Delta),\, y\in C_{j+1}(\Delta)} k_{j+1}(x,y)} \frac{\lambda_{d_s}(g_{j+1}^{\Delta}\psi_{j+2,i}^{\Delta})}{\tilde{\pi}_{j,j-1}^{\Delta}(\lambda_{d_s}(g_{j+1}^{\Delta}\psi_{j+2,i}^{\Delta}))}
$$

$$
\leq \frac{1}{\varepsilon_{j+1}(\Delta)}, \tag{4.9}
$$

where the last inequality follows from Lemma 2.1. In order to bound the first product term in (4.8), we write

$$
\tilde{\pi}_{j-1}^{\Delta}K_j(g_j^{\Delta}) \geq \int_{C_{j-1}(\Delta)}\left(\int_{C_j(\Delta)} g_j(x_j)k_j(x_{j-1},x_j)\lambda_{d_s}(\mathrm{d}x_j)\right)\tilde{\pi}_{j-1}^{\Delta}(\mathrm{d}x_{j-1})
$$

$$
\geq \lambda_{d_s}(C_j(\Delta))\tilde{\pi}_{j-1}^{\Delta}(C_{j-1}(\Delta)) \inf_{x\in C_j(\Delta)} g_j(x) \inf_{\substack{y\in C_j(\Delta)\\ x\in C_{j-1}(\Delta)}} k_j(x,y). \tag{4.10}
$$

Let $B_{d_s}(x,r)$ denote a ball with radius $r > 0$ centred at $x \in \mathbb{R}^{d_s}$. Then, for all $x \in B_{d_s}(\tilde{h}_j^{-1}(Y_j), \Delta_0/\beta_0)$, we have

$$
\|Y_j - \tilde{h}_j(x)\| = \|\tilde{h}_j(\tilde{h}_j^{-1}(Y_j)) - \tilde{h}_j(x)\| \leq \beta_0\|\tilde{h}_j^{-1}(Y_j) - x\| \leq \Delta_0.
$$

Therefore, for all $\Delta > \Delta_0$, $B_{d_s}(\tilde{h}_j^{-1}(Y_j), \Delta_0/\beta_0) \subset C_j(\Delta_0) \subset C_j(\Delta)$, which implies that

$$
\lambda_{d_s}(C_j(\Delta)) \geq \lambda_{d_s}(C_j(\Delta_0)) \geq \lambda_{d_s}\left(B_{d_s}\left(\tilde{h}_j^{-1}(Y_j), \frac{\Delta_0}{\beta_0}\right)\right). \tag{4.11}
$$

Moreover, according to assumptions (A2) and (A4),

$$
\inf_{x\in C_j(\Delta)} g_j(x) \geq \inf_{x\in C_j(\Delta)} m_2 \exp(-a_2\|Y_j - h_j(x)\|^{b_2})
$$

$$
\geq \inf_{x\in C_j(\Delta)} m_2 \exp(-a_2(\|Y_j - \tilde{h}_j(x)\| + \|\bar{h}_j(x)\|)^{b_2})
$$

$$
\geq m_2 \exp(-a_2(\Delta + \gamma)^{b_2}). \tag{4.12}
$$

Also, because, for all $j > 0$, $\tilde{\pi}_j^{\Delta}(C_j(\Delta)) = 1$, we have $\tilde{\pi}_{j-1}^{\Delta}(C_{j-1}(\Delta)) \geq \pi_0(C_0(\Delta_0))$. Therefore, by observing that $J_{j,i}(\Delta) \leq 1$, and by combining (4.7), (4.8), (4.9), (4.10), (4.11), and (4.12), we have

$$
J_j(\Delta) := \min\left(1, \frac{\Gamma(\Delta)}{\varepsilon_j(\Delta)\varepsilon_{j+1}(\Delta)}\right) \geq J_{j,i}(\Delta), \tag{4.13}
$$

where

$$
\Gamma(\Delta) := c_{11}\exp(a_2(\Delta + \gamma)^{b_2} - a_4\Delta^{b_2})
$$

and $c_{11} := M_2 M_4/(\lambda_{d_s}(B_{d_s}(\tilde{h}_j^{-1}(Y_j), \Delta_0/\beta_0))m_1 m_2 \pi_0(C_0(\Delta_0)))$. By induction, it can be shown that

$$
(1 - \alpha_{C_j(\Delta)}(S_{j+1,i}^{\Delta}\cdots S_{i,i}^{\Delta})) \leq \prod_{n=j+1}^{i}(1 - \alpha_{C_{n-1}(\Delta)}(S_{n,i}^{\Delta})) \leq \prod_{n=j+2}^{i}(1 - \varepsilon_n(\Delta)),
$$

and, therefore, because $J_j(\Delta)$ is $\mathcal{F}_{j+1}$-measurable, it follows, from (4.6) and (4.13), that

$$
\mathrm{E}\left[\sup_{\|\varphi\|_\infty \le 1} \mathrm{E}[|\tilde{\pi}_{i,j}^\Delta \varphi - \tilde{\pi}_{i,j-1}^\Delta \varphi| \mid \mathcal{H}_{j-1}]\right] \le 4\,\mathrm{E}\left[\mathrm{E}\left[\prod_{n=j+2}^i (1 - \varepsilon_n(\Delta)) \,\middle|\, \mathcal{F}_{j+1}\right] J_j(\Delta)\right]
$$

$$
\le 4(1 - \tilde{\varepsilon}(\Delta))^{(i-j-2)^+}\,\mathrm{E}[J_j(\Delta)]. \qquad (4.14)
$$

