



HAL
open science

Adaptive methods for sequential importance sampling with application to state space models

Julien Cornebise, Eric Moulines, Jimmy Olsson

► **To cite this version:**

Julien Cornebise, Eric Moulines, Jimmy Olsson. Adaptive methods for sequential importance sampling with application to state space models. 2008. hal-00259951v1

HAL Id: hal-00259951

<https://hal.science/hal-00259951v1>

Preprint submitted on 29 Feb 2008 (v1), last revised 22 Aug 2008 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ADAPTIVE METHODS FOR SEQUENTIAL IMPORTANCE SAMPLING WITH APPLICATION TO STATE SPACE MODELS

JULIEN CORNEBISE, ÉRIC MOULINES, AND JIMMY OLSSON

ABSTRACT. In this paper we discuss new adaptive proposal strategies for sequential Monte Carlo algorithms—also known as particle filters—relying on criterions evaluating the quality of the proposed particles. The choice of the proposal distribution is a major concern and can dramatically influence the quality of the estimates. Thus, we show how the long-used coefficient of variation (suggested by Kong *et al.*, 1994) of the weights can be used for estimating the chi-square distance between the target and instrumental distributions of the auxiliary particle filter. As a by-product of this analysis we obtain an auxiliary adjustment multiplier weight type for which this chi-square distance is minimal. Moreover, we establish an empirical estimate of linear complexity of the Kullback-Leibler divergence between the involved distributions. Guided by these results, we discuss adaptive designing of the particle filter proposal distribution, e.g., by means of population Monte Carlo techniques, and illustrate the methods on several numerical examples.

1. INTRODUCTION

Easing the role of the user by tuning automatically the key parameters of *sequential Monte Carlo* (SMC) *algorithms* has been a long-standing topic in the community, notably through adaptation of the particle sample size or the way the particles are proposed. In this paper we focus on the latter issue and develop methods for adjusting adaptively the importance sampling distribution of the particle filter.

Adaptation of the number of particles has been treated by several authors. Legland and Oudjane (2006) (and later Hu *et al.*, 2008, Section IV) increase the size of the particle sample until the total weight mass reaches a positive threshold, avoiding a situation where all particles are located in regions of the state space having zero posterior probability. Another approach, suggested by Fox (2003) and refined in (Soto, 2005; Straka and Simandl, 2006), consists in increasing the sample size until the *Kullback-Leibler divergence* (KLD) between the true and estimated target distributions is below a given threshold.

Unarguably, setting an appropriate sample size is a key ingredient of any statistical estimation procedure, and there are cases where the methods mentioned above may be used for designing this size satisfactorily; however increasing the sample size only is far from being always sufficient for achieving efficient variance reduction. Indeed, as in any algorithm based

2000 *Mathematics Subject Classification.* Primary 65C05; Secondary 60G35.

Key words and phrases. Adaptive Monte Carlo, auxiliary particle filter, chi-square distance, coefficient of variation, Kullback-Leibler divergence, Population Monte Carlo, sequential Monte Carlo, Shannon entropy, state space models.

on importance sampling, a significant discrepancy between the proposal and target distributions may require an unreasonably large number of samples for decreasing the variance of the estimate under a specified value. For a very simple illustration, consider importance sampling estimation of the mean m of a normal distribution using as importance distribution another normal distribution having zero mean and same variance: in this case, the variance of the estimate grows like $\exp(m^2)/N$, N denoting the number of draws, implying that the sample size required for ensuring a given variance grows exponentially fast with m .

This points to the need for adapting the importance distributions of the particle filter, e.g., by adjusting the proposal kernels. Less work has been done on this topic, with the notable exception of (Pitt and Shephard, 1999) (see also Doucet *et al.*, 2000), in which methods for approximating the so-called *optimal kernel* are developed, and, more recently, (Chan *et al.*, 2003), in which the authors aim at minimizing the expectation of a cost function, such as the mean square error or the negative of the *effective sample size*, over a parametric family of kernels.

Most of the algorithms described above require tools, such as, e.g., the *coefficient of variation* (CV) proposed by Kong *et al.* (1994), for evaluating on-line the quality of the particle swarm. In this article we justify theoretically (Theorem 4.1(ii)) that the CV can be used for estimating sequentially the *chi-square distance* (CSD) between the auxiliary SMC target and importance distributions. We also propose a new empirical estimate of the KLD (Theorem 4.1(i)) having, like the CV, a computational complexity which is linear in the number of particles. Moreover, by examining the derived asymptotic CSD we identify a type of auxiliary SMC adjustment multiplier weights which minimize this CSD for a given proposal kernel (Corollary 4.1). In Section 5 we use—and this is the main objective of the paper—these empirical estimates to design adaptive algorithms to construct particle approximations. Finally, in the implementation section (Section 6), we apply the proposed algorithms to approximate the filtering distribution in several state-space models.

2. SOME NOTATION AND DEFINITIONS

In the following we assume that all random variables are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let, for any general state space $(\Xi, \mathcal{B}(\Xi))$, $\mathcal{P}(\Xi)$ and $\mathbb{B}(\Xi)$ be the sets of probability measures on $(\Xi, \mathcal{B}(\Xi))$ and measurable functions from Ξ to \mathbb{R} , respectively.

A kernel K from $(\Xi, \mathcal{B}(\Xi))$ to some other state space $(\tilde{\Xi}, \mathcal{B}(\tilde{\Xi}))$ is called *finite* if $K(\xi, \tilde{\Xi}) < \infty$ for all $\xi \in \Xi$ and *Markovian* if $K(\xi, \tilde{\Xi}) = 1$ for all $\xi \in \Xi$. Moreover, K induces two operators, one transforming a function $f \in \mathbb{B}(\Xi \times \tilde{\Xi})$ satisfying $\int_{\tilde{\Xi}} |f(\xi, \tilde{\xi})| K(\xi, d\tilde{\xi}) < \infty$ into another function

$$\xi \mapsto K(\xi, f) \triangleq \int_{\tilde{\Xi}} f(\xi, \tilde{\xi}) K(\xi, d\tilde{\xi})$$

in $\mathbb{B}(\tilde{\Xi})$; the other transforms a measure $\nu \in \mathcal{P}(\Xi)$ into another measure

$$A \mapsto \nu K(A) \triangleq \int_{\Xi} K(\xi, A) \nu(d\xi) \tag{2.1}$$

in $\mathcal{P}(\tilde{\Xi})$. Furthermore, for any probability measure $\mu \in \mathcal{P}(\Xi)$ and function $f \in \mathbb{B}(\Xi)$ satisfying $\int_{\Xi} |f(\xi)| \mu(d\xi) < \infty$, we write $\mu(f) \triangleq \int_{\Xi} f(\xi) \mu(d\xi)$.

The *outer product* of the measure γ and the kernel T , denoted by $\gamma \otimes T$, is defined as the measure on the product space $\Xi \times \tilde{\Xi}$, equipped with the product σ -algebra $\mathcal{B}(\Xi) \otimes \mathcal{B}(\tilde{\Xi})$, satisfying

$$\nu \otimes K(A) \triangleq \iint_{\Xi \times \tilde{\Xi}} \gamma(d\xi) T(\xi, d\xi') \mathbb{1}_A(\xi, \xi') \quad (2.2)$$

for any $A \in \mathcal{B}(\Xi) \otimes \mathcal{B}(\tilde{\Xi})$. For a non-negative function $f \in \mathbb{B}(\Xi)$, we define the measure $\gamma[f] \in \mathcal{P}(\Xi)$ by

$$\nu[f](A) \triangleq \nu(f \mathbb{1}_A), \quad (2.3)$$

for any $A \in \mathcal{B}(\Xi)$.

Finally, a set \mathcal{C} of real-valued functions on Ξ is said to be *proper* if the following conditions hold: **i)** \mathcal{C} is a linear space; **ii)** if $g \in \mathcal{C}$ and f is measurable with $|f| \leq |g|$, then $f \in \mathcal{C}$; **iii)** for all $c \in \mathbb{R}$, the constant function $f \equiv c$ belongs to \mathcal{C} .

We call a set $\{\xi_{N,i}\}_{i=1}^{M_N}$ of random variables, referred to as *particles* and taking values in Ξ , and nonnegative weights $\{\omega_{N,i}\}_{i=1}^{M_N}$ a *weighted sample* on Ξ . Here M_N is a (possibly random) natural number. From (Douc and Moulines, 2005) we adopt the following definition, where $\Omega_N \triangleq \sum_{j=1}^{M_N} \omega_{N,j}$.

Definition 2.1. A weighted sample $\{(\xi_{N,i}, \omega_{N,i})\}_{i=1}^{M_N}$ on Ξ is said to be consistent for the probability measure $\nu \in \mathcal{P}(\Xi)$ and the proper set \mathcal{C} if, for any $f \in \mathcal{C}$, as $N \rightarrow \infty$,

$$\begin{aligned} \Omega_N^{-1} \sum_{i=1}^{M_N} \omega_{N,i} f(\xi_{N,i}) &\xrightarrow{\mathbb{P}} \nu(f), \\ \Omega_N^{-1} \max_{1 \leq i \leq M_N} \omega_{N,i} &\xrightarrow{\mathbb{P}} 0. \end{aligned}$$

Alternatively, we will sometimes say that the weighted sample in Definition 2.1 *targets* the measure ν .

3. AUXILIARY SEQUENTIAL IMPORTANCE SAMPLING

As mentioned in the introduction, the objective of this paper is the development of adaptive methods for the *auxiliary sequential importance sampling* (ASIS) *algorithm* (originally proposed by Pitt and Shephard, 1999) which is recalled below.

