A Robustification Approach to Stability and to Uniform Particle Approximation of Nonlinear Filters : the Example of Pseudo–Mixing Signals

François Le Gland, Nadia Oudjane

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A Robustification Approach to Stability and to Uniform Particle Approximation of Nonlinear Filters : the Example of Pseudo–Mixing Signals

François Le Gland, Nadia Oudjane

N°4431

Mars 2002

THÈME 4

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--4431--FR+ENG

A Robustification Approach to Stability and to Uniform Particle Approximation of Nonlinear Filters : the Example of Pseudo–Mixing Signals

François Le Gland*, Nadia Oudjane**

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet SIGMA2 (σ^2)

Rapport de recherche n ° 4431 — Mars 2002 — 24 pages

Abstract: We propose a new approach to study the stability of the optimal filter w.r.t. its initial condition, by introducing a "robust" filter, which is exponentially stable and which approximates the optimal filter uniformly in time. The "robust" filter is obtained here by truncation of the likelihood function, and the robustification result is proved under the assumption that the Markov transition kernel satisfies a pseudo–mixing condition (weaker than the usual mixing condition), and that the observations are "sufficiently good". This robustification approach allows us to prove also the uniform convergence of several particle approximations to the optimal filter, in some cases of nonergodic signals.

Key-words: nonlinear filter, particle filter, stability, Hilbert metric, mixing, pseudo–mixing, robustification.

(Résumé : tsvp)

This work was partially supported by the CNRS programmes *Modélisation et Simulation Numérique* and *MathSTIC*.

* IRISA / INRIA, Campus de Beaulieu, 35042 RENNES Cédex, France. — legland@irisa.fr
** EDF, Division R&D, 1 avenue du Général de Gaulle, 92141 CLAMART Cédex, France. — nadia.oudjane@edf.fr

Une approche par robustification pour la stabilité et l'approximation particulaire uniforme des filtres non–linéaires : l'example des signaux pseudo–mélangeants

Résumé : Nous proposons une nouvelle approche pour étudier la stabilité du filtre optimal par rapport à sa condition initiale, en introduisant un filtre "robuste", exponentiellement stable et approchant le filtre optimal uniformément en temps. Le filtre "robuste" est obtenu içi en tronquant la fonction de vraisemblance, et le résultat de robustification est obtenu sous l'hypothèse que le noyau de transition markovien vérifie une condition de pseudo–mélange (plus faible que la condition de mélange habituelle), et que les observations soient "suffisament précises". Cette approche par robustification nous permet aussi de montrer la convergence uniforme en temps de plusieurs approximations particulaires du filtre optimal, dans certains cas de signaux non ergodiques.

Mots-clé : filtre non linéaire, filtre particulaire, stabilité, métrique de Hilbert, mélange, pseudo–mélange, robustification.

1 Introduction

The stability of the optimal nonlinear filter has been recently the subject of many works. The first stability result has been obtained by Ocone and Pardoux [14] who have used the approach of Stettner [17] and Kunita [10] to prove that the optimal filter forgets its initial condition in the L^p sense, when the signal itself is ergodic : however, their method of proof cannot provide a rate of convergence. A new approach based on the Hilbert projective metric has been recently introduced by Da Prato, Furhman and Malliavin [6, 7], and successfully used by Atar and Zeitouni [2] to obtain some results on the exponential stability of the optimal filter w.r.t. its initial condition. Independently, Del Moral and Guionnet [8] have developped another approach based on semi–group techniques and on the Dobrushin ergodic coefficient, to derive exponential stability results of the optimal filter w.r.t. its initial condition, which they have used to prove uniform convergence of the approximate interacting particle filters to the optimal filter, under some mixing condition on the Markov transition kernel.

However, the mixing condition is a rather strong condition, and the main objective of this paper is to relax it, using a robustification approach. In full generality, the idea behind robustification is as follows : if a perturbed sequence of probability distributions can be found,

- which is exponentially stable itself, e.g. because it satisfies some mixing condition,
- and which approximates the optimal filter in some sense, uniformly in time,

then the optimal filter is stable (but not exponentially stable). In this paper, the perturbed sequence of probability distributions is obtained by truncation of the likelihood function, as in Oudjane and Rubenthaler [15], and assuming that the Markov transition kernel satisfies a weaker pseudo–mixing condition, and that the observations are “sufficiently good”, we show that the robustification approach can be implemented effectively. The robustification result is proved in Theorem 5.6, where error bounds, averaged over observation sequences, are obtained for the approximation of the optimal filter by the “robust” filter. This result has two important consequences : (i) it allows us to obtain stability properties of the optimal filter w.r.t. its initial condition, see Theorem 6.2 and Theorem 6.3 where an a.s. result is obtained, and (ii) it is used in Section 7 to build particle filter approximations to the optimal filter, with error bounds, averaged over observation sequences, which are uniform in time.

The paper is organized as follows : In the next two sections, we define the framework of the nonlinear filtering problem, and we recall some notations and stability results obtained in LeGland and Oudjane [11] under the mixing condition. In Section 4, we introduce the weaker pseudo–mixing condition, and we give some examples of pseudo–mixing Markov kernels. A “robust” filter is introduced and studied in Section 5, for a model of a pseudo–mixing signal observed in additive, not necessarily Gaussian, white noise. In Section 6, we obtain the stability of the optimal filter w.r.t. its initial condition, as a consequence of the robustification result, and in Section 7, we define two particle filters for which we prove convergence to the optimal filter, uniformly in time.