Following a similar principle as in the proof of Theorem 3.1, we derive an upper bound for $\mathrm{E}[J_j(\Delta)]$ which is independent of j and, therefore, can be brought outside the summation. For the remaining sum, (3.3) holds. According to (3.4),

$$
\mathrm{E}[J_j(\Delta)] \le \mathrm{E}\left[\min\left(1, \frac{\sqrt{\Gamma(\Delta)}}{\varepsilon_j(\Delta)}\right)\right] + \mathrm{E}\left[\min\left(1, \frac{\sqrt{\Gamma(\Delta)}}{\varepsilon_{j+1}(\Delta)}\right)\right]. \qquad (4.15)
$$

Because of Proposition 3.1, it suffices to consider only one of the expectations on the right-hand side. Similarly as in (3.8), we have, for all $\theta, \epsilon > 0$,

$$
\mathrm{E}\left[\min\left(1, \frac{\sqrt{\Gamma(\Delta)}}{\varepsilon_j(\Delta)}\right)\right] \le \frac{c_{12}(\epsilon)\sqrt{\Gamma(\Delta)}}{\varepsilon(\Delta, \theta\Delta)} + c_8(\epsilon, \theta) \exp((-A_3 + 2\epsilon)(\theta\Delta)^{B_1}). \qquad (4.16)
$$

By combining (3.3), (4.14), (4.15), and (4.16), we can check, by using the definitions of Γ and ε, that if (i) or (ii) is satisfied then, for all $\theta > 0$, there exist $c_{13}, c_{14} > 0$ such that

$$
\sum_{j=1}^i \mathrm{E}\left[\sup_{\|\varphi\|_\infty \le 1} \mathrm{E}[|\tilde{\pi}_{i,j}^\Delta \varphi - \tilde{\pi}_{i,j-1}^\Delta \varphi| \mid \mathcal{H}_{j-1}]\right] \le c_{13} \exp(-c_{14}\Delta^{B_1}).
$$

In case (iii), θ must satisfy

$$
\frac{a_4}{2} > \frac{a_2}{2} + a_1(\alpha\beta + \beta + \theta)^{b_1} + a_1(\alpha\beta + \beta)^{b_1},
$$

$$
A_3\theta^{B_1} > a_1(\alpha\beta + \beta)^{b_1}.
$$

A solution to this system of inequalities exists if $a_4 > a_4^*$. The claim then follows by applying Theorem 3.1 to the first term on the right-hand side of (4.5).

For estimation purposes, we are often interested in the mean of the posterior distribution. The following corollary of Theorem 4.1 establishes the uniform convergence of the approximate posterior mean to the exact posterior mean in the sense of Euclidean distance. The expected Euclidean distance of posterior means will also be used as a measure of error in the simulations described in Section 6. For this purpose, we let $\hat{x}_i$ and $\hat{x}_i^\Delta$ denote the means of π_i and $\tilde{\pi}_i^\Delta$, respectively.

Corollary 4.1. *For all $\epsilon > 0$ there exist $c_{15}, c_{16} > 0$ such that*

$$
\sup_{i > 0} \mathrm{E}[\|\hat{x}_i - \hat{x}_i^\Delta\|] \le c_{15} \exp(-c_{16}\Delta^{B_1}).
$$

Proof. Let $I : \mathbb{R}^{d_s} \to \mathbb{R}^{d_s}$ denote the identity mapping, let $I_j : \mathbb{R}^{d_s} \to \mathbb{R}$ denote the projection to the jth axis, let $\tilde{x} := \tilde{h}_i^{-1}(Y_i)$, and let $\tilde{x}_j := I_j(\tilde{x})$. In this case,

$$
\mathrm{E}[\|\hat{x}_i - \hat{x}_i^\Delta\|] \le \sum_{j=1}^{d_s} \mathrm{E}[|\pi_i(I_j) - \tilde{\pi}_i^\Delta(I_j)|],
$$

where

$$\mathrm{E}[|\pi_i(I_j) - \tilde{\pi}_i^\Delta(I_j)|] \leq \mathrm{E}[\mathrm{E}[|\pi_i(\mathbf{1}_{C_i(\Delta)}(I_j - \tilde{x}_j)) - \tilde{\pi}_i^\Delta(\mathbf{1}_{C_i(\Delta)}(I_j - \tilde{x}_j))| \mid \mathcal{Y}_i]]$$
$$+ \mathrm{E}[|\pi_i(\mathbf{1}_{\overline{C_i(\Delta)}}(I_j - \tilde{x}_j))|],$$

where we have also used the fact that $\tilde{\pi}_i^\Delta(\mathbf{1}_{\overline{C_i(\Delta)}}(I_j - \tilde{x}_j)) = 0$. For all $x \in C_i(\Delta)$,

$$\beta\Delta \geq \beta\|Y_i - \tilde{h}_i(x)\| \geq \|\tilde{h}_i^{-1}(Y_i) - x\| \geq |\tilde{x}_j - I_j(x)|,$$

and, thus, $\mathbf{1}_{C_i(\Delta)}(x)(I_j(x) - \tilde{x}_j) < \beta\Delta$ for all $x \in \mathbb{R}^{d_s}$. Consequently,

$$\mathrm{E}[|\pi_i(I_j) - \tilde{\pi}_i^\Delta(I_j)|] \leq \beta\Delta\,\mathrm{E}\left[\sup_{\|\varphi\|_\infty \leq 1} \mathrm{E}[|\pi_i\varphi - \tilde{\pi}_i^\Delta\varphi| \mid \mathcal{Y}_i]\right] + \mathrm{E}[|\pi_i(\mathbf{1}_{\overline{C_i(\Delta)}}(I_j - \tilde{x}_j))|].$$
$$(4.17)$$