Suppose that we have at hand a weighted sample $\{(\xi_{N,i}, \omega_{N,i})\}_{i=1}^{M_N}$ targeting $\nu \in \mathcal{P}(\Xi)$. We wish to transform this sample into a new weighted particle sample approximating the probability measure

$$\mu(\cdot) \triangleq \frac{\nu L(\cdot)}{\nu L(\tilde{\Xi})} = \frac{\int_{\Xi} L(\xi, \cdot) \nu(d\xi)}{\int_{\Xi} L(\xi', \tilde{\Xi}) \nu(d\xi')} \quad (3.1)$$

on some other state space $(\tilde{\Xi}, \mathcal{B}(\tilde{\Xi}))$. Here L is a finite transition kernel from $(\Xi, \mathcal{B}(\Xi))$ to $(\tilde{\Xi}, \mathcal{B}(\tilde{\Xi}))$. A natural strategy for achieving this is to replace ν in (3.1) by its particle approximation, yielding

$$\mu_N(\cdot) \triangleq \sum_{i=1}^{M_N} \frac{\omega_{N,i} L(\xi_{N,i}, \tilde{\Xi})}{\sum_{j=1}^{M_N} \omega_{N,j} L(\xi_{N,j}, \tilde{\Xi})} \left[L(\xi_{N,i}, \cdot) / L(\xi_{N,i}, \tilde{\Xi}) \right]$$

as an approximation of μ , and simulate \tilde{M}_N new particles from this distribution; however, in many applications direct simulation from μ_N is infeasible without the application of the computationally expensive auxiliary accept-reject techniques introduced by Hürzeler and Künsch (1998) and thoroughly analyzed by Künsch (2005). This difficulty can be overcome by simulating new particles $\{\tilde{\xi}_{N,i}\}_{i=1}^{\tilde{M}_N}$ from the instrumental mixture distribution

$$\pi_N(\cdot) \triangleq \sum_{i=1}^{M_N} \frac{\omega_{N,i} \psi_{N,i}}{\sum_{j=1}^{M_N} \omega_{N,j} \psi_{N,j}} R(\xi_{N,i}, \cdot),$$

where $\{\psi_{N,i}\}_{i=1}^{M_N}$ are positive numbers referred to as *adjustment multiplier weights* and R is a Markovian kernel, and associating these particles with weights $\{d\mu_N/d\pi_N(\tilde{\xi}_{N,i})\}_{i=1}^{\tilde{M}_N}$. In this setting, a new particle position is simulated from the stratum $R(\xi_{N,i}, \cdot)$ with probability proportional to $\omega_{N,i} \psi_{N,i}$. Haplessly, the Radon-Nikodym derivative $d\mu_N/d\pi_N$ is expensive to evaluate since this involves summing over M_N terms. Thus, we introduce, as suggested by Pitt and Shephard (1999), an *auxiliary variable* corresponding to the selected stratum, and target instead the measure

$$\mu_N^{\text{aux}}(\{i\} \times A) \triangleq \frac{\omega_{N,i} L(\xi_{N,i}, \tilde{\Xi})}{\sum_{j=1}^{M_N} \omega_{N,j} L(\xi_{N,j}, \tilde{\Xi})} \left[L(\xi_{N,i}, A) / L(\xi_{N,i}, \tilde{\Xi}) \right]$$

on the product space $\{1, \dots, M_N\} \times \tilde{\Xi}$. Since μ_N is the marginal distribution of μ_N^{aux} with respect to the particle position, we may sample from μ_N by simulating instead a set $\{(I_{N,i}, \tilde{\xi}_{N,i})\}_{i=1}^{\tilde{M}_N}$ of indices and particle positions from the instrumental distribution

$$\pi_N^{\text{aux}}(\{i\} \times A) \triangleq \frac{\omega_{N,i} \psi_{N,i}}{\sum_{j=1}^{M_N} \omega_{N,j} \psi_{N,j}} R(\xi_{N,i}, A)$$

and assigning each draw $(I_{N,i}, \tilde{\xi}_{N,i})$ the weight

$$\tilde{\omega}_{N,i} \triangleq \psi_{N,I_{N,i}}^{-1} \frac{dL(\xi_{N,I_{N,i}}, \cdot)}{dR(\xi_{N,I_{N,i}}, \cdot)(\tilde{\xi}_{N,i})} \quad (3.2)$$

being proportional to $d\mu_N^{\text{aux}}/d\pi_N^{\text{aux}}(I_{N,i}, \tilde{\xi}_{N,i})$. Hereafter, we discard the indices and take $\{(\tilde{\xi}_{N,i}, \tilde{\omega}_{N,i})\}_{i=1}^{\tilde{M}_N}$ as an approximation of μ . The algorithm is summarized in Algorithm 1.

Note that setting, for all $1 \leq i \leq M_N$, $\psi_{N,i} \equiv 1$ in Algorithm 1 yields the standard *bootstrap particle filter* presented in (Gordon *et al.*, 1993).

Algorithm 1 Nonadaptive ASIS

Require: $\{(\xi_{N,i}, \omega_{N,i})\}_{i=1}^{M_N}$ targets ν .

1: Draw $\{I_{N,i}\}_{i=1}^{M_N} \sim \mathcal{M}(\tilde{M}_N, \{\omega_{N,j}\psi_{N,j} / \sum_{\ell=1}^{M_N} \omega_{N,\ell}\psi_{N,\ell}\}_{j=1}^{M_N})$,

2: simulate $\{\tilde{\xi}_{N,i}\}_{i=1}^{M_N} \sim \bigotimes_{i=1}^{M_N} R(\xi_{N,I_{N,i}}, \cdot)$,

3: set, for all $i \in \{1, \dots, M_N\}$,

$$\tilde{\omega}_{N,i} \leftarrow \psi_{N,I_{N,i}}^{-1} dL(\xi_{N,I_{N,i}}, \cdot) / dR(\xi_{N,I_{N,i}}, \cdot)(\tilde{\xi}_{N,i}) .$$

4: take $\{(\tilde{\xi}_{N,i}, \tilde{\omega}_{N,i})\}_{i=1}^{M_N}$ as an approximation of μ .

We may expect that the efficiency of Algorithm 1 depends highly on the choice of adjustment multiplier weights and proposal kernel. The former issue was dealt with by Douc *et al.* (2007b) (see also Olsson *et al.*, 2007) who identified adjustment multiplier weights for which the increase of asymptotic variance at a single iteration of the algorithm is minimal. In this article we focus on the latter issue and discuss strategies for adaptive designing of the proposal kernel. Unlike Douc *et al.* (2007b), we base our methods on the results of the next section describing the asymptotic KLD and CSD between the target and importance distributions of the auxiliary SMC algorithm.

4. THEORETICAL RESULTS

Consider the following assumptions.

(A1) The initial sample $\{(\xi_{N,i}, \omega_{N,i})\}_{i=1}^{M_N}$ is consistent for (ν, \mathbf{C}) .

(A2) There exists a function $\Psi : \Xi \rightarrow \mathbb{R}^+$ such that $\psi_{N,i} = \Psi(\xi_{N,i})$; moreover, $\Psi \in \mathbf{C} \cap \mathbf{L}^1(\Xi, \nu)$ and $L(\cdot, \tilde{\Xi}) \in \mathbf{C}$.

Under these assumptions we define for $(\xi, \tilde{\xi}) \in \Xi \times \tilde{\Xi}$ the weight function

$$\Phi(\xi, \tilde{\xi}) \triangleq \Psi^{-1}(\xi) \frac{dL(\xi, \cdot)}{dR(\xi, \cdot)}(\tilde{\xi}) . \quad (4.1)$$

The following result describes how the consistency property is passed through one step of the ASIS algorithm. A somewhat less general version of this result was also proved by Douc *et al.* (2007b, Theorem 3.1).

Proposition 4.1. Assume **(A1, A2)**. Then the weighted sample $\{(\tilde{\xi}_{N,i}, \tilde{\omega}_{N,i})\}_{i=1}^{M_N}$ is consistent for $(\nu, \tilde{\mathbf{C}})$, where $\tilde{\mathbf{C}} \triangleq \{f \in \mathbf{L}^1(\tilde{\Xi}, \mu), L(\cdot, |f|) \in \mathbf{C}\}$.

The result above is a direct consequence of Lemma A.1 and the fact that the set $\tilde{\mathbf{C}}$ is proper.

Let μ and ν be two probability measures in $\mathcal{P}(\mathbf{\Lambda})$ such that μ is absolutely continuous with respect to ν . We then recall that the KLD and the CSD are given, respectively, by

$$d_{\text{KL}}(\mu||\nu) \triangleq \int_{\mathbf{\Lambda}} \log[d\mu/d\nu(\lambda)] \mu(d\lambda) ,$$

$$d_{\chi^2}(\mu||\nu) \triangleq \int_{\mathbf{\Lambda}} [d\mu/d\nu(\lambda) - 1]^2 \nu(d\lambda) .$$

We will use the following quantities to compute *empirical* estimates of the KLD and CSD between μ_N^{aux} and π_N^{aux} . Indeed, define

$$\mathcal{E}(\{\tilde{\omega}_{N,i}\}_{i=1}^{\tilde{M}_N}) \triangleq \tilde{\Omega}_N^{-1} \sum_{i=1}^{\tilde{M}_N} \tilde{\omega}_{N,i} \log \left(\tilde{M}_N \tilde{\Omega}_N^{-1} \tilde{\omega}_{N,i} \right) ,$$

$$\text{CV}^2(\{\tilde{\omega}_{N,i}\}_{i=1}^{\tilde{M}_N}) \triangleq \tilde{M}_N \tilde{\Omega}_N^{-2} \sum_{i=1}^{\tilde{M}_N} \tilde{\omega}_{N,i}^2 - 1 .$$

The criterion \mathcal{E} is the negated Shannon entropy of the importance weights. The Shannon entropy is maximal when all the weights are equal and minimal when all of the weights are zero except one. The criterion CV^2 is the square of the coefficient of variation of the importance weights, which was suggested by Kong *et al.* (1994) as a means for detecting weight degeneracy. If all the weights are equal, then CV^2 is equal to zero. On the other hand, if all the weights are zero except for one, then the coefficient of variations is equal to $\tilde{M}_N - 1$ which is its maximum value.

The next theorem, which is the main result of this section and whose proof is found in the appendix shows that relates \mathcal{E} and CV^2 to two measures of the discrepancy between the proposal distribution and the target distribution, namely, the KLD for the \mathcal{E} criterion and the CSD for the CV^2 criterion.