2 Signal and observation model

On the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we consider two sequences of random variables : the signal sequence $\{X_n \, , \, n \geq 0\}$ and the observation sequence $\{Y_n \, , \, n \geq 1\}$, taking values in $E = \mathbb{R}^m$ and in $F = \mathbb{R}^d$, respectively :

- The signal $\{X_n \, , \, n \geq 0\}$ is an homogeneous Markov chain, with initial probability distribution μ_0, and transition kernel Q, i.e. for all $n \geq 1$

$$\mathbb{P}[X_n \in dx \mid X_{0:n-1} = x_{0:n-1}] = \mathbb{P}[X_n \in dx \mid X_{n-1} = x_{n-1}] = Q(x_{n-1}, dx) \ ,$$

  where $x_{0:n-1} = (x_0, \cdots, x_{n-1})$.

- The observation sequence $\{Y_n \, , \, n \geq 1\}$ is related to the signal $\{X_n \, , \, n \geq 0\}$ by

$$Y_n = h(X_n) + V_n \ ,$$

  for all $n \geq 1$, where $\{V_n \, , \, n \geq 1\}$ is a sequence of i.i.d. random variables, not necessarily Gaussian, with common probability density g w.r.t. the Lebesgue measure, independent of the signal $\{X_n \, , \, n \geq 0\}$. We define the corresponding likelihood function by

$$\Psi_n(x) = g(Y_n - h(x)) \ ,$$

  for any $x \in E$, which depends implicitly on the current observation Y_n.

The nonlinear filtering problem is to compute at each time n the optimal filter, i.e. the conditional probability distribution

$$\mu_n(dx) = \mathbb{P}[X_n \in dx \mid Y_{1:n}] \; ,$$

of the signal X_n, given a realization $Y_{1:n} = (Y_1, \cdots, Y_n)$ of the observation sequence up to the current time n, with the convention that $Y_{1:0} = \emptyset$. The transition from μ_{n-1} to μ_n involves the optimal prediction filter, i.e. the conditional probability distribution

$$\mu_{n|n-1}(dx) = \mathbb{P}[X_n \in dx \mid Y_{1:n-1}] \; ,$$

of the signal X_n, given a realization $Y_{1:n-1} = (Y_1, \cdots, Y_{n-1})$ of the observation sequence up to the previous time $(n-1)$. Whereas it is in general difficult to compute the optimal filter, its evolution is surprisingly easy to describe, and consists of the following two steps :

- In the *prediction* step, the prior knowledge on the signal provided by the transition kernel Q, is used to express $\mu_{n|n-1}$ in terms of μ_{n-1} as follows

  $$\mu_{n|n-1}(dx') = \int_E \mu_{n-1}(dx)\, Q(x, dx') = (Q\, \mu_{n-1})(dx') \; .$$

  Notice that the prediction step is linear w.r.t. μ_{n-1}.

- In the *correction* step, the posterior information provided by the incoming observation Y_n through the likelihood function Ψ_n, is used via the Bayes formula to update $\mu_{n|n-1}$ into μ_n as follows

  $$\mu_n(dx) = \frac{\Psi_n(x)\, \mu_{n|n-1}(dx)}{\displaystyle\int_E \Psi_n(x')\, \mu_{n|n-1}(dx')} = (\Psi_n \cdot \mu_{n|n-1})(dx) \; ,$$

  where $\cdot$ denotes the projective product. Notice that the correction step is nonlinear w.r.t. $\mu_{n|n-1}$.

Overall, the evolution of the optimal filter is summarized by the following diagram

$$\mu_{n-1} \xrightarrow[\text{prediction}]{} \mu_{n|n-1} = Q\, \mu_{n-1} \xrightarrow[\text{correction}]{} \mu_n = \Psi_n \cdot \mu_{n|n-1} \; .$$

This evolution can be described in terms of the nonnegative kernel

$$R_n(x, dx') = Q(x, dx')\, \Psi_n(x') \; ,$$

and of the associated integral operator R_n and normalized operator $\bar{R}_n$, both acting on the set $\mathcal{M}^+(E)$ of nonnegative and finite measures on E, and defined respectively by

$$(R_n\, \mu)(dx') = \int_E \mu(dx)\, R_n(x, dx') = \int_E \mu(dx)\, Q(x, dx')\, \Psi_n(x') \; ,$$

and $\bar{R}_n(\mu) = (R_n\, \mu)/(R_n\, \mu)(E)$ if $(R_n\, \mu)(E) > 0$, and $\bar{R}_n(\mu) = 0$ otherwise. Notice that R_n depends on the observation Y_n through the likelihood function Ψ_n. With these definitions, the evolution of the optimal filter on the set $\mathcal{P}(E)$ of probability distributions on E can be described by the single formula

$$\mu_n = \frac{R_n\, \mu_{n-1}}{(R_n\, \mu_{n-1})(E)} = \bar{R}_n(\mu_{n-1}) = \Psi_n \cdot (Q\, \mu_{n-1}) \; , \tag{1}$$

which by induction yields

$$\mu_n = \bar{R}_n(\mu_{n-1}) = \bar{R}_n \circ \cdots \circ \bar{R}_m(\mu_{m-1}) = \bar{R}_{n:m}(\mu_{m-1}) \; .$$

Equation (1) shows clearly that the nonlinearity in the evolution of the optimal filter is only due to the normalization in the Bayes formula, occuring in the correction step. The Hilbert projective metric has been introduced in [6, 7, 2] precisely to get rid of this normalization term, and to reduce the problem to the analysis of a linear evolution on $\mathcal{M}^+(E)$.