For the second term, we observe similarly as in the proof of Theorem 3.1 that

$$\mathrm{E}[|\pi_i(\mathbf{1}_{\overline{C_i(\Delta)}}(I_j - \tilde{x}_j))|] \leq \mathrm{E}[\pi_i(\mathbf{1}_{\overline{C_i(\Delta)}}\|I - \tilde{x}\|)]$$
$$= \mathrm{E}\left[\iint_{\overline{D_i(\Delta)}} g_{i,y}(x)\|x - \tilde{h}_i^{-1}(y)\|\lambda_{d_m}(\mathrm{d}y)\pi_{i-1}K_i(\mathrm{d}x)\right],$$

and, for the inner integral, we have, by a change of variables,

$$\int_{\overline{D_i(\Delta)}} g_{i,y}(x)\|x - \tilde{h}_i^{-1}(y)\|\lambda_{d_m}(\mathrm{d}y) \leq \int_{\|z\| > \Delta - \gamma} \beta(\|z\| + \gamma)\rho_{V_i}(z)\lambda_{d_m}(\mathrm{d}z)$$
$$\leq c_{17}(\epsilon)\exp((-A_2 + \epsilon)(\Delta - \gamma)^{B_2}). \qquad (4.18)$$

The claim then follows by applying Theorem 4.1 to the first term on the right-hand side of (4.17) and (4.18) to the second term on the right-hand side of (4.17).

5. An example of a uniformly convergent particle filter

In this section we describe a feasible filter approximation $\tilde{\pi}^\Delta$ which is based on the Monte Carlo method and can be parameterised to satisfy assumptions (A5) and (A6) of the previous section. We define $\tilde{\pi}^\Delta$ to be a random probability measure-valued process satisfying the following definitions.

(i) *Initialisation.* Define $\{\bar{X}_0^j\}_{j=1}^N$ to be a set of independent random variables with common distribution π_0.

(ii) *Importance sampling.* For all $i > 0$ and $0 < j \leq N$, define X_i^j to be an independent random variable with distribution $\tilde{K}_i(\bar{X}_{i-1}^j, \cdot)$ such that the Radon–Nikodým derivative

$$w_i^\Delta(x, y) := \frac{\mathrm{d}K_i^\Delta(x, \cdot)}{\mathrm{d}\tilde{K}_i(x, \cdot)}(y)$$

exists, is positive for all $x \in C_{i-1}(\Delta)$, and satisfies $\sup_{i>0}\|w_i^\Delta\|_\infty < \infty$.

(iii) *Weight update.* For all $i > 0$,

$$\tilde{\pi}_i^\Delta := \sum_{j=1}^N W_i^j \delta_{\{X_i^j\}},$$

where $\delta_{\{x\}}$ denotes the unit point mass located at $x \in \mathbb{R}^{d_s}$ and

$$W_i^j := \frac{g_i^\Delta(X_i^j) w_i^\Delta(\bar{X}_{i-1}^j, X_i^j)}{\sum_{n=1}^N g_i^\Delta(X_i^n) w_i^\Delta(\bar{X}_{i-1}^n, X_i^n)}.$$

(iv) *Resampling.* For all $i > 0$ and $0 < j \le N$, define random variables $\{\bar{X}_i^j\}_{j=1}^N$ to be a set of random variables such that

$$\mathrm{E}[\zeta_i \mid \mathcal{H}_i] = N W_i,$$

and there exists $c_{18} > 0$ such that, for all $z = (z_1, z_2, \ldots, z_N)^\top$, where $|z_i| \le 1$, $i = 1, \ldots, N$, we have

$$z^\top \mathrm{E}[(\zeta_i - N W_i)(\zeta_i - N W_i)^\top \mid \mathcal{H}_i]z \le c_{18}N,$$

where $W_i := (W_i^1, W_i^2, \ldots, W_i^N)^\top$, $\zeta_i := (\zeta_i^1, \zeta_i^2, \ldots, \zeta_i^N)^\top$, and

$$\zeta_i^j := \sum_{\ell=1}^N \mathbf{1}_{\{X_i^j\}}(\bar{X}_i^\ell).$$

In order to verify that $\tilde{\pi}^\Delta$ satisfies assumptions (A5) and (A6), we first observe that $\tilde{\pi}^\Delta$ is of the form (4.1), where

$$\nu_i^\Delta := \frac{1}{N} \sum_{j=1}^N w_i^\Delta(\bar{X}_{i-1}^j, X_i^j)\delta_{\{X_i^j\}}.$$