Theorem 4.1. *Assume (A1, A2). Then the following holds as $N \rightarrow \infty$.*

i) *If $L(\cdot, |\log \Phi|) \in \mathbf{C} \cap \mathbf{L}^1(\mathbf{\Xi}, \nu)$, then*

$$\left| d_{\text{KL}}(\mu_N^{\text{aux}}||\pi_N^{\text{aux}}) - \mathcal{E}(\{\tilde{\omega}_{N,i}\}_{i=1}^{\tilde{M}_N}) \right| \xrightarrow{\mathbb{P}} 0 .$$

In addition,

$$d_{\text{KL}}(\mu_N^{\text{aux}}||\pi_N^{\text{aux}}) \xrightarrow{\mathbb{P}} \eta_{\text{KL}}(\Psi) \triangleq \nu \otimes L\{\log[\Phi\nu(\Psi)/\nu L(\tilde{\mathbf{\Xi}})]\}/\nu L(\tilde{\mathbf{\Xi}}) , \quad (4.2)$$

ii) *If $L(\cdot, \Phi) \in \mathbf{C}$, then*

$$\left| d_{\chi^2}(\mu_N^{\text{aux}}||\pi_N^{\text{aux}}) - \text{CV}^2(\{\tilde{\omega}_{N,i}\}_{i=1}^{\tilde{M}_N}) \right| \xrightarrow{\mathbb{P}} 0 .$$

In addition,

$$d_{\chi^2}(\mu_N^{\text{aux}}||\pi_N^{\text{aux}}) \xrightarrow{\mathbb{P}} \eta_{\chi^2}(\Psi) \triangleq \nu(\Psi) \nu \otimes L(\Phi)/[\nu L(\tilde{\mathbf{\Xi}})]^2 - 1 . \quad (4.3)$$

The expressions appearing in the RHS of Eqs. (4.2) and (4.3) may be interpreted as the KLD and CSD between probability distributions. Indeed, consider the following probability measures on the product space $(\Xi \times \tilde{\Xi}, \mathcal{B}(\Xi) \otimes \mathcal{B}(\tilde{\Xi}))$

$$\frac{\nu \otimes L}{\nu L(\tilde{\Xi})}(A) = \frac{\iint_{\Xi \times \tilde{\Xi}} \nu(d\xi) L(\xi, d\xi') \mathbb{1}_A(\xi, \xi')}{\iint_{\Xi \times \tilde{\Xi}} \nu(d\xi) L(\xi, d\xi')}, \quad (4.4)$$

$$\frac{\nu[\Psi] \otimes R}{\nu(\Psi)}(A) = \frac{\iint_{\Xi \times \tilde{\Xi}} \nu(d\xi) \Psi(\xi) R(\xi, d\xi') \mathbb{1}_A(\xi, \xi')}{\iint_{\Xi \times \tilde{\Xi}} \nu(d\xi) \Psi(\xi) R(\xi, d\xi')}, \quad (4.5)$$

where $A \in \mathcal{B}(\Xi) \otimes \mathcal{B}(\tilde{\Xi})$ and the outer product \otimes of a measure and a kernel is defined in (2.2).

Proposition 4.2. *Assume (A1, A2). Then the following holds.*

- i) *The limiting Kullback-Leibler divergence in (4.2) is the KLD between the probability distributions $\nu \otimes L/\nu L(\tilde{\Xi})$ and $\nu[\Psi] \otimes \nu(\Psi)/\nu(\Psi)$, that is,*

$$\eta_{\text{KL}}(\Psi) = d_{\text{KL}} \left(\frac{\nu \otimes L}{\nu L(\tilde{\Xi})} \parallel \frac{\nu[\Psi] \otimes R}{\nu(\Psi)} \right).$$

- ii) *The limiting chi-square distance in (4.3) is the CSD between the probability distributions $\nu \otimes L/\nu L(\tilde{\Xi})$ and $\nu[\Psi] \otimes \nu(\Psi)/\nu(\Psi)$, that is,*

$$\eta_{\chi^2}(\Psi) = d_{\chi^2} \left(\frac{\nu \otimes L}{\nu L(\tilde{\Xi})} \parallel \frac{\nu[\Psi] \otimes R}{\nu(\Psi)} \right).$$

The expressions of the limiting measures in Proposition 4.2 then allow for deriving the adjustment multipliers weights Ψ which minimize the corresponding discrepancy measures. The next corollary provides the adjustment multiplier weight function minimizing, for a given proposal kernel R , the asymptotic KLD and CSD. In the KLD case, the best that we can do is to choose Ψ such that the two marginal distributions $\frac{\nu \otimes L}{\nu L(\tilde{\Xi})}(\cdot \times \tilde{\Xi})$ and $\frac{\nu[\Psi] \otimes R}{\nu(\Psi)}(\cdot \times \tilde{\Xi})$ are equal on the space $(\Xi, \mathcal{B}(\Xi))$. This is a direct consequence of the so-called *chain rule for entropy* (see Cover and Thomas, 1991, Theorem 2.2.1). The optimization is slightly less obvious for the chi-square distance, yet an explicit expression for the adjustment multiplier weights can be found in this case. Again the proof is found in the appendix.

Corollary 4.1. *Assume (A1, A2). Then the following holds.*

- i) *If $L(\cdot, |\log \Phi|) \in \mathcal{C} \cap \mathcal{L}^1(\Xi, \nu)$, then $\arg \min_{\Psi} \eta_{\chi^2}(\Psi) = \Psi_{\text{KL}, R}^*$, where*

$$\Psi_{\text{KL}, R}^*(\xi) \triangleq L(\xi, \tilde{\Xi}).$$

- ii) *If $L(\cdot, \Phi) \in \mathcal{C}$, then $\arg \min_{\Psi} \eta_{\chi^2}(\Psi) = \Psi_{\chi^2, R}^*$, where*

$$\Psi_{\chi^2, R}^*(\xi) \triangleq \sqrt{\int_{\tilde{\Xi}} \frac{dL(\xi, \cdot)}{dR(\xi, \cdot)}(\tilde{\xi}) L(\xi, d\tilde{\xi})}.$$

It is worthwhile to notice that the optimal adjustment weights for the KLD do not depend on the proposal kernel R . The minimal value $\eta_{\text{KL}}(\Psi_{\text{KL},R}^*)$ of the limiting KLD is the conditional relative entropy between $\nu \otimes L/\nu L(\tilde{\Xi})$ and $\nu[\Psi] \otimes \nu(\Psi)/\nu(\Psi)$.

In both cases, letting $R(\cdot, A) = L(\cdot, A)/L(\cdot, \tilde{\Xi})$ yields, as we may expect, the chi-square optimal adjustment multiplier weight function $\Psi_{\text{KL},R}^*(\cdot) = \Psi_{\chi^2,R}^*(\cdot, \tilde{\Xi}) = L(\cdot, \tilde{\Xi})$, resulting in uniform importance weights $\tilde{\omega}_{N,i} \equiv 1$.

5. ADAPTIVE IMPORTANCE SAMPLING

5.1. ASIS adaptation by minimization of estimated KLD and CSD over a parametric family. In the light of Theorem 4.1, a natural strategy for adaptive design of π_N^{aux} is to minimize the empirical estimate \mathcal{E} (or CV^2) of the KLD (or CSD) under consideration over all proposal kernels belonging to some parametric family $\{R_\theta\}_{\theta \in \Theta}$. Thus, assume that there exists a random noise variable ϵ , having distribution λ on some measurable space $(\Lambda, \mathcal{B}(\Lambda))$, and a family $\{F_\theta\}_{\theta \in \Theta}$ of mappings from $\Xi \times \Lambda$ to $\tilde{\Xi}$ such that we are able to simulate $\tilde{\xi} \sim R_\theta(\xi, \cdot)$, for $\xi \in \Xi$, by simulating $\epsilon \sim \lambda$ and letting $\tilde{\xi} = F_\theta(\xi, \epsilon)$. We denote by Φ_θ the importance weight function associated with R_θ and set $\Phi_\theta \circ F_\theta(\xi, \epsilon) \triangleq \Phi_\theta(\xi, F_\theta(\xi, \epsilon))$.

In this setting, assume that **(A1)** holds and suppose that we have simulated, as in the first step of Algorithm 1, indices $\{I_{N,i}\}_{i=1}^{\tilde{M}_N}$ and noise variables $\{\epsilon_{N,i}\}_{i=1}^{\tilde{M}_N} \sim \lambda^{\otimes \tilde{M}_N}$. Now, keeping these indices and noise variables fixed, we can form an idea of how the KLD varies with θ via the mapping $\theta \mapsto \mathcal{E}(\{\Phi_\theta \circ F_\theta(\xi_{N,I_{N,i}}, \epsilon_{N,i})\}_{i=1}^{\tilde{M}_N})$. Similarly, the CSD can be studied by using CV^2 instead of \mathcal{E} . This suggests an algorithm in which the particles are repropagated using R_{θ_*} , where $\theta_* = \arg \min_{\theta \in \Theta} \mathcal{E}(\{\Phi_\theta \circ F_\theta(\xi_{N,I_{N,i}}, \epsilon_{N,i})\}_{i=1}^{\tilde{M}_N})$. The minimum θ_* exists if, e.g., the parameter space Θ is compact and the mapping $\theta \mapsto \Phi_\theta \circ F_\theta(\xi, \epsilon)$ is continuous for all (ξ, ϵ) , or when Θ is finite. The algorithm is summarized below, and its modification for minimization of the empirical CSD is straightforward.