3 Stability and robustification

In this section, we define the mixing property and the Hilbert metric, we describe roughly the robustification approach, and we recall some stability results obtained in [11] under the mixing assumption, that will be useful in the sequel.

Definition 3.1. *Two nonnegative measures $\mu, \mu' \in \mathcal{M}^+(E)$ are* comparable, *if there exists positive constants $0 < a \leq b < \infty$, such that*

$$a\,\mu'(A) \leq \mu(A) \leq b\,\mu'(A)\ ,$$

for any Borel subset $A \subset E$.

Definition 3.2 (Mixing). *A nonnegative kernel K defined on $\mathcal{M}^+(E)$ is* mixing, *if there exists a positive constant $0 < \varepsilon \leq 1$, and a nonnegative measure $\lambda \in \mathcal{M}^+(E)$, such that*

$$\varepsilon\,\lambda(A) \leq K(x, A) \leq \frac{1}{\varepsilon}\,\lambda(A)\ ,$$

for any $x \in E$, and any Borel subset $A \subset E$. The constant ε is called a mixing constant, *and the measure λ is called a* mixing measure *associated with the mixing kernel K.*

The mixing property is related to a kernel that is little dependent on the initial state x. A special and extreme example of a mixing kernel is when $K(x, dx')$ does not depend on x at all, in which case $\varepsilon = 1$.

Definition 3.3. *The* Hilbert metric *on $\mathcal{M}^+(E)$ is defined by*

$$h(\mu, \mu') = \begin{cases} \log \dfrac{\displaystyle\sup_{A\,:\,\mu'(A)>0} \frac{\mu(A)}{\mu'(A)}}{\displaystyle\inf_{A\,:\,\mu'(A)>0} \frac{\mu(A)}{\mu'(A)}}\ , & \text{if } \mu \text{ and } \mu' \text{ are comparable,} \\ 0\ , & \text{if } \mu = \mu' \equiv 0, \\ +\infty\ , & \text{otherwise.} \end{cases}$$

Notice that h is a projective metric, i.e. it is invariant under multiplication by positive constants, hence $h(\mu, \mu') = h(\dfrac{\mu}{\mu(E)}, \dfrac{\mu'}{\mu'(E)})$, for any μ and $\mu' \in \mathcal{M}^+(E)$. In the nonlinear filtering context, this projective property allows us to consider the linear transformation $\mu \longmapsto R_n\,\mu$ instead of the nonlinear transformation $\mu \longmapsto \bar{R}_n(\mu) = (R_n\,\mu)/(R_n\,\mu)(E)$.

Lemma 3.4. *If μ and μ' are comparable, then for any nonnegative kernels K and K' it holds*

$$\frac{(K\,\mu)(E)}{(K\,\mu')(E)}\,\frac{(K'\,\mu')(E)}{(K'\,\mu)(E)} \leq \exp(h(\mu, \mu'))\ .$$

Proof of Lemma 3.4. By definition, if μ and μ' are comparable, then there exist constants $0 < a \leq b < \infty$ such that

$$a\,\mu'(A) \leq \mu(A) \leq b\,\mu'(A)\ ,$$

for any Borel subset $A \subset E$. The optimal values for the constants a and b are

$$a = \inf_{x\in E} \frac{d\mu}{d\mu'}(x) = 1\ /\ \sup_{x\in E} \frac{d\mu'}{d\mu}(x) \qquad \text{and} \qquad b = \sup_{x\in E} \frac{d\mu}{d\mu'}(x)\ ,$$

and it holds $h(\mu, \mu') = \log \dfrac{b}{a}$. For any nonnegative kernel K defined on E

$$a\ (K\,\mu')(E) \leq (K\,\mu)(E) = \int_E K(x, E)\,\mu(dx) \leq b\ (K\,\mu')(E)\ ,$$

hence

$$\frac{(K\,\mu)(E)}{(K\,\mu')(E)} \leq b \qquad \text{and} \qquad \frac{(K\,\mu')(E)}{(K\,\mu)(E)} \leq \frac{1}{a} \; .$$

If K' is another nonnegative kernel defined on E, then

$$\frac{(K\,\mu)(E)}{(K\,\mu')(E)} \, \frac{(K'\,\mu')(E)}{(K'\,\mu)(E)} \leq \frac{b}{a} = \exp(h(\mu,\mu')) \; . \qquad \Box$$

The analysis of stability properties of the optimal filter is of great interest. Indeed, one has rarely access in practice to the initial probability distribution of the signal, and it is important to know whether the filter is sensitive or not to this condition. More generally, the parameters of the signal / observation model, such as the densities of the observation noise and signal noise are usually not available, and it is crucial to know whether the optimal filter is robust w.r.t. such model errors. Finally, to compute the optimal filter we usually introduce errors in the transitions of the filter, because the true transitions are not practically computable. Before entering into details, let us make precise what we mean here by stability.