According to (ii), $w_i^\Delta > 0$, P-a.s. and, hence, assumption (A5) is satisfied. In order to verify assumption (A6), we define σ-fields $\mathcal{H}_i$ and $\bar{\mathcal{H}}_i$ as

$$\mathcal{H}_i := \sigma(Y_m, X_n^j, \bar{X}_{n-1}^j, 1 \le n \le i, 1 \le j \le N, 0 < m),$$
$$\bar{\mathcal{H}}_i := \sigma(Y_m, X_n^j, \bar{X}_\ell^j, 1 \le n \le i, 1 \le j \le N, 0 \le \ell \le i, 0 < m),$$

and a random probability measure $\bar{\pi}_i^\Delta$ as

$$\bar{\pi}_i^\Delta := \frac{1}{N} \sum_{j=1}^N \zeta_i^j \delta_{\{X_i^j\}}.$$

In this case, because $\mathcal{H}_{i-1} \subset \bar{\mathcal{H}}_{i-1}$, we can write

$$\mathrm{E}[|\tilde{\pi}_{i-1}^\Delta K_i^\Delta \varphi - \nu_i^\Delta \varphi| \mid \mathcal{H}_{i-1}] \le \mathrm{E}[\mathrm{E}[|\bar{\pi}_{i-1}^\Delta K_i^\Delta \varphi - \nu_i^\Delta \varphi| \mid \bar{\mathcal{H}}_{i-1}] \mid \mathcal{H}_{i-1}] \quad (5.1)$$
$$+ \mathrm{E}[|\tilde{\pi}_{i-1}^\Delta K_i^\Delta \varphi - \bar{\pi}_{i-1}^\Delta K_i^\Delta \varphi| \mid \mathcal{H}_{i-1}]. \quad (5.2)$$

Because $\mathrm{E}[\nu_i^\Delta \varphi \mid \bar{\mathcal{H}}_{i-1}] = \bar{\pi}_i^\Delta K_i^\Delta \varphi$ and the random variables $\{X_i^j\}_{j=1}^N$ are independent conditioned on $\bar{\mathcal{H}}_{i-1}$, it can be shown that

$$\mathrm{E}[|\bar{\pi}_{i-1}^\Delta K_i^\Delta \varphi - \nu_i^\Delta \varphi| \mid \bar{\mathcal{H}}_{i-1}] \le \frac{\|w_i^\Delta \varphi\|_\infty}{\sqrt{N(\Delta)}}. \quad (5.3)$$

Also, it can be shown that if (iv) holds then there exists $c_{19} > 0$ such that

$$\mathrm{E}[|\tilde{\pi}_i^\Delta \varphi - \bar{\pi}_i^\Delta \varphi| \mid \mathcal{H}_i] \leq \frac{c_{19}\|\varphi\|_\infty}{\sqrt{N(\Delta)}}. \tag{5.4}$$

The proof of this inequality can be found, for example, in [2]. Because

$$\sup_{i>0} \sup_{\|\varphi\|_\infty \leq 1} \|K_i\varphi\|_\infty = 1$$

and, according to (ii), $\sup_{i>0}\|w_i^\Delta\|_\infty < \infty$, the substitution of (5.3) and (5.4) into (5.1) shows that there exists $c_{20}(\Delta) > 0$ such that

$$\sup_{i>0} \sup_{\|\varphi\|_\infty \leq 1} \mathrm{E}[|\tilde{\pi}_i^\Delta K_i^\Delta \varphi - \nu_i^\Delta \varphi| \mid \mathcal{H}_{i-1}] \leq \frac{c_{20}(\Delta)}{\sqrt{N(\Delta)}},$$

from which assumption (A6) follows if

$$N(\Delta) \geq \left(\frac{c_{20}(\Delta)}{M_4}\right)^2 \exp(2a_4 \Delta^{b_2}). \tag{5.5}$$

The approximation specified by (i)–(iv) above has resemblance to the well-known *sampling/importance resampling (SIR)* algorithm. Essentially, the only difference to the SIR filter is the requirement imposed by assumption (A5), that the importance distribution specified by the transition probabilities $\tilde{K}_i$ must assign zero probability to the set $\overline{C_i(\Delta)}$. As a general rule for specifying such importance distribution, we can set, for all $x \in \mathbb{R}^{d_s}$,

$$\tilde{K}_i(x, \cdot) = \mathrm{P}_{\tilde{h}_i^{-1}(Z)}(\cdot), \tag{5.6}$$

where $\mathrm{P}_{\tilde{h}_i^{-1}(Z)}$ denotes the distribution of the random variable $\tilde{h}_i^{-1}(Z)$ and Z is a random variable with uniform distribution on the ball $B_{d_m}(Y_i, \Delta)$. This choice of $\tilde{K}_i$ obviously satisfies assumption (A5). Moreover, by assuming that, for all $x \in \mathbb{R}^{d_s}$, we have $|\det(\tilde{h}_i'(x))| > 0$, where $\tilde{h}_i'(x)$ denotes the Jacobian matrix of $\tilde{h}_i$ at $x \in \mathbb{R}^{d_s}$, the density $\rho_{X_i^j}$ of the random variable X_i^j with respect to the Lebesgue measure on $\mathcal{B}(\mathbb{R}^{d_s})$ satisfies

$$\rho_{X_i^j}(x) = \frac{\mathbf{1}_{C_i(\Delta)}(x)|\det \tilde{h}_i'(x)|}{V_{d_s}(\Delta)},$$