Algorithm 2 Adaptive ASIS

Require: (A1)

- 1: Draw $\{I_{N,i}\}_{i=1}^{\tilde{M}_N} \sim \mathcal{M}(\tilde{M}_N, \{\omega_{N,j} \psi_{N,j} / \sum_{\ell=1}^{\tilde{M}_N} \omega_{N,\ell} \psi_{N,\ell}\}_{j=1}^{\tilde{M}_N})$,
 - 2: simulate $\{\epsilon_{N,i}\}_{i=1}^{\tilde{M}_N} \sim \lambda^{\otimes \tilde{M}_N}$,
 - 3: $\theta_* \leftarrow \arg \min_{\theta \in \Theta} \mathcal{E}(\{\Phi_\theta \circ F_\theta(\xi_{N,I_{N,i}}, \epsilon_{N,i})\}_{i=1}^{\tilde{M}_N})$,
 - 4: set $\tilde{\xi}_{N,i} \stackrel{\forall i}{\leftarrow} F_{\theta_*}(\xi_{N,I_{N,i}}, \epsilon_{N,i})$ and $\tilde{\omega}_{N,i} \stackrel{\forall i}{\leftarrow} \Phi_{\theta_*}(\xi_{N,I_{N,i}}, \tilde{\xi}_{N,i})$,
 - 5: let $\{(\tilde{\xi}_{N,i}, \tilde{\omega}_{N,i})\}_{i=1}^{\tilde{M}_N}$ approximate μ .
-

Remark 5.1. *In order to correctly deal with particles having null weights when implementing the empirical estimate of the KLD in practice, one should ensure that the convention $0 \log 0 = 0$ is supported by the compiler or enforced by the user.*

Remark 5.2. *A slight modification of Algorithm 2, lowering the added computational burden, is to apply the adaptation mechanism only when the estimated KLD (or CSD) is above a chosen threshold κ , that is replacing*

$$\theta_* \leftarrow \arg \min_{\theta \in \Theta} \mathcal{E}(\{\Phi_\theta \circ F_\theta(\xi_{N,I_{N,i}}, \epsilon_{N,i})\}_{i=1}^{\tilde{M}_N})$$

in Algorithm 2 by

if $\mathcal{E}(\{\Phi_{\theta_0} \circ F_{\theta_0}(\xi_{N,I_{N,i}}, \epsilon_{N,i})\}_{i=1}^{\tilde{M}_N}) \geq \kappa$ **then**

$$\theta_* \leftarrow \arg \min_{\theta \in \Theta} \mathcal{E}(\{\Phi_\theta \circ F_\theta(\xi_{N,I_{N,i}}, \epsilon_{N,i})\}_{i=1}^{\tilde{M}_N}),$$

else

$$\theta_* \leftarrow \theta_0,$$

end if

where the parameter $\theta_0 \in \Theta$ corresponds to the proposal kernel used when the adaptation operation is put on standby. Here the threshold value κ may be designed by, e.g., estimating prefatorily the KLDs over a relatively long simulated observation record and relating these empirical values to the ratios between the MSE values of the nonadaptive filter and those of the adaptive filter. More specifically, say that we want the adaptation mechanism to be automatically activated when this may decrease the error by a factor α ; we then identify all time steps in the calibration record for which the MSE ratios exceed α and let κ take the minimal value among the corresponding empirical KLD estimates.

Remark 5.3. *It is possible to establish a law of large numbers as well as a central limit theorem for the algorithm above, similarly to what has been done for the nonadaptive auxiliary particle filter by Douc et al. (2007b) and Olsson et al. (2007). More specifically, suppose that $\tilde{M}_N/M_N \rightarrow \ell \in [0, \infty]$ as $N \rightarrow \infty$ and that the initial sample $\{(\xi_{N,i}, \omega_{N,i})\}_{i=1}^{M_N}$ satisfies, for all f belonging to a given class \mathbf{A} of functions, the central limit theorem*

$$a_N \Omega_N^{-1} \sum_{i=1}^{M_N} \omega_{N,i} [f(\xi_{N,i}) - \mu(f)] \xrightarrow{\mathcal{D}} \mathcal{N}[0, \sigma^2(f)],$$

where the sequence $\{a_N\}_N$ is such that $a_N M_N \rightarrow \beta \in [0, \infty)$ as $N \rightarrow \infty$ and $\sigma : \mathbf{A} \rightarrow \mathbb{R}^+$ is a functional. Then the sample $\{(\tilde{\xi}_{N,i}, \tilde{\omega}_{N,i})\}_{i=1}^{\tilde{M}_N}$ produced in Algorithm 2 is asymptotically normal for a class of functions $\tilde{\mathbf{A}}$ in the sense that, for all $f \in \tilde{\mathbf{A}}$,

$$\tilde{\Omega}_N^{-1} \sum_{i=1}^{\tilde{M}_N} \tilde{\omega}_{N,i} [f(\tilde{\xi}_{N,i}) - \mu(f)] \xrightarrow{\mathcal{D}} \mathcal{N}[0, \tilde{\sigma}^2(f)],$$

where

$$\tilde{\sigma}^2(f) \triangleq \beta \ell^{-1} \nu(\Psi R_{\theta_*} \{\cdot, \Phi_{\theta_*}^2[f - \mu(f)]\}) \nu(\Psi) / [\nu L(\tilde{\Xi})]^2 + \sigma^2(L\{\cdot, [f - \mu(f)]\}) / [\nu L(\tilde{\Xi})]^2$$

and θ_* minimizes the asymptotic KLD. The complete proof of this result is however omitted for brevity.

5.2. ASIS adaptation by means of population Monte Carlo methods. Another route to effective adaptive particle proposition goes via the *D-kernel population Monte Carlo* (DPMC) *algorithm* presented in (Douc *et al.*, 2007a) (and being a refinement of the basic population Monte Carlo algorithm introduced by Cappé *et al.*, 2004) in which the weights of a proposal mixture comprising D stratas are adapted towards a better fit with the target distribution. Applying the DPMC method in our context will yield an algorithm in which, as soon as the empirical KLD is indicating a large discrepancy between μ_N^{aux} and π_N^{aux} , each proposed pair $(I_{N,i}, \tilde{\xi}_{N,i})$ is repropoed according to $\sum_{d=1}^D \alpha_d Q_d((I_{N,i}, \tilde{\xi}_{N,i}), \cdot)$, that is, a mixture of Markovian kernels $\{Q_d\}_{d=1}^D$ on the product space $\{1, \dots, M_N\} \times \tilde{\Xi}$. In practice (cf. Algorithm 3) only the particle components are moved, while the index transitions follow the identity kernel. After this, the updated pairs are resampled multinomially with respect to weights proportional to the Radon-Nikodym derivative $d\mu_N^{\text{aux}}/d[\sum_{d=1}^D \alpha_d Q_d((I_{N,i}, \tilde{\xi}_{N,i}), \cdot)]$ evaluated at the pairs. The procedure is repeated, say, T times and at each step the mixture weights $\{\alpha_d\}_{d=1}^D$ are updated to favour kernels Q_d associated with a high acceptance probability at the preceding DPMC resampling operation. More specifically, we set α_d proportional to the survival rate of the corresponding Q_d . There is nothing that prevents us from using a number m_N of index-particle pairs within the DPMC adaptation loop which is larger than \tilde{M}_N . Referring to the theoretical results obtained by Douc *et al.* (2007a), we may expect (a proof is in progress) that this scheme, as the number of DPMC samples and iterations increases, provides us with a set of mixture weights minimizing the KLD between the measures $\mu_N^{\text{aux}} \otimes \mu_N^{\text{aux}}$ and $\mu_N^{\text{aux}} \otimes \sum_{d=1}^D \alpha_d Q_d$. The algorithmic details are presented in the scheme below, in which we assume that the indices evolve according to the identity kernel within the DPMC adaptation loop. We also assume that L and the Q_d 's (which here describe the DPMC transitions of the particle positions only) admit densities (denoted by the same symbols) with respect to a common reference measure on $\tilde{\Xi}$.

Remark 5.4. *Similarly to what is pointed out in Remark 5.2, the additional computational work implied by the DPMC adaptation can be lightened by entering the DPMC loop (lines 5–13 in the scheme above) only when the empirical KLD exceeds a chosen threshold κ .*

6. APPLICATION TO STATE SPACE MODELS

For an illustration, we apply our methods to *optimal filtering* in *state space models* of type

$$\begin{aligned} X_{k+1} &= f_k(X_k, W_{k+1}), \quad k \geq 0, \\ Y_k &= h_k(X_k, V_k), \quad k \geq 0, \end{aligned}$$

where $\{f_k\}_{k=0}^\infty$ and $\{h_k\}_{k=0}^\infty$ are sets of known \mathbb{R} -valued functions, $\{W_k\}_{k=1}^\infty$ and $\{V_k\}_{k=0}^\infty$ are mutually independent sets of standard normal-distributed variables such that W_{k+1} is independent of $\{(X_i, Y_i)\}_{i=0}^k$ and V_k is independent of X_k and $\{(X_i, Y_i)\}_{i=0}^{k-1}$. In this setting, we wish to approximate the *filter distributions*, that is, the posterior distributions $\phi_k(\cdot) \triangleq \mathbb{P}(X_k \in \cdot | Y_0, \dots, Y_k)$, $k \geq 0$, which in general lack closed form expressions. By the *filtering*

Algorithm 3 AIS with PMC-adaptation**Require: (A1)**

- 1: Run steps 1 to 3 of Algorithm 1,
- 2: **if** $\mathcal{E}(\{\tilde{\omega}_{N,i}\}_{i=1}^{\tilde{M}_N}) \geq \kappa$ **then**
- 3: set $\alpha_d \stackrel{\forall d}{\leftarrow} 1/D$,
- 4: draw $\{J_{N,i}\}_{i=1}^{m_N} \sim \mathcal{M}(m_N, \{\tilde{\omega}_{N,j} \tilde{\Omega}_N^{-1}\}_{j=1}^{\tilde{M}_N})$,
- 5: **for** $t = 1, \dots, T$ **do**
- 6: draw $\{K_{N,i}\}_{i=1}^{m_N} \sim \mathcal{M}(m_N, \{\alpha_d\}_{d=1}^D)$,
- 7: draw $\{\tilde{\xi}'_{N,i}\}_{i=1}^{m_N} \sim \bigotimes_{i=1}^{m_N} Q_{K_{N,i}}(\tilde{\xi}_{N,J_{N,i}}, \cdot)$,
- 8: set $I_{N,i} \stackrel{\forall i}{\leftarrow} I_{N,J_{N,i}}$,
- 9: update

$$\tilde{\omega}'_{N,i} \stackrel{\forall i}{\leftarrow} \omega_{N,I_{N,i}} L(\xi_{N,I_{N,i}}, \tilde{\xi}'_{N,i}) / \sum_{d=1}^D \alpha_d Q_d(\tilde{\xi}_{N,J_{N,i}}, \tilde{\xi}'_{N,i}),$$

- 10: set $\tilde{\xi}_{N,i} \stackrel{\forall i}{\leftarrow} \tilde{\xi}'_{N,i}$,
- 11: update

$$\alpha_d \stackrel{\forall d}{\leftarrow} \tilde{\Omega}_N^{-1} \sum_{i=1}^{m_N} \tilde{\omega}'_{N,i} \mathbb{1}_{K_{N,i}}(d),$$