Definition 3.5 (Stability). *A sequence $\{S_n \, , \, n \geq 1\}$ of nonlinear transformations on the metric space $(\mathcal{P}(E), d)$ is*

(i) stable, *if for any two sequences $\{\mu_n \, , \, n \geq 0\}$ and $\{\mu'_n \, , \, n \geq 0\}$ defined on $\mathcal{P}(E)$ by the same recursion $\mu_n = S_n(\mu_{n-1})$ and $\mu'_n = S_n(\mu'_{n-1})$ for any $n \geq 1$, with possibly different initial conditions, it holds*

$$\lim_{n\to\infty} d(\mu_n, \mu'_n) = 0 \; , \tag{2}$$

(ii) stable w.r.t. local perturbations, *if for any sequence $\{\mu_n \, , \, n \geq 0\}$ defined on $\mathcal{P}(E)$ by the recursion $\mu_n = S_n(\mu_{n-1})$, and for any sequence $\{\mu'_n \, , \, n \geq 0\}$ defined on $\mathcal{P}(E)$, such that the local error is uniformly controlled, i.e. such that for any $n \geq 1$*

$$d(\mu'_n, S_n(\mu'_{n-1})) \leq \delta < \infty \; ,$$

it holds

$$\limsup_{n\to\infty} d(\mu_n, \mu'_n) \leq C\,\delta \; . \tag{3}$$

A sufficient condition for stability can be formulated in terms of contraction coefficients : indeed, if the sequence $\{S_n \, , \, n \geq 1\}$ is uniformly contracting, in the sense that

$$\sup_{\mu,\mu' \in \mathcal{P}(E) \, : \, \mu \neq \mu'} \frac{d(S_n(\mu), S_n(\mu'))}{d(\mu,\mu')} \leq \tau < 1 \; ,$$

for any $n \geq 1$, then $(\star)$ can be replaced by

$$d(\mu_n, \mu'_n) \leq \tau^n \, d(\mu_0, \mu'_0) \; ,$$

which implies (exponential) stability, and $(\star\star)$ can be replaced by

$$d(\mu_n, \mu'_n) \leq \frac{\delta}{1-\tau} + \tau^n \, d(\mu_0, \mu'_0) \; ,$$

which implies stability w.r.t. local perturbations.

If the sequence $\{S_n \, , \, n \geq 1\}$ is not uniformly contracting, stability can still hold, and can be proved sometimes using a robustification approach. Assume that a perturbed sequence $\{S^h_n \, , \, n \geq 1\}$ of nonlinear transformations can be found, which is uniformly contracting itself, i.e. such that for any $n \geq 1$

$$\sup_{\mu,\mu' \in \mathcal{P}(E) \, : \, \mu \neq \mu'} \frac{d(S^h_n(\mu), S^h_n(\mu'))}{d(\mu,\mu')} \leq \tau_h < 1 \; ,$$

and such that the local error is uniformly controlled, in the sense that for any $n \geq 1$

$$\sup_{\mu \in \mathcal{P}(E)} d(S_n(\mu), S^h_n(\mu)) \leq \delta_h < \infty \; .$$

(In general $\delta_h \xrightarrow{h} 0$ and $\tau_h \xrightarrow{h} 1$, otherwise the sequence $\{S_n\,,\, n \geq 1\}$ would be uniformly contracting). Then the approximation is uniform : for any two sequences $\{\mu_n\,,\, n \geq 0\}$ and $\{\mu_n^h\,,\, n \geq 0\}$ defined on $\mathcal{P}(E)$ by the recursion $\mu_n = S_n(\mu_{n-1})$ and $\mu_n^h = S_n^h(\mu_{n-1}^h)$ respectively for any $n \geq 1$, it holds

$$d(\mu_n, \mu_n^h) \leq \frac{\delta_h}{1 - \tau_h} + \tau_h^n \, d(\mu_0, \mu_0^h) \;,$$

and moreover the original sequence $\{S_n\,,\, n \geq 1\}$ is stable (but not exponentially stable) : for any two sequences $\{\mu_n\,,\, n \geq 0\}$ and $\{\mu'_n\,,\, n \geq 0\}$ defined on $\mathcal{P}(E)$ by the same recursion $\mu_n = S_n(\mu_{n-1})$ and $\mu'_n = S_n(\mu'_{n-1})$ for any $n \geq 1$, with possibly different initial conditions, it holds

$$d(\mu_n, \mu'_n) \leq d(\mu_n, \mu_n^h) + d(\mu_n^h, \mu_n'^h) + d(\mu_n'^h, \mu'_n) \leq \frac{2\,\delta_h}{1 - \tau_h} + \tau_h^n \, d(\mu_0, \mu'_0) \;,$$

where the two sequences $\{\mu_n^h\,,\, n \geq 0\}$ and $\{\mu_n'^h\,,\, n \geq 0\}$ are defined on $\mathcal{P}(E)$ by the same recursion $\mu_n^h = S_n^h(\mu_{n-1}^h)$ and $\mu_n'^h = S_n^h(\mu_{n-1}'^h)$ for any $n \geq 1$, with initial conditions $\mu_0^h = \mu_0$ and $\mu_0'^h = \mu'_0$, hence

$$\limsup_{n \to \infty} d(\mu_n, \mu'_n) \leq \frac{2\,\delta_h}{1 - \tau_h} \;.$$

Since the left hand side does not depend on the perturbation parameter h, it holds

$$\lim_{n \to \infty} d(\mu_n, \mu'_n) = 0 \;,$$

provided that

$$\frac{\delta_h}{1 - \tau_h} \xrightarrow{\;h\;} 0 \;.$$

In [11, Section 4] we have proved some results of stability of the optimal filter w.r.t. its initial condition and w.r.t. local perturbations, under the mixing assumption, using the Hilbert metric. We summarize here the main results that will be useful in the sequel.