where $V_{d_s}(\Delta)$ denotes the volume of a d_s-dimensional ball of radius Δ. From assumption (A2), it follows that

$$\sup_{i>0} \|w_i^\Delta\|_\infty = \sup_{\substack{x,y \in \mathbb{R}^{d_s} \\ i>0}} \frac{V_{d_s}(\Delta)\,\mathbf{1}_{C_i(\Delta)}(y)k_i(x,y)}{|\det \tilde{h}_i'(x)|} \leq c_{21}\Delta^{d_s}$$

for some $c_{21} > 0$. The substitution of this approximation into (5.5) yields a lower bound for the sample size $N(\Delta)$, which, according to the discussion above, implies assumption (A6). This particular choice of importance distribution was made because the generation of such a random variate is feasible in higher dimensions without resorting to the rejection method. Indeed, it may be possible to generate random variables from a more efficient importance distribution using

the rejection method, but then, as pointed out in Section 1, we must also take the randomness of the computational cost into account.

It is also observed that the *branching* or *resampling scheme* in (iv) is left unspecified as well. From the practical point of view, we content ourselves with pointing out that the well-known *multinomial resampling* scheme as well as the *tree-based branching* scheme both satisfy (iv). For the proof, see, e.g. [2] and the references therein.

6. Numerical experiments

In this section we illustrate Theorem 4.1 and Corollary 4.1 by some numerical experiments. The experiments are carried out using the approximate filter $\tilde{\pi}^\Delta$ described in the previous sections and it employs the multinomial resampling scheme and an importance distribution of the form given in (5.6).

6.1. Linear model

In the first experiment we consider the model

$$X_{i+1} = X_i + W_{i+1}, \qquad Y_{i+1} = 4X_{i+1} + V_{i+1},$$

where $X_0 \sim \mathcal{N}(0, 1)$, $W_i \sim \mathcal{N}(0, 2)$, and $V_i \sim \mathcal{N}(0, 1)$ independently. Here $\mathcal{N}(x, y)$ denotes the normal distribution with mean equal to x and covariance equal to y, but it should be noted that this normality assumption is not crucial for the validity of Theorem 3.1 or Theorem 4.1. Indeed, the distributions of X_0, W_i, and V_i could also be, for example, double exponential, Gaussian mixtures, convolutions of Gaussian, or double exponential distributions with distributions having a bounded support. Also, it should be noted that, for this linear and Gaussian model, the filtering distribution π_i is known to be equal to a normal distribution whose mean and covariance can be computed by the Kalman filter. Therefore, its numerical approximations are not of any interest in practice, and it has been included in the experiments only because it enables exact computations of the approximation errors.

The simulations consisted of 1000 time steps and the filtering distributions were approximated with eight different values of Δ ($\Delta = 0.625, 1.250, 1.875, 2.500, 3.125, 3.750, 4.375, 5.000$). This range of Δ was found to give a reasonable illustration of the behaviour of the approximations as a function of Δ. For each Δ, four different sample sizes ($N = 10^2, 10^3, 10^4, 10^5$) were used.

The error under consideration is the expected distance between the exact and approximate conditional means, i.e. $E[|\hat{x}_i - \hat{x}_i^\Delta|]$. Because the exact evaluation of the expectation is intractable, it was approximated by the Monte Carlo method with 50 simulations. Let this approximation be denoted by $e_i(\Delta)$. Figure 2a depicts $e_i(\Delta)$ for four different pairs (Δ, N).

For a fixed value of Δ, Figure 2a suggests that $e_i(\Delta)$ is approximately a constant with respect to i, which is in accordance with Corollary 4.1. Therefore, the time average of $e_i(\Delta)$, denoted by $e(\Delta)$, is expected to be a relatively good approximation of the uniform bound of error. Figure 2b shows the time averages for all experimented values of Δ and N. To study the behaviour of the algorithm on a longer time scale, the experiment was repeated for a simulation consisting of 10^5 time steps for two different values of Δ, each with $N = 10^5$. Figure 3 depicts the error $e_i(\Delta)$ based on 50 simulations and seems to confirm that the error is uniformly bounded.
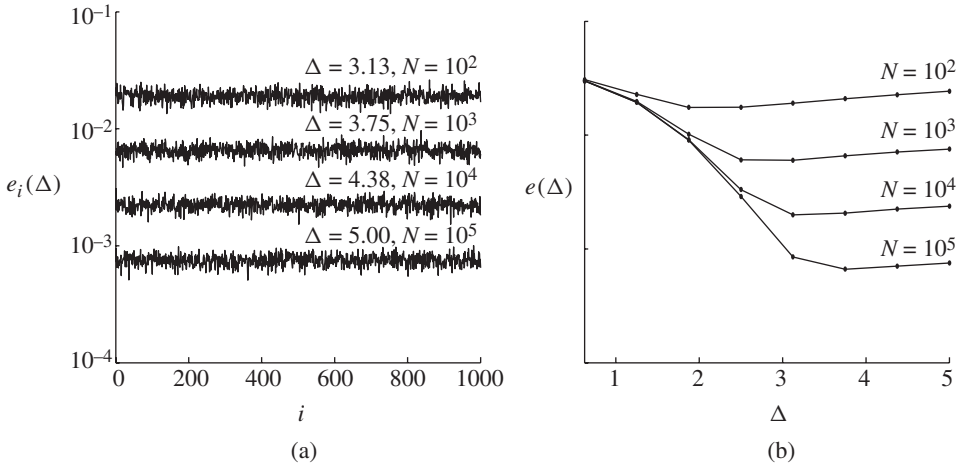
FIGURE 2: (a) The error $e_i(\Delta)$ for four different pairs (Δ, N). (b) Time averages of $e_i(\Delta)$ for different values of N and Δ.
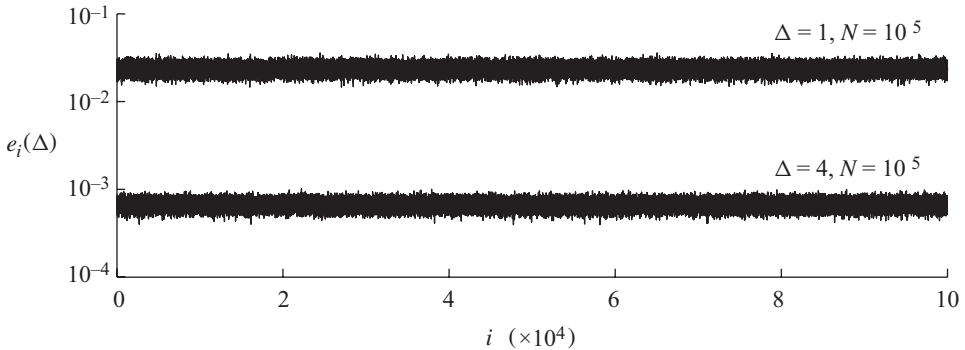