- 12: draw $\{J_{N,i}\}_{i=1}^{m_N} \sim \mathcal{M}(m_N, \{\tilde{\omega}'_{N,j} \tilde{\Omega}_N^{-1}\}_{j=1}^{m_N})$,
- 13: **end for**
- 14: let $\{(\tilde{\xi}'_{N,i}, \tilde{\omega}'_{N,i})\}_{i=1}^{m_N}$ approximate μ ,
- 15: **else**
- 16: let $\{(\tilde{\xi}_{N,i}, \tilde{\omega}_{N,i})\}_{i=1}^{\tilde{M}_N}$ approximate μ .
- 17: **end if**

recursion it holds that

$$\phi_{k+1}(A) = \frac{\iint_{A \times \mathbb{R}} g(x_{k+1}, Y_{k+1}) Q(x_k, dx_{k+1}) \phi_k(dx_k)}{\iint_{\mathbb{R} \times \mathbb{R}} g(x_{k+1}, Y_{k+1}) Q(x_k, dx_{k+1}) \phi_k(dx_k)}, \quad (6.1)$$

where Q is the transition kernel of the *unobservable* chain $\{X_k\}_{k=0}^{\infty}$ and $g(x, \cdot)$ is the density of the distribution $\mathbb{P}(Y_k \in \cdot | X_k = x)$, that is, the distribution of the *observation* Y_k given the hidden state $X_k = x$. From (6.1) we conclude that this filtering problem can, with $\Xi = \tilde{\Xi} = \mathbb{R}$, $\nu = \phi_k$, $\mu = \phi_{k+1}$, and $L_k(x, A) = \int_A g(x', Y_{k+1}) Q(x, dx')$, be perfectly cast into the framework of Section 3, rendering sequential particle approximation of the filter measures possible. At each time step we propose new particles according to the dynamics of the hidden chain, that is, $R_k = Q$.

6.1. Nonlinear autoregressive model observed in noise. As a first example we consider a first order (possibly nonlinear) autoregression observed in noise, for which $f_k(x, w) =$

$m(x) + \sigma_w(x)w$ and $h_k(x, v) = x + \sigma_v v$. Note that the bivariate Markov chain $\{(X_k, Y_k)\}_{k=0}^\infty$ is time homogenous since neither f_k nor h_k vary with k in this case. For a model of this type the optimal adjustment multiplier weight function $\Psi_{\chi^2, Q}^*$ can be expressed on closed form:

$$\Psi_{\chi^2, Q}^*(x) \propto \sqrt{\frac{2\sigma_v^2}{2\sigma_w^2(x) + \sigma_v^2}} \exp\left(-\frac{Y_{k+1}^2}{\sigma_v^2} + \frac{m(x)}{2\sigma_w^2(x) + \sigma_v^2}[2Y_{k+1} - m(x)]\right) \quad (6.2)$$

We will study the following two special cases of the model in question:

- $m_k(x) \equiv 0$, $\sigma_w(x) = \sqrt{\beta_0 + \beta_1 x^2}$

This is the classical Gaussian autoregressive conditional heteroscedasticity (ARCH) model observed in noise (see Bollerslev *et al.*, 1994). In this case an experiment was conducted where we compared: (i) a plain nonadaptive particle filter for which $\Psi \equiv 1$, that is, the bootstrap particle filter of Gordon *et al.* (1993), (ii) a nonadaptive auxiliary filter based on the chi-square optimal weights $\Psi_{\chi^2, Q}^*$, and adaptive bootstrap filters with uniform adjustment multiplier weights using direct minimization of (iii) the empirical CSD and (iv) the empirical KLD (Algorithm 2). Here the CV^2 and \mathcal{E} were minimized over the family of kernels obtained by scaling the variance of Q , so that $\xi_{N,i}^{k+1} = F_\theta(\xi_{N,I_{N,i}^k}^k, W_{k+1})$, with $F_\theta(\xi, w) = \theta\sqrt{\beta_0 + \beta_1 \xi^2}w$ and $\theta \in \Theta \triangleq (0, 8]$. This experiment was repeated for impetuously varying as well as close to constant volatilities of the hidden process, corresponding to the parameter vectors $(\beta_0, \beta_1, \sigma_v^2) = (0.1, 2, 0.1)$ and $(\beta_0, \beta_1, \sigma_v^2) = (0.5, 0.1, 0.25)$, respectively. In order to design a challenging test of the adaptation procedures we replaced the simulated observations at $k = 7$ by outliers located at distances $50\beta_0$ and $6\beta_0$ from the mean level (zero) of $\{X_k\}_{k=0}^\infty$ in the two cases, respectively. For these observation records, each particle filter approximated 400 filter means using 5,000 particles. The mean square errors in the figures are based on reference values obtained using a bootstrap particle filter with as many as 500,000 particles. The outcome is plotted in Figures 1(a) and 1(b), from which it is evident that the two adaptive particle filters outperform the other two in the case of a slowly varying volatility. In this case the optimal weight function in (6.2) is almost constant at all time steps, making the two nonadaptive filters close to equivalent. In the case of impetuously varying observations, the auxiliary filter based on optimal weights perform almost as well as the adaptive filters. It is also to be noted that the adaptive filters based on CV^2 and \mathcal{E} perform equivalently. The values, presented in Table 1, of the optimal parameters θ_* associated with these filters at different time steps show that the adaptively chosen proposal kernels are close to identical for the two approaches.

- $m(x) = \varphi x$, $\sigma_w(x) \equiv \sigma$

TABLE 1. Mean values of the optimal parameter θ_* over time for the CSD-based as well as the KLD-based versions of Algorithm 2. The model under consideration was the ARCH model with informative observations and each value is based on 400 runs of each filter. The corresponding MSE are plotted in Figure 1(b)

Adaptation method	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
CSD-based	1.49	0.59	1.18	0.92	1.50
KLD-based	1.46	0.57	1.13	0.87	1.46
Adaptation method	$k = 6$	$k = 7$	$k = 8$	$k = 9$	$k = 10$
CSD-based	0.81	2.90	0.67	0.70	1.74
KLD-based	0.76	2.85	0.62	0.66	1.69

A similar trial was conducted also for a standard linear/Gaussian model, and this time we involved (i) a standard bootstrap filter, adaptive bootstrap filters minimizing (ii) the empirical CSD and (iii) the empirical KLD, and (iv) a bootstrap filter using DPMC adaptation. For this model, exact posterior filter means can be obtained using the *Kalman filter*, and these exact values were used as reference when computing mean square errors. For the DPMC-based adaptive filter, we let the Q_d 's be *independence samplers* located at grid points $m_d = -3 + 0.1d$, where $1 \leq d \leq 60$; more specifically, we let $Q_d(x, x') = \exp(-(x' - m_d)^2/0.2)/\sqrt{0.1\pi}$. The number of DPMC adaptation iterations was $T = 30$. Also in this example the CV^2 and \mathcal{E} were minimized over the family of kernels obtained by scaling the variance of the prior kernel Q . We set the model parameters to $(\varphi, \sigma_v^2, \sigma_w^2) = (0.9, 0.01, 0.1)$, yielding informative observations, and the observation record under consideration was $(Y_0, \dots, Y_4) = (0.69 \ 0.39 \ 0.34 \ 3 \ 0.54)$; here an outlier was placed at $k = 3$. The result is displayed in Fig 1(c) from which it is clear that the three adaptation strategies lead to drastic improvements of the SMC filter mean estimates vis-à-vis the bootstrap filter. Again, the mean square errors were based on 400 runs of each filter (using still 5,000 particles).

6.2. Growth model. We now turn to the univariate *growth model*, discussed by Kitagawa (1987), where the observations are nonlinear in the hidden state. More specifically, $h_k(x, v) = bx^2 + \sigma_v v$ and $f_k(x, w) = a_k(x) + \sigma_w w$, with

$$a_k(x) = \alpha_0 x + \alpha_1 \frac{x}{1+x^2} + \alpha_2 \cos(1.2k), \quad k \geq 0,$$

where $(\alpha_0, \alpha_1, \alpha_2, b, \sigma_w^2, \sigma_v^2) = (0.5, 25, 8, 0.05, 1, 1)$, corresponding to the non-informative case discussed in (Cappé *et al.*, 2005, pp. 230–231). The initial state is known and set to $X_0 = 0.1$. For this model the optimal weights given in Corollary 4.1 lack closed form expressions. The filters run were same as in the previous example. The CV^2 and \mathcal{E} were once again