Theorem 3.6. *Let $\{\mu_n\,,\, n \geq 0\}$ be the optimal filter as defined in Section 2. Assume that for any $k \geq 1$, the nonnegative kernel R_k is mixing with $\varepsilon_k > 0$.*

(i) Let $\{\mu'_n\,,\, n \geq 0\}$ be a wrongly initialized filter, i.e. $\mu'_n = \bar{R}_{n:1}(\mu'_0)$ for any $n \geq 1$, then

$$\|\mu_n - \mu'_n\| \leq \frac{2}{\log 3} \prod_{k=2}^{n} (1 - \varepsilon_k^2) \, \frac{1}{\varepsilon_1^2} \, \|\mu_0 - \mu'_0\| \;,$$

where $\|\cdot\|$ denotes the total variation norm on the space of signed measures on E.

(ii) Let $\{\mu'_n\,,\, n \geq 0\}$ be a sequence of probability measures such that $\mu'_0 = \mu_0$, and

$$\mathbb{E}[F(\mu'_k) \mid Y_{1:n}] = \mathbb{E}[F(\mu'_k) \mid Y_{1:k}] \;,$$

for any bounded measurable function F defined on $\mathcal{P}(E)$. If the local error in the total variation sense is controlled for any $k \geq 1$ by

$$\mathbb{E}[\, \|\mu'_k - \bar{R}_k(\mu'_{k-1})\| \mid Y_{1:k}\,] \leq \delta_k^{\rm TV} \;,$$

then

$$\mathbb{E}[\, \|\mu_n - \mu'_n\| \mid Y_{1:n}\,] \leq \delta_n^{\rm TV} + \frac{2}{\log 3} \sum_{k=1}^{n-1} [\, \prod_{\ell=k+2}^{n} (1 - \varepsilon_\ell^2) \,] \, \frac{\delta_k^{\rm TV}}{\varepsilon_{k+1}^2} \;.$$

If the local error in the weak sense is controlled for any $k \geq 1$ by

$$\sup_{\phi \,:\, \|\phi\|=1} \mathbb{E}[\, | \langle \mu'_k - \bar{R}_k(\mu'_{k-1}), \phi \rangle | \mid Y_{1:k}\,] \leq \delta_k^{\rm W} \;,$$

then

$$\sup_{\phi \,:\, \|\phi\|=1} \mathbb{E}[\, | \langle \mu_n - \mu'_n, \phi \rangle | \mid Y_{1:n}\,] \leq \delta_n^{\rm W} + 2\, \frac{\delta_{n-1}^{\rm W}}{\varepsilon_n^2} + \frac{4}{\log 3} \sum_{k=1}^{n-2} [\, \prod_{\ell=k+3}^{n} (1 - \varepsilon_\ell^2) \,] \, \frac{\delta_k^{\rm W}}{\varepsilon_{k+2}^2 \, \varepsilon_{k+1}^2} \;.$$

A convenient way to approximate numerically the optimal filter is to use a particle method that provides a sequence of random empirical probability distributions $\{\mu'_n \,,\, n \geq 0\}$ whose evolution is both close to the evolution of the optimal filter and computable. The results of Theorem 3.6 show that, if we are able to control uniformly in time the local errors committed at each time step by using the wrong evolution $\mu'_{k-1} \longmapsto \mu'_k$ instead of the true evolution $\mu'_{k-1} \longmapsto \bar{R}_k(\mu'_{k-1})$, then we are also able to control uniformly in time the resulting global error between $\{\mu_n \,,\, n \geq 0\}$ and $\{\mu'_n \,,\, n \geq 0\}$, provided that the signal is mixing. Unfortunatly the mixing condition implies in general a strong ergodicity assumption on the signal, which practically requires that the state space is compact. The aim of this paper is precisely to relax this assumption, using the robustification approach, and still obtain results similar to those of Theorem 3.6. In this sense, numerical approximation of the optimal filter, e.g. using particle methods, provides another motivation for the robustification approach, see Section 7 below or [11, Section 6]. Indeed, it is sometimes a good idea to approximate rather the "robust" filter, defined by a perturbed wrong evolution, especially when it enjoys some additional regularity property : usually in such cases, the local errors can be estimated more precisely, and their propagation under the perturbed evolution is better controlled. This results in better convergence properties, which of course need to be balanced with the residual error resulting from the approximation of the optimal filter by the "robust" filter.

4 Pseudo–mixing signals

Our objective in this paper is to present an extension of the stability results recalled above, and in this section we introduce the pseudo–mixing assumption, which will allow us to relax the mixing assumption and still obtain stability properties for the optimal filter. There are already some results available in the literature, which prove the stability of the filter without assuming ergodicity of the signal : Budhiraja and Ocone [4, 5] have proved that the optimal filter forgets its initial condition with a rate which is asymptotically exponential in some special cases where the signal is not necessarily ergodic but the observations are "sufficiently good". The interest of our result is that it provides a bound for the rate of convergence which is nonasymptotic, i.e. valid at each time instant. This fact will be used in Section 7 to provide uniform particle approximations to the optimal filter.