FIGURE 3: The error $e_i(\Delta)$ for two different values of Δ and $N = 10^5$.

6.2. Nonlinear model

In the second experiment we consider the model

$$X_{i+1} = \frac{X_i}{2} + 25\frac{X_i}{1 + X_i^2} + 8\cos(1.2i) + W_{i+1},$$

$$Y_{i+1} = 4X_{i+1} + 4\sin(2X_{i+1}) + V_{i+1},$$

where $X_0 \sim \mathcal{N}(0, 1)$, $W_i \sim \mathcal{N}(0, 2)$, and $V_i \sim \mathcal{N}(0, 0.006)$. This example is adapted from the well-known example in [8] and it does not allow the exact computation of the posterior mean $\hat{x}_i$. The simulation consisted of 1000 time steps with eight different truncation radii ($\Delta = 3.25$, 3.50, 3.75, 4.00, 4.25, 4.50, 4.75, 5.00) and four different sample sizes. The sample sizes are the same as in the linear case, but because of the bounded component $\bar{h}_i(x) = 4\sin(2x)$ in the observation model, the values of Δ were chosen differently to give a reasonable illustration of the behaviour of the approximation as a function of Δ.
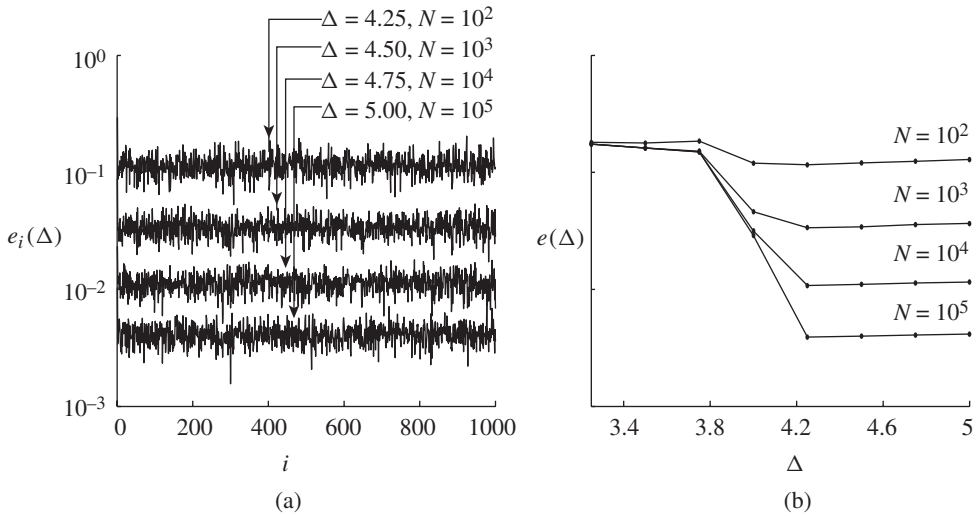
FIGURE 4: (a) The error $e_i(\Delta)$ for four different pairs (Δ, N). (b) Time averages of $e_i(\Delta)$ for different values of N and Δ.

Because the posterior mean cannot be evaluated exactly for the nonlinear model, a conventional SIR filter with 10^6 particles was used as a reference mean. Figure 4a depicts the approximation $e_i(\Delta)$ of the expected distance between the approximate posterior mean and the reference mean. Again, $e_i(\Delta)$ is the Monte Carlo approximation with 50 samples.

Similarly as for the first example, the errors appear to be nearly constant with respect to the time. Figure 4b depicts the time averages of the errors.