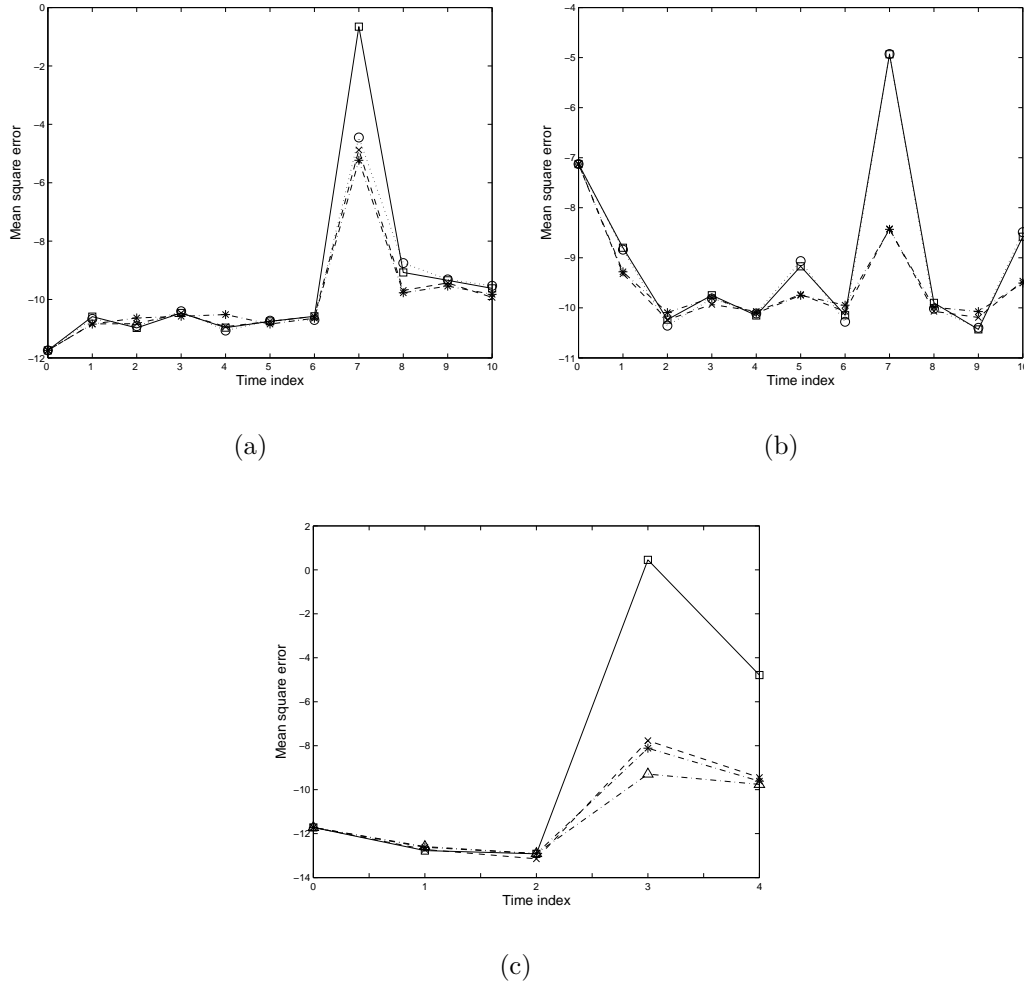


FIGURE 1. Plot of MSE performances (on log-scale) of the bootstrap filter (\square), the auxiliary filter based on optimal weights (\circ), adaptive filters minimizing the empirical KLD ($*$) and CSD (\times), and DPMC-based adaptive filter (Δ). In figures (a) and (b) the ARCH model is considered, while the model in (c) is linear/Gaussian. The MSE values are computed using 5,000 particles and 400 runs for each algorithm.

minimized by scaling the variance of Q , so that $F_{k,\theta}(\xi, w) = a_k(\xi) + \theta\sigma_w w$ for all $k \geq 0$ and $\theta \in \Theta \triangleq (0, 6]$. The DPMC kernels were similar to those used for the linear/Gaussian model, that is, $Q_d(x, x') = \exp(-(x' - m_d)^2) / \sqrt{2\pi}$ with $m_d = 0.5(d-1)$ and $1 \leq d \leq 40$. The number of DPMC adaptation iterations was still $T = 30$. For an observation record containing the outlying observation $Y_7 = bX_7^2 + 10\sigma_v$, the MSEs were estimated over 400 runs of the filters, each filter using 5,000 particles. As displayed in Fig 2, the filters minimizing directly the

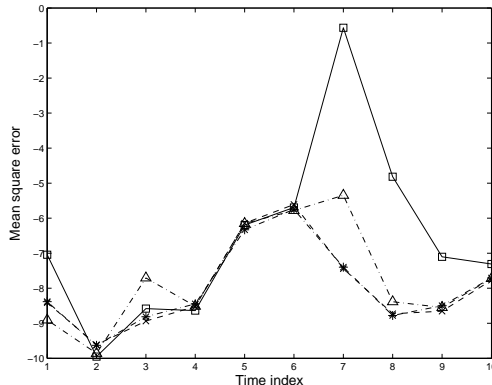


FIGURE 2. Plot of MSE performances (on log-scale) of the bootstrap filter (\square), adaptive filters minimizing the empirical KLD ($*$) as well as CSD (\times), and DPMC-based adaptive filter (\triangle) for the growth model in the presence of an outlying observation at time 7.

empirical CSD and KLD robustify equally well the filter to the outlying observation. Also the DPMC-based filter exhibits a similar performance, even though its efficiency may benefit from a more sensible choice of D -kernel mixture.

Of special interest in practice are the cases where the state space dynamics (Q, g) used for filtering differ from that generating the input observations. We thus tested the adaptive algorithms on the growth model where the observations were simulated under $(\alpha_0, \alpha_1, \alpha_2, b, \sigma_w^2, \sigma_v^2) = (0.5, 25, 8, 0.05, 1, 1)$, but where the evolution of the particles followed the dynamics determined by the vector $(\alpha_0, \alpha_1, \alpha_2, b, \sigma_w^2, \sigma_v^2) = (5, 25, 8, 0.1, 1, 1)$, that is, with heavily perturbed leading parameters of the state and observation equations. As displayed in Figure 3, the general improvement obtained by optimization over the simple parametric family used in the previous is blatant, reducing the MSE by up a factor 158 (at time step $k = 3$). The behaviour at the extreme time step $k = 6$ can be explained by the rather limited selection of proposal kernels offered by the parametric family in this case; actually, it does not provide any kernel being better suited for this extreme situation than the prior kernel Q . This is also confirmed by the values of the optimal adjustment parameters θ_* figured in Table 2, which are, on an average, close to one at this time step. However, Algorithm 2 is utterly generic and allows for significantly more sophisticated families of kernels. The study of this aspect is however beyond the scope of this paper and will be the subject of future work.

APPENDIX A. PROOFS

A.1. **Proof of Theorem 4.1.** We first prove an introductory lemma.

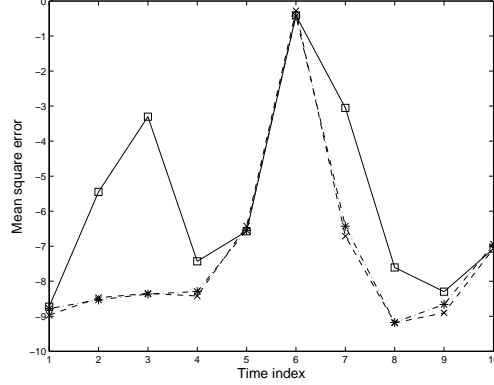


FIGURE 3. Plot of MSE performances (on log-scale) of the bootstrap filter (\square) and adaptive filters minimizing the empirical KLD ($*$) as well as CSD (\times) for input observations generated under pertubated parameters in the growth model. The MSE values are computed using 5,000 particles and 200 runs of each algorithm.

TABLE 2. Mean values of the optimal parameter θ_* over time for the CSD-based as well as the KLD-based versions of Algorithm 2. The model under consideration was the Growth model and each value is based on 200 runs of each filter, corresponding to Figure 3.

Adaptation method	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
CSD-based	1.13	2.79	3.19	1.04	1.10
KLD-based	1.08	2.84	3.21	1.03	1.15
Adaptation method	$k = 6$	$k = 7$	$k = 8$	$k = 9$	$k = 10$
CSD-based	1.03	3.43	1.92	1.22	1.01
KLD-based	1.07	3.36	1.93	1.15	1.00

Lemma A.1. Assume **(A1, A2)** and let $C^* \triangleq \{f \in \mathbb{B}(\Xi \times \tilde{\Xi}) : L(\cdot, |f|) \in \mathcal{C} \cap \mathcal{L}^1(\Xi, \nu)\}$. Then, for all $f \in C^*$, as $N \rightarrow \infty$,

$$\tilde{\Omega}_N^{-1} \sum_{i=1}^{\tilde{M}_N} \tilde{\omega}_{N,i} f(\xi_{N,I_{N,i}}, \tilde{\xi}_{N,i}) \xrightarrow{\mathbb{P}} \nu \otimes L(f) / \nu L(\tilde{\Xi})$$

Proof. It is enough to prove that

$$\tilde{M}_N^{-1} \sum_{i=1}^{\tilde{M}_N} \tilde{\omega}_{N,i} f(\xi_{N,I_{N,i}}, \tilde{\xi}_{N,i}) \xrightarrow{\mathbb{P}} \nu \otimes L(f) / \nu(\Psi). \quad (\text{A.1})$$

for all $f \in \mathbf{C}^*$; indeed, since the function $f \equiv 1$ belongs to \mathbf{C}^* under **(A2)**, the result of the lemma will follow from (A.1) by Slutsky's theorem. Define the measure $\varphi(A) \triangleq \nu(\Psi \mathbb{1}_A) / \nu(\Psi)$, with $A \in \mathcal{B}(\Xi)$. By applying Theorem 1 in Douc and Moulines (2005) we conclude that the weighted sample $\{(\xi_{N,i}, \psi_{N,i})\}_{i=1}^{\tilde{M}_N}$ is consistent for $(\varphi, \{f \in \mathbf{L}^1(\Xi, \varphi) : \Psi|f| \in \mathbf{C}\})$. Moreover, by Theorem 2 in the same paper this is also true for the uniformly weighted sample $\{(\xi_{N,i}, 1)\}_{i=1}^{\tilde{M}_N}$ (see the proof of Theorem 3.1 in Douc *et al.* (2007b) for details). By definition, for $f \in \mathbf{C}^*$, $\varphi \otimes R(\Phi|f|) \nu(\Psi) = \nu \otimes L(|f|) < \infty$ and $\Psi R(\cdot, \Phi|f|) = L(\cdot, |f|) \in \mathbf{C}$. Hence, we conclude that $R(\cdot, \Phi|f|)$ and thus $R(\cdot, \Phi f)$ belong to the proper set $\{f \in \mathbf{L}^1(\Xi, \varphi) : \Psi|f| \in \mathbf{C}\}$. This implies the convergence

$$\begin{aligned} \tilde{M}_N^{-1} \sum_{i=1}^{\tilde{M}_N} \mathbb{E} \left[\tilde{\omega}_{N,i} f(\xi_{N,i}, \tilde{\xi}_{N,i}) \middle| \mathcal{F}_N \right] \\ = \tilde{M}_N^{-1} \sum_{i=1}^{\tilde{M}_N} R(\xi_{N,i}, \Phi f) \xrightarrow{\mathbb{P}} \varphi \otimes R(\Phi f) = \nu \otimes L(f) / \nu(\Psi), \end{aligned} \quad (\text{A.2})$$

where $\mathcal{F}_N \triangleq \sigma(\{\xi_{N,i}, \tilde{\xi}_{N,i}\}_{i=1}^{\tilde{M}_N})$ denotes the σ -algebra generated by the selected particles. It thus suffices to establish that

$$\tilde{M}_N^{-1} \sum_{i=1}^{\tilde{M}_N} \left\{ \mathbb{E} \left[\tilde{\omega}_{N,i} f(\xi_{N,i}, \tilde{\xi}_{N,i}) \middle| \mathcal{F}_N \right] - \tilde{\omega}_{N,i} f(\xi_{N,i}, \tilde{\xi}_{N,i}) \right\} \xrightarrow{\mathbb{P}} 0, \quad (\text{A.3})$$