Definition 4.1 (Pseudo–mixing). *A nonnegative kernel K defined on E is* pseudo–mixing, *if for any compact set C in E with diameter D large enough, there exist a positive constant $0 < \varepsilon(D) \leq 1$, depending only on the diameter D, and a nonnegative measure $\lambda_C \in \mathcal{M}^+(E)$, such that*

$$\varepsilon(D)\, \lambda_C(A) \leq K(x,A) \leq \frac{1}{\varepsilon(D)}\, \lambda_C(A) \;, \tag{4}$$

for any $x \in C$, and any Borel subset $A \subset E$. A Markov chain with a pseudo–mixing transition kernel is called a pseudo–mixing *signal.*

Remark 4.2. If the nonnegative kernel K is Markov on C, i.e. if $K(x,E) = 1$ for any $x \in C$, then without loss of generality the pseudo–mixing measure λ_C can be taken as a probability distribution. Indeed, taking $A = E$ in the pseudo–mixing equation (4) yields

$$\varepsilon(D)\, \lambda_C(E) \leq K(x,E) = 1 \leq \frac{1}{\varepsilon(D)}\, \lambda_C(E) \;,$$

hence

$$\varepsilon(D) \leq \lambda_C(E) \leq \frac{1}{\varepsilon(D)} \;,$$

and

$$\varepsilon^2(D)\, \frac{\lambda_C(A)}{\lambda_C(E)} \leq K(x,A) \leq \frac{1}{\varepsilon^2(D)}\, \frac{\lambda_C(A)}{\lambda_C(E)} \;,$$

for any $x \in C$, and any Borel subset $A \subset E$.

Example 4.3. To illustrate the pseudo–mixing property, we can for instance consider kernels Q of the form

$$Q(x,dx') = q(x'-x)\, dx' = \ell(|x'-x|)\, dx' \;,$$

where ℓ is a bounded positive function defined on $[0,\infty)$. If there exists a constant $M > 0$ large enough such that

$$\begin{cases} \dfrac{\ell(u+v)}{\ell(u)\,\ell(v)} \geq a > 0 \;, \quad \text{for any } u \geq M \text{ and any } v \geq M, \\ \\ \qquad\qquad \text{and } \ell \text{ is decreasing to zero on } [M,\infty) \end{cases} \tag{5}$$

then Q is a pseudo–mixing kernel. Precisely, for any compact set C with diameter $D \geq M$ large enough

$$\varepsilon(D)\,\lambda_C(dx') \leq q(x'-x)\,dx' \leq \frac{1}{\varepsilon(D)}\,\lambda_C(dx') \;, \tag{6}$$

for any $x \in C$ and any $x' \in E$, with pseudo–mixing constant $\varepsilon(D) = a\,\ell(D+M)$ and absolutely continuous pseudo–mixing measure

$$\lambda_C(dx') = [\mathbf{1}_{(d(x',C)\,\leq\, M)} + \mathbf{1}_{(d(x',C)\,>\,M)}\,q(x'-z)]\,dx' \;,$$

for any $x' \in E$, where z is an arbitrary element of C. Indeed, if $d(x',C) > M$, then for any $x, z \in C$ it holds

$$M \leq |x-x'| \leq |x-z| + |z-x'| \leq D + |z-x'| \;,$$

and

$$M \leq |z-x'| \leq |z-x| + |x-x'| \leq D + |x-x'| \;,$$

hence, since $\ell(D) \geq \ell(D+M)$

$$q(x'-x) \geq \ell(D+|z-x'|) \geq a\,\ell(D+M)\,q(x'-z) \;,$$

and

$$q(x'-z) \geq \ell(D+|x-x'|) \geq a\,\ell(D+M)\,q(x'-x) \;.$$

It follows that

$$a\,\ell(D+M)\,q(x'-z) \leq q(x'-x) \leq \frac{1}{a\,\ell(D+M)}\,q(x'-z) \;, \tag{7}$$

for any $x, z \in C$ and any $x' \in E$ such that $d(x',C) > M$. On the other hand, if $d(x',C) \leq M$, then for any $x \in C$ it holds

$$0 \leq |x-x'| \leq D+M \;,$$

hence

$$\inf_{0 \leq u \leq D+M} \ell(u) \leq q(x'-x) \leq \sup_{0 \leq u \leq D+M} \ell(u) \;,$$

and for D large enough

$$\inf_{0 \leq u \leq D+M} \ell(u) = \min(\inf_{0 \leq u \leq M} \ell(u), \ell(D+M)) = \ell(D+M) \;,$$

and

$$\sup_{0 \leq u \leq D+M} \ell(u) = \sup_{0 \leq u \leq M} \ell(u) \leq \frac{1}{\ell(D+M)} \;.$$

Without loss of generality, we can assume that $a \leq 1$ in (5) (otherwise, take $\min(a,1)$ instead), and it follows that

$$a\,\ell(D+M) \leq q(x'-x) \leq \frac{1}{a\,\ell(D+M)} \;, \tag{8}$$

for any $x \in C$ and any $x' \in E$ such that $d(x', C) \leq M$. Combining (7) and (8) finally yields (6).