6.3. Discussion

The simulations described above were implemented in MATLAB® (the resampling was implemented in C) and executed on a 1 GHz PowerPC G4 with 768 MB of memory. For the linear model, the computation time of processing ten data sets of length 100 with $N = 10^5$ was approximately 172 seconds. For the nonlinear model, this time was approximately 214 seconds. The computation time of the uniformly convergent filter was approximately 1.1 times that of the conventional SIR filter. Although the computation time is highly dependent on implementational details, and there might be room for more efficient implementations, it should be noted that the uniformly convergent filter and SIR filter are roughly equal in terms of computational cost.

Figure 2b and Figure 4b also illustrate the twofold construction of the approximation error. For the smallest values of Δ, the error reaches a level which apparently cannot be exceeded by increasing the sample size. The remaining error is expected to be explained by the error due to the truncation and, therefore, it cannot be affected by the sample size. For both examples, such behaviour does not appear for the largest values of Δ. Also, Figure 2b and Figure 4b show that the error has a tendency to increase for large enough Δ. This is the reason why the sample size must be defined as an increasing function of Δ in order to ensure that the error converges to 0 as $\Delta \to \infty$.

In principle, we could approximate the constants in (5.5) that determine the rate at which the sample size should be increased. Moreover, bounds for the errors could also be evaluated according to the analysis given earlier. However, these bounds are expected to be relatively loose and, therefore, only of limited interest in practice.

Appendix A. Proof of Proposition 3.1

The following lemma establishes the tails for the convolution of two densities of positive random variables with exponential tails.

Lemma A.1. *Let* $a_X, a_Y, b_X, b_Y > 0$, *and let* $\rho_X, \rho_Y \colon \mathbb{R} \to \mathbb{R}_+$ *be probability density functions with respect to the Lebesgue measure* λ_1 *on* $\mathcal{B}(\mathbb{R})$ *such that, for all* $x \leq 0$, $\rho_X(x) = \rho_Y(x) = 0$ *and, for all* $\epsilon \in (0, \min(a_X, a_Y))$, *there exist* $M_X = M_X(\epsilon) > 0$ *and* $M_Y = M_Y(\epsilon) > 0$ *such that*

$$\rho_X(x) \leq M_X \exp((-a_X + \epsilon)x^{b_X}),$$
$$\rho_Y(x) \leq M_Y \exp((-a_Y + \epsilon)x^{b_Y}),$$

for all $x > 0$. *Then, for all* $\epsilon > 0$, *there exists* $c_{22} = c_{22}(\epsilon, a_X, a_Y, b_X, b_Y) > 0$ *such that, for all* $x > 0$,

$$(\rho_X * \rho_Y)(x) \leq c_{22} \exp((-a + \epsilon)x^b),$$

where $b = \min(b_X, b_Y)$ *and*

$$
a = \begin{cases}
a_X a_Y (a_X^{1/b} + a_Y^{1/b})^{-b}, & b_X = b_Y, \\
a_X, & b_X < b_Y, \\
a_Y, & b_X > b_Y.
\end{cases}
$$

Proof. If $\epsilon > \max(a_X, a_Y) \geq a$, the claim follows trivially. Therefore, we assume that $\epsilon_1 \in (0, a_X)$ and $\epsilon_2 \in (0, a_Y)$. For all $t \in (0, 1)$ and $y \geq 0$,

$$(\rho_X * \rho_Y)(y) \leq M_X(\epsilon_1) M_Y(\epsilon_2)(J_1(\epsilon_1, \epsilon_2, y) + J_2(\epsilon_1, \epsilon_2, y)), \tag{A.1}$$

where

$$J_1(\epsilon_1, \epsilon_2, y) = \int_0^{ty} \exp((-a_X + \epsilon_1)(y - x)^{b_X} + (-a_Y + \epsilon_2)x^{b_Y}) \, dx,$$
$$J_2(\epsilon_1, \epsilon_2, y) = \int_{ty}^{y} \exp((-a_X + \epsilon_1)(y - x)^{b_X} + (-a_Y + \epsilon_2)x^{b_Y}) \, dx.$$

By defining

$$c_{23}(\epsilon) := \int_0^{\infty} \exp((-a_Y + \epsilon)x^{b_Y}) \, dx,$$

we observe that

$$J_1(\epsilon_1, \epsilon_2, y) \leq c_{23}(\epsilon_2) \exp((-a_X + \epsilon_1)(1 - t)^{b_X} y^{b_X}). \tag{A.2}$$

Because, for all $\delta_0 > 0$, there exists $c_{24}(\epsilon_2, \delta_0) > 0$ such that if $ty > \delta_0$ then

$$J_2(\epsilon_1, \epsilon_2, y) \leq \int_{ty}^{\infty} \exp((-a_Y + \epsilon_2)x^{b_Y}) \, dx \leq c_{24}(\epsilon_2, \delta_0) \exp((-a_Y + 2\epsilon_2)t^{b_Y} y^{b_Y}),$$

it can be shown that by defining

$$c_{25}(\epsilon_1, \epsilon_2) = \max(c_{24}(\epsilon_2, \delta_0), c_{21}(\epsilon_2) \exp((a_Y - 2\epsilon_2)\delta_0^{b_Y}))$$

we have, for all $y > 0$,

$$J_2(\epsilon_1, \epsilon_2, y) \le c_{25}(\epsilon_1, \epsilon_2) \exp((-a_Y + 2\epsilon_2)t^{b_Y} y^{b_Y}). \tag{A.3}$$