and we do this, following the lines of the proof of Theorem 1 in Douc and Moulines (2005), by verifying the two conditions of Theorem 11 in the same work. We first conclude that the sequence $\{\tilde{M}_N^{-1} \sum_{i=1}^{\tilde{M}_N} \mathbb{E}[\tilde{\omega}_{N,i} |f(\xi_{N,i}, \tilde{\xi}_{N,i})| \middle| \mathcal{F}_N]\}_N$ is tight since it tends to $\nu \otimes L(|f|) / \nu(\Psi)$ in probability (cf. (A.2)). Thus, the first condition is satisfied. To verify the second condition, take $\epsilon > 0$ and consider, for any $C > 0$, the decomposition

$$\begin{aligned} \tilde{M}_N^{-1} \sum_{i=1}^{\tilde{M}_N} \mathbb{E} \left[\tilde{\omega}_{N,i} |f(\xi_{N,i}, \tilde{\xi}_{N,i})| \mathbb{1}_{\{\tilde{\omega}_{N,i} |f(\xi_{N,i}, \tilde{\xi}_{N,i})| \geq \epsilon\}} \middle| \mathcal{F}_N \right] \\ \leq \tilde{M}_N^{-1} \sum_{i=1}^{\tilde{M}_N} R(\xi_{N,i}, \Phi|f| \mathbb{1}_{\{\Phi|f| \geq C\}}) + \mathbb{1}_{\{\epsilon \tilde{M}_N < C\}} \tilde{M}_N^{-1} \sum_{i=1}^{\tilde{M}_N} \mathbb{E} \left[\tilde{\omega}_{N,i} |f(\xi_{N,i}, \tilde{\xi}_{N,i})| \middle| \mathcal{F}_N \right]. \end{aligned}$$

Since $R(\cdot, \Phi f)$ belongs to the proper set $\{f \in \mathbf{L}^1(\Xi, \varphi) : \Psi|f| \in \mathbf{C}\}$, so does the function $R(\cdot, \Phi|f| \mathbb{1}_{\{\Phi|f| \geq C\}})$. Thus, since the indicator $\mathbb{1}_{\{\epsilon \tilde{M}_N < C\}}$ tends to zero, we conclude that the upper bound above has the limit $\varphi \otimes R(\Phi|f| \mathbb{1}_{\{\Phi|f| \geq C\}})$; however, by dominated convergence this limit can be made arbitrarily small by increasing C . Hence

$$\tilde{M}_N^{-1} \sum_{i=1}^{\tilde{M}_N} \mathbb{E} \left[\tilde{\omega}_{N,i} |f(\xi_{N,i}, \tilde{\xi}_{N,i})| \mathbb{1}_{\{\tilde{\omega}_{N,i} |f(\xi_{N,i}, \tilde{\xi}_{N,i})| \geq \epsilon\}} \middle| \mathcal{F}_N \right] \xrightarrow{\mathbb{P}} 0,$$

which verifies the second condition of Theorem 11 in Douc and Moulines (2005). Thus, (A.3) follows. \square

Proof of Theorem 4.1. We start with i). In order to establish the limit

$$d_{\text{KL}}(\mu_N^{\text{aux}} || \pi_N^{\text{aux}}) \xrightarrow{\mathbb{P}} \nu \otimes L\{\log[\Phi\nu(\Psi)/\nu L(\tilde{\Xi})]\}/\nu L(\tilde{\Xi}), \quad (\text{A.4})$$

as $N \rightarrow \infty$, recall the definition (given in Section 4) of the KLD and write, for any index $m \in \{1, \dots, \tilde{M}_N\}$,

$$d_{\text{KL}}(\mu_N^{\text{aux}} || \pi_N^{\text{aux}}) = \sum_{i=1}^{M_N} \mathbb{E}_{\mu_N^{\text{aux}}} \left[\log \Phi(\xi_{N,I_{N,m}}, \tilde{\xi}_{N,m}) \Big| I_{N,m} = i \right] \mu_N^{\text{aux}}(\{i\} \times \tilde{\Xi}) + \log \left[\frac{\sum_{j=1}^{M_N} \omega_{N,j} \psi_{N,j}}{\sum_{\ell=1}^{M_N} \omega_{N,\ell} L(\xi_{N,\ell}, \tilde{\Xi})} \right], \quad (\text{A.5})$$

where $\mathbb{E}_{\mu_N^{\text{aux}}}$ denotes the expectation associated with the random measure μ_N^{aux} . For each term of the sum in (A.5) we have

$$\mathbb{E}_{\mu_N^{\text{aux}}} \left[\log \Phi(\xi_{N,I_{N,m}}, \tilde{\xi}_{N,m}) \Big| I_{N,m} = i \right] \mu_N^{\text{aux}}(\{i\} \times \tilde{\Xi}) = \frac{\omega_{N,i} L(\xi_{N,i}, \log \Phi)}{\sum_{j=1}^{M_N} \omega_{N,j} L(\xi_{N,j}, \tilde{\Xi})},$$

and by using the consistency of $\{(\xi_{N,i}, \omega_{N,i})\}_{i=1}^{M_N}$ (under **(A1)**) we obtain the limit

$$\sum_{i=1}^{M_N} \mathbb{E}_{\mu_N^{\text{aux}}} \left[\log \Phi(\xi_{N,I_{N,m}}, \tilde{\xi}_{N,m}) \Big| I_{N,m} = i \right] \mu_N^{\text{aux}}(\{i\} \times \tilde{\Xi}) \xrightarrow{\mathbb{P}} \nu \otimes L(\log \Phi) / \nu L(\tilde{\Xi}),$$

where we used that $L(\cdot, |\log \Phi|) \in \mathbf{C}$ by assumption, implying, since \mathbf{C} is proper, $L(\cdot, \log \Phi) \in \mathbf{C}$. Moreover, under **(A2)**, by the continuous mapping theorem,

$$\log \left[\frac{\sum_{j=1}^{M_N} \omega_{N,j} \psi_{N,j}}{\sum_{\ell=1}^{M_N} \omega_{N,\ell} L(\xi_{N,\ell}, \tilde{\Xi})} \right] \xrightarrow{\mathbb{P}} \log[\nu(\Psi) / \nu L(\tilde{\Xi})],$$

which establishes (A.4). We prove that $\mathcal{E}(\{\tilde{\omega}_{N,i}\}_{i=1}^{\tilde{M}_N})$ tends to the same limit as $d_{\text{KL}}(\mu_N^{\text{aux}} || \pi_N^{\text{aux}})$. However, applying directly Lemma A.1 for $f = \log \Phi$ (which belongs to \mathbf{C}^* by assumption) and the limit (A.1) for $f \equiv 1$ yields, by the continuous mapping theorem,

$$\begin{aligned} \mathcal{E}(\{\tilde{\omega}_{N,i}\}_{i=1}^{\tilde{M}_N}) &= \tilde{\Omega}_N^{-1} \sum_{i=1}^{\tilde{M}_N} \tilde{\omega}_{N,i} \log \tilde{\omega}_{N,i} + \log(\tilde{M}_N \tilde{\Omega}_N^{-1}) \\ &\xrightarrow{\mathbb{P}} \nu \otimes L(\log \Phi) / \nu L(\tilde{\Xi}) + \log[\nu(\Psi) / \nu L(\tilde{\Xi})] \\ &= \nu \otimes L\{\log[\Phi\nu(\Psi) / \nu L(\tilde{\Xi})]\} / \nu L(\tilde{\Xi}). \end{aligned}$$

This completes the proof of i).

To prove *ii*) we first show that

$$d_{\chi^2}(\mu_N^{\text{aux}} || \pi_N^{\text{aux}}) \xrightarrow{\mathbb{P}} \nu(\Psi)\nu L(\Phi)/[\nu L(\tilde{\Xi})]^2 - 1. \quad (\text{A.6})$$

Indeed, recall the definition of the CSD and write, for any index $m \in \{1, \dots, \tilde{M}_N\}$,

$$\begin{aligned} d_{\chi^2}(\mu_N^{\text{aux}} || \pi_N^{\text{aux}}) &= \mathbb{E}_{\mu_N^{\text{aux}}} \left[\frac{d\mu_N^{\text{aux}}}{d\pi_N^{\text{aux}}}(\xi_{N,I_{N,m}}, \tilde{\xi}_{N,m}) \right] - 1 \\ &= \sum_{i=1}^{M_N} \mathbb{E}_{\mu_N^{\text{aux}}} \left[\frac{d\mu_N^{\text{aux}}}{d\pi_N^{\text{aux}}}(\xi_{N,I_{N,m}}, \tilde{\xi}_{N,m}) \middle| I_{N,m} = i \right] \mu_N^{\text{aux}}(\{i\} \times \tilde{\Xi}) - 1. \end{aligned}$$

Here

$$\begin{aligned} \mathbb{E}_{\mu_N^{\text{aux}}} \left[\frac{d\mu_N^{\text{aux}}}{d\pi_N^{\text{aux}}}(\xi_{N,I_{N,m}}, \tilde{\xi}_{N,m}) \middle| I_{N,m} = i \right] \mu_N^{\text{aux}}(\{i\} \times \tilde{\Xi}) \\ = \omega_{N,i} L(\xi_{N,i}, \Phi) \left[\sum_{j=1}^{M_N} \omega_{N,i} L(\xi_{N,j}, \tilde{\Xi}) \right]^{-2} \sum_{j=1}^{M_N} \omega_{N,i} \psi_{N,i}, \end{aligned}$$

and using the consistency of $\{(\xi_{N,i}, \omega_{N,i})\}_{i=1}^{M_N}$ yields the limit

$$\sum_{i=1}^{M_N} \mathbb{E}_{\mu_N^{\text{aux}}} \left[\frac{d\mu_N^{\text{aux}}}{d\pi_N^{\text{aux}}}(\xi_{N,I_{N,m}}, \tilde{\xi}_{N,m}) \middle| I_{N,m} = i \right] \mu_N^{\text{aux}}(\{i\} \times \tilde{\Xi}) \xrightarrow{\mathbb{P}} \nu(\Psi)\nu L(\Phi)/[\nu L(\tilde{\Xi})]^2.$$

which proves (A.6).