We can for example consider the following classical densities defined for any $x \in \mathbb{R}$ by

$$q(x) = \frac{p-1}{2p}\,(1 + \frac{|x|}{p})^{-p} \;, \qquad \text{with } p \geq 2,$$

$$q(x) = \tfrac{1}{2}\,\exp(-|x|) \;, \qquad \text{(exponential density)},$$

$$q(x) = \tfrac{1}{2}\,\frac{1}{2 + \exp(|x|) + \exp(-|x|)} \;, \qquad \text{(logistic density)},$$

which all satisfy (5).

As we can see, the property (5) requires some conditions on the tails of the densities, but unfortunately this property is not satisfied by the Gaussian densities because of their too light tails.

Remark 4.4. Let Q be a pseudo–mixing kernel, and let f be a Lipschitz continuous function defined on E and taking values in E, i.e. there exists a positive constant $a > 0$ such that

$$|f(x) - f(x')| \leq a\,|x - x'| \;,$$

for any x, $x' \in E$. Then, the nonnegative kernel Q_f defined by $Q_f(x, A) = Q(f(x), A)$ for any $x \in E$, and any Borel subset $A \subset E$, is also pseudo–mixing. Indeed, let C be a compact set in E, with diameter D, and let $f(C) = \{x' \in E \,:\, x' = f(x) \text{ for some } x \in C\}$ denote the image of C by f. Then, the set $f(C)$ is compact, with diameter smaller than $a\,D$, and since Q satisfies the peudo–mixing property (4), it holds

$$\varepsilon(a\,D)\,\lambda_{f(C)}(A) \leq Q_f(x, A) = Q(f(x), A) \leq \frac{1}{\varepsilon(a\,D)}\,\lambda_{f(C)}(A) \;,$$

for any $x \in C$, and any Borel subset $A \subset E$. This remark allows us to extend the simple examples of pseudo–mixing kernels given in Example 4.3 above to the case of signals with dynamics of the form

$$X_{n+1} = f(X_n) + W_n \;,$$

where $\{W_n \,,\, n \geq 0\}$ is a sequence of i.i.d. random variables with common probability density q of the form given in Example 4.3, and where f is a Lipschitz continuous function defined on E.

5 Approximation of the optimal filter by a "robust" filter

In this section, we show that if the transition kernel of the signal is pseudo–mixing, and the observations are "sufficiently good", then we can approximate the optimal filter uniformly in time by an exponentially stable sequence of probability distributions. This exponentially stable sequence is called a "robust" filter because it is an approximation to the optimal filter, which is much less sensitive to perturbations than the optimal filter itself. This robustification approach is the main contribution of the paper : it will imply that the optimal filter itself forgets its initial condition at a rate that will be precised in Section 6, and in Section 7, it will provide particle filter approximations which converge uniformly in time to the optimal filter.

We consider the following state / observation model, where the signal $\{X_n \,,\, n \geq 0\}$ is a Markov chain with initial probability distribution μ_0, and transition kernel Q, observed in additive noise, i.e.

$$Y_n = h(X_n) + V_n \;,$$

for all $n \geq 1$, where $\{V_n \,,\, n \geq 1\}$ is a sequence of i.i.d. random variables, not necessarily Gaussian, with common probability density g w.r.t. the Lebesgue measure, independent of the signal $\{X_n \,,\, n \geq 0\}$. We introduce the following assumptions.

Assumption A1 The transition kernel Q is pseudo–mixing, and for any compact C with diameter $D \geq D_0$, let $\varepsilon(D)$ denote the corresponding mixing constant.

Assumption A2 For any $y \in F$ and any $\Delta > 0$, the sublevel set

$$C(y,\Delta) = \{x \in E \, : \, |y - h(x)| \leq \Delta\}$$

is *compact*, and its diameter is bounded by a finite constant $D(\Delta)$ independent of $y \in F$.

Assumption A3 For any $\Delta > 0$, the following quantity characterizes the tails of the observation noise density

$$\alpha(\Delta) = \int_F \mathbf{1}_{(|u| > \Delta)} \, g(u) \, du \ ,$$

and satisfies

$$\lim_{\Delta \to \infty} \frac{\alpha(\Delta)}{\varepsilon^4(D(\Delta))} = 0 \ .$$

Assumption A1 replaces the stronger mixing assumption, and can be satisfied by nonergodic signals. Assumption A2 is a relatively strong assumption, which in some sense requires that the signal is completely observed, see Remark 5.1 below. Finally, Assumption A3 suggests that the tails of the observation noise are light w.r.t. the tails of the signal noise : indeed in Section 4, we have seen some examples of signals with additive noise, where the mixing constant $\varepsilon(D)$ can be related to the tails of the signal noise.

Remark 5.1. If the function h is injective from E in F with Lipschitz continuous inverse, i.e. if there exists a positive constant $b > 0$ such that

$$|x - x'| \leq b \, |h(x) - h(x')| \ ,$$

for any $x, x' \in \mathbb{R}^m$, then Assumption A2 holds with $D(\Delta) = 2 \, b \, \Delta$. This is only a sufficient condition, which obviously is not necessary, as the following two examples show, where $E = F = \mathbb{R}$. In the first example, depicted in Figure 1, the observation function h is noninjective, but it is "injective at infinity" in the sense that, outside a compact interval of length L, it is injective and has a constant slope $1/b$. Assumption A2 holds there with $D(\Delta) = 2 \, b \, \Delta + L$.