By combining (A.1), (A.2), and (A.3), we find that, for all $\epsilon > 0$, there exist $c_{26}(\epsilon) > 0$ and $c_{27}(\epsilon) > 0$ such that

$$(\rho_X * \rho_Y)(y) \le c_{26}(\epsilon) \exp((-a_X + \epsilon)(1 - t)^{b_X} y^{b_X}) + c_{27}(\epsilon) \exp((-a_Y + \epsilon)t^{b_Y} y^{b_Y}). \tag{A.4}$$

Let us then consider the case in which $b_X < b_Y$. Since convolution commutes, b_X and b_Y are interchangeable, and the same reasoning applies to the case in which $b_Y < b_X$ as well. If we define

$$t(z) = \begin{cases} \dfrac{1}{2} - \dfrac{1}{2}\left(\dfrac{a_X - 2z}{a_X - z}\right)^{1/b_X} & \text{if } 0 < z \le a_X/2, \\ \dfrac{1}{2} & \text{if } a_X/2 < z, \end{cases}$$

it follows that $a_X - 2z < (a_X - z)(1 - t(z))^{b_X}$, and, therefore, by (A.4),

$$(\rho_X * \rho_Y)(y) \le c_{28}(\epsilon) \exp((-a_X + 2\epsilon)y^{b_X}),$$

where

$$c_{28}(\epsilon) = \sup_{y > 0}(c_{26}(\epsilon) \exp((-a_X + \epsilon)(1 - t(\epsilon))^{b_X} y^{b_X} + (a_X - 2\epsilon)y^{b_X})$$
$$+ c_{27}(\epsilon) \exp((-a_Y + 2\epsilon)t^{b_Y} y^{b_Y} + (a_X - 2\epsilon)y^{b_X})),$$

and we have the claim for $b_X \ne b_Y$. In the case in which $b_X = b_Y = b$ we define $t(z)$ for all $z \in (0, \min(a_X, a_Y))$ as

$$t(z) = \frac{(a_X - z)^{1/b}}{(a_Y - z)^{1/b} + (a_X - z)^{1/b}},$$

and observe that

$$\lim_{z \to 0}(-a_X + z)(1 - t(z))^b = \lim_{z \to 0}(-a_Y + z)t(z)^b = -a.$$

Therefore, for all $\epsilon > 0$, there exists $z \in (0, \min(a_X, a_Y))$ such that $(-a_X + z)(1 - t(z))^b \le -a + \epsilon$ and $(-a_Y + z)t(z)^b \le -a + \epsilon$. The substitution of these inequalities into (A.4) yields the claim for $b_X = b_Y$.

References

[1] ATAR, R. AND ZEITOUNI, O. (1997). Exponential stability for nonlinear filtering. *Ann. Inst. H. Poincaré Prob. Statist.* **33**, 697–725.

[2] CRISAN, D. (2001). Particle filters - a theoretical perspective. In *Sequential Monte Carlo Methods in Practice*, eds A. Doucet *et al.*, Springer, New York, pp. 17–41.

[3] CRISAN, D. AND HEINE, K. (2008). Stability of the discrete time filter in terms of the tails of noise distributions. *J. London Math. Soc.* **78**, 441–458.

[4] DEL MORAL, P. (1998). A uniform convergence theorem for the numerical solving of the nonlinear filtering problem. *J. Appl. Prob.* **4**, 873–884.

[5] DEL MORAL, P. AND GUIONNET, A. (2001). On the stability of interacting processes with applications to filtering and genetic algorithms. *Ann. Inst. H. Poincaré Prob. Statist.* **37,** 155–194.

[6] DOBRUSHIN, R. L. (1956). Central limit theorem for non-stationary Markov chains. I. *Theory Prob. Appl.* **1,** 72–89.

[7] HEINE, K. (2007). On the stability of the discrete time filter and the uniform convergence of its approximations. Doctoral Thesis, Tampere University of Technology.

[8] KITAGAWA, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Comput. Graph. Statist.* **5,** 1–25.

[9] KONTOYIANNIS, I. AND MEYN, S. P. (2005). Large deviations asymptotics and the spectral theory of multiplicatively regular Markov processes. *Electron. J. Prob.* **10,** 61–123.

[10] KÜNSCH, H. (2005). Recursive Monte Carlo filters: algortihms and theoretical analysis. *Ann. Statist.* **33,** 1983–2021.

[11] LEGLAND, F. AND OUDJANE, N. (2003). A robustification approach to stability and to uniform particle approximation of nonlinear filters: the example of pseudo-mixing signals. *Stoch. Process. Appl.* **106,** 279–316.

[12] LEGLAND, F. AND OUDJANE, N. (2004). Stability and uniform approximation of nonlinear filters using the Hilbert metric, and applications to particle filters. *Ann. Appl. Prob.* **14,** 144–187.

[13] OUDJANE, N. AND RUBENTHALER, S. (2005). Stability and uniform particle approximation of nonlinear filters in case of non ergodic signals. *Stoch. Anal. Appl.* **23,** 421–448.

[14] PAPAVASILIOU, A. (2005). A uniformly convergent adaptive particle filter. *J. Appl. Prob.* **4,** 1053–1068.

[15] WU, L. (2000). Some notes on large deviations of Markov processes. *Acta Math. Sin. (Engl. Ser.)*, **16,** 369–394.