Finally, we prove that $\text{CV}^2(\{\tilde{\omega}_{N,i}\}_{i=1}^{\tilde{M}_N})$ tends to the same limit as $d_{\chi^2}(\mu_N^{\text{aux}} || \pi_N^{\text{aux}})$. Indeed, since Φ belongs to \mathbf{C}^* by assumption, we obtain, by applying Lemma A.1 together with (A.1),

$$\text{CV}^2(\{\tilde{\omega}_{N,i}\}_{i=1}^{\tilde{M}_N}) = (\tilde{M}_N \tilde{\Omega}_N^{-1}) \tilde{\Omega}_N^{-1} \sum_{i=1}^{\tilde{M}_N} \tilde{\omega}_{N,i}^2 - 1 \xrightarrow{\mathbb{P}} \nu(\Psi)\nu L(\Phi)/[\nu L(\tilde{\Xi})]^2 - 1. \quad (\text{A.7})$$

From this *ii*) follows. \square

A.2. Proof of Proposition 4.2. We denote by $q(\xi, \xi')$ the Radon-Nikodym derivative of the probability measure $\nu \otimes L/\nu L(\tilde{\Xi})$ with respect to $\nu \otimes R$, where the outer product \otimes of measure and a kernel is defined in (2.2):

$$q(\xi, \xi') \triangleq \frac{\frac{dL(\xi, \cdot)}{dR(\xi, \cdot)}(\xi')}{\iint_{\Xi \times \tilde{\Xi}} \nu(d\xi) L(\xi, d\xi')}, \quad (\text{A.8})$$

and by $p(\xi)$ the Radon-Nikodym derivative of the probability measure $\nu[\Psi] \otimes R/\nu(\Psi)$ with respect to $\nu \otimes R$:

$$p(\xi) = \frac{\Psi(\xi)}{\nu(\Psi)}. \quad (\text{A.9})$$

Using these notations and the definition (4.1) of the weight function Φ , we have

$$\frac{\Phi(\xi, \xi')\nu(\Psi)}{\nu L(\tilde{\Xi})} = \frac{\nu(\Psi) \frac{dL(\xi, \cdot)}{dR(\xi, \cdot)}(\xi')}{\Psi(\xi)\nu L(\tilde{\Xi})} = q(\xi, \xi')/p(\xi),$$

which implies that

$$\nu_{\text{KL}}(\Psi) = \iint_{\Xi \times \tilde{\Xi}} \nu(d\xi) R(\xi, d\xi') q(\xi, \xi') \log [q(\xi, \xi')/p(\xi)].$$

This relation establishes the first assertion (4.2). Similarly, we may write

$$\begin{aligned} \nu_{\chi^2}(\Psi) &= \frac{\iint_{\Xi \times \tilde{\Xi}} \nu(\Psi) R(\xi, d\xi') \left[\frac{dL(\xi, \cdot)}{dR(\xi, \cdot)}(\xi') \right]^2 \Psi^{-1}(\xi)}{[\nu L(\tilde{\Xi})]^2} - 1 \\ &= \iint_{\Xi \times \tilde{\Xi}} \nu(d\xi) R(\xi, d\xi') q^2(\xi, \xi')/p(\xi) - 1, \end{aligned}$$

which shows (4.2).

A.3. Proof of Corollary 4.1. Denote by $q(\xi) \triangleq \int_{\tilde{\Xi}} R(\xi, d\xi') q(\xi, \xi')$ the marginal density of the measure

$$A \in \mathcal{B}(\Xi) \mapsto \frac{\nu \otimes L}{\nu L(\tilde{\Xi})}(A \times \tilde{\Xi})$$

on $(\Xi, \mathcal{B}(\Xi))$. We denote by $q(\xi'|\xi) = q(\xi, \xi')/q(\xi)$ the conditional distribution. By the chain rule of entropy (the entropy of a pair of random variables is the entropy of one plus the conditional entropy of the other), we may split the KLD between the measures $\nu \otimes L/\nu L(\tilde{\Xi})$ and $\nu[\Psi] \otimes \nu(\Psi)/\nu(\Psi)$ as follows,

$$\eta_{\text{KL}}(\Psi) = \int_{\Xi} \nu(d\xi) q(\xi) \log[q(\xi)/p(\xi)] + \iint_{\Xi \times \tilde{\Xi}} \nu(d\xi) R(\xi, d\xi') q(\xi, \xi') \log(\xi|\xi').$$

The second term in the RHS of the previous equation does not depend on the adjustment multiplier weight Ψ . The first term is canceled if we set $p = q$, i.e., if

$$\frac{\Psi(\xi)}{\nu(\Psi)} = \int_{\tilde{\Xi}} R(\xi, d\xi') q(\xi, \xi') = L(\xi, \tilde{\Xi})/\nu L(\tilde{\Xi})$$

which establishes Assertion (4.1).

Consider now Assertion (4.1). Note first that

$$\begin{aligned}
& \iint_{\Xi \times \tilde{\Xi}} \nu(d\xi) R(\xi, d\xi') q^2(\xi, \xi')/p(\xi) - 1 \\
&= \int_{\Xi} \nu(d\xi) g^2(\xi)/p(\xi) - 1 \\
&= \nu^2(g) \left[\int_{\Xi} \nu(d\xi) \frac{g^2(\xi)}{p(\xi)\nu^2(g)} - 1 \right] + \nu^2(g) - 1, \tag{A.10}
\end{aligned}$$

where

$$g^2(\xi) = \int_{\tilde{\Xi}} R(\xi, d\xi') q^2(\xi, \xi').$$

The first term in the RHS of (A.10) is the CSD between the probability distributions with densities $g/\nu(g)$ and $\Psi/\nu(\Psi)$ with respect to ν . The second term does not depend on Ψ . The optimal value of the adjustment multiplier weight is obtained by canceling the first term. This establishes assertion (4.1).

REFERENCES

- Bollerslev, T., Engle, R. F., and Nelson, D. B. (1994) ARCH models. In *The Handbook of Econometrics* (R. F. Engle and D. MacFadden eds.), **4**, pp. 2959–3038.
- Cappé, O., Guillin, A., Marin, J., and Robert, C. (2004) Population Monte Carlo. *J. Comput. Graph. Statist.*, **13**, pp. 907–929.
- Cappé, O., Moulines, É., and Rydén, T. (2005) *Inference in Hidden Markov Models*, Springer.
- Chan, B., Doucet, A., and Tadic, V. B. (2003) Optimization of particle filters using simultaneous perturbation stochastic approximation. *Proc. IEEE ICASSP'03*.
- Cover, T. M., and Thomas, J. A. (1991) *Elements of Information Theory*, Wiley.
- Douc, R., and Moulines, É. (2005) Limit theorems for weighted samples with applications to sequential Monte Carlo methods. To appear in *Ann. Stat.*
- Douc, R., Guillin, A., Marin, J. M., and Robert, C. P. (2007a) Convergence of adaptive mixtures of importance sampling schemes. *Ann. Stat.*, **35**, pp. 420–448.
- Douc, R., Moulines, É., and Olsson, J. (2007b) On the auxiliary particle filter. Technical report ENST. URL <http://arXiv.org/abs/0709.3448>.
- Doucet, A., Godsill, S., and Andrieu, C. (2000) On sequential Monte-Carlo sampling methods for Bayesian filtering. *Stat. Comput.*, **10**, pp. 197–208.
- Fox, D. (2003) Adapting the sample size in particle filters through KLD-sampling. *International Journal of Robotics Research*, **22**, pp. 985–1004.
- Gordon, N. J., Salmond, D. J., and Smith, A. F. M. (1993) Novel approach to non-linear/non-Gaussian Bayesian state estimation. *IEEE Proc. Comm. Radar Signal Proc.*, **140**, pp. 107–113.
- Hürzeler, M., and Künsch, H. R. (1998) Monte Carlo approximations for general state space models. *J. Comput. Graph. Statist.*, **7**, pp. 175–193.

- Hu, X.-L., Schön, T. B., and Ljung, L. (2008) A basic convergence result for particle filtering. To appear in *IEEE Trans. Signal Proc.*
- Kitagawa, G. (1987) Non-Gaussian state space modeling of nonstationary time series. *J. Am. Statist. Assoc.*, **82**, pp. 1023–1063.
- Kong, A., Liu, J. S., and Wong, W. (1994) Sequential imputation and Bayesian missing data problems. *J. Am. Statist. Assoc.*, **89**, pp. 278–288.
- Künsch, H. R. (2005) Recursive Monte Carlo filters: algorithms and theoretical analysis. *Ann. Stat.*, **33**, pp. 1983–2021.
- Legland, F., and Oudjane, N. (2006) A sequential particle algorithm that keeps the particle system alive. Rapport de Recherche 5826, INRIA. URL <ftp://ftp.inria.fr/INRIA/publication/publi-pdf/RR/RR-5826.pdf>.
- Olsson, J., Moulines, É., and Douc, R. (2007) Improving the performance of the two-stage sampling particle filter: a statistical perspective. *Proceedings of the IEEE/SP 14th Workshop on Statistical Signal Processing*, Madison, USA, pp. 284–288.
- Pitt, M. K., and Shephard, N. (1999) Filtering via simulation: Auxiliary particle filters. *J. Am. Statist. Assoc.*, **87**, pp. 493–499.
- Soto, A. (2005) Self adaptive particle filter. (2005) *Proceedings of the 19th International Joint Conference on Artificial Intelligence* (L. P. Kaelbling and A. Saffiotti eds.), Edinburgh, Scotland, pp. 1398–1406.
- Straka, O., and Simandl, M. (2006) Particle filter adaptation based on efficient sample size. *Proceedings of the 14th IFAC Symposium on System Identification*, Newcastle, Australia, pp. 991–996.

(J. Cornebise) INSTITUT DES TÉLÉCOMS, TÉLÉCOM PARISTECH, 46 RUE BARRAULT, 75634 PARIS CEDEX 13, FRANCE
E-mail address: julien.cornebise@telecom-paristech.fr

(É. Moulines) INSTITUT DES TÉLÉCOMS, TÉLÉCOM PARISTECH, 46 RUE BARRAULT, 75634 PARIS CEDEX 13, FRANCE
E-mail address: moulines@telecom-paristech.fr

(J. Olsson) CENTER OF MATHEMATICAL SCIENCES, LUND UNIVERSITY, BOX 118, SE-22100 LUND, SWEDEN
E-mail address: jimmy@maths.lth.se