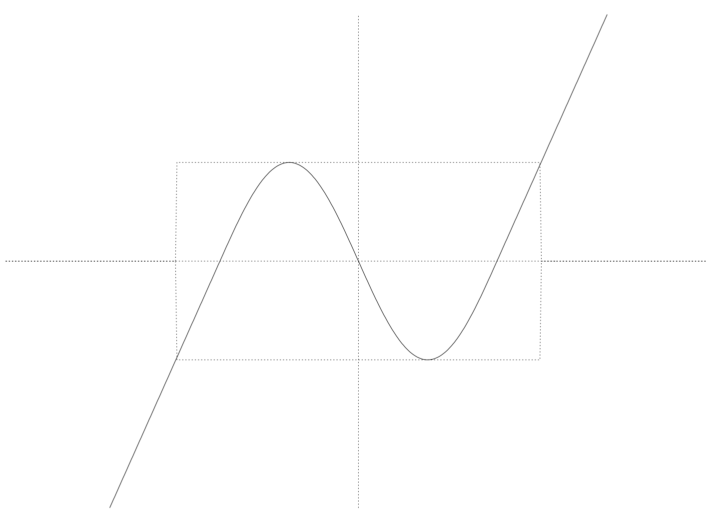

Figure 1: Noninjective observation function.

In the second example, depicted in Figure 2, the observation function h is injective, but its inverse is not Lipschitz continuous : it is defined by $h(x) = \phi(x)$ if $x \geq 0$, and $h(x) = -\phi(-x)$ if $x \leq 0$, where the function ϕ defined on $[0, \infty)$ is continuous, strictly convex, monotonically increasing, and satifies $\phi(0) = 0$. Assumption A2 holds there with $D(\Delta) = 2 \, \phi^{-1}(\Delta)$.

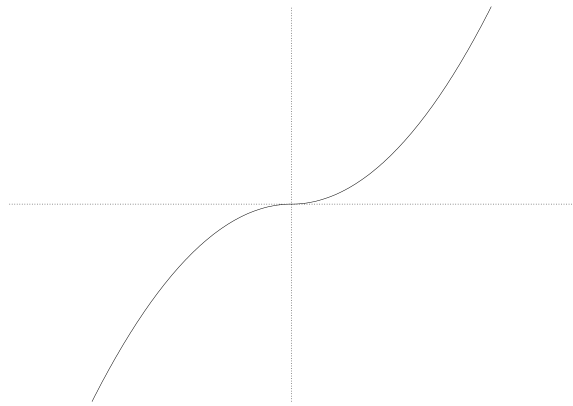

Figure 2: Injective observation function, with non Lispschitz inverse.

Let $\{\mu_n\ ,\ n \geq 0\}$ denote the optimal filter associated with the above model. We recall that

$$\mu_n = \bar{R}_n(\mu_{n-1}) = \bar{R}_{n:1}(\mu_0)\ ,$$

where for any $n \geq 1$,

$$R_n(x, dx') = Q(x, dx')\, \Psi_n(x') = Q(x, dx')\, g(Y_n - h(x'))\ ,$$

for any $x, x' \in E$.

For any nondecreasing sequence $\Delta = \{\Delta_k\ ,\ k \geq 1\}$, we introduce the following notations, under Assumptions A1 and A2 : for any $k \geq 1$

- $C_k = C(Y_k, \Delta_k)$ denotes the compact set

  $$C_k = \{x \in \mathbb{R}^m\ :\ |Y_k - h(x)| \leq \Delta_k\}\ ,$$

  with diameter $D_k \leq D(\Delta_k)$,

- $\lambda_k = \lambda_{C_{k-1}}$ and $\varepsilon(D_{k-1}) \geq \varepsilon(D(\Delta_{k-1})) = \varepsilon_k$ denote the mixing probability measure and the mixing constant respectively, associated with the pseudo–mixing kernel Q on the compact set C_{k-1},

- Ψ_k^Δ denotes the truncated likelihood function defined by

  $$\Psi_k^\Delta(x') = \mathbf{1}_{C_k}(x')\, \Psi_k(x')\ ,$$

  for any $x' \in E$,

- R_k^Δ denotes the nonnegative kernel defined by

  $$R_k^\Delta(x, dx') = \begin{cases} Q(x, dx')\, \Psi_k^\Delta(x')\ , & \text{if } x \in C_{k-1}, \\[2ex] \lambda_k(dx')\, \Psi_k^\Delta(x')\ , & \text{if } x \notin C_{k-1}. \end{cases}$$

Notice that R_k^Δ depends on two successive observations Y_{k-1} and Y_k, and since the kernel Q is pseudo–mixing, the following mixing property is satisfied

$$\varepsilon_k\, \Psi_k^\Delta(x')\, \lambda_k(dx') \leq R_k^\Delta(x, dx') \leq \frac{1}{\varepsilon_k}\, \Psi_k^\Delta(x')\, \lambda_k(dx')\ ,$$

hence for any $\mu \in \mathcal{P}(E)$

$$(R_k^\Delta\, \mu)(E) \geq \varepsilon_k \int_E \Psi_k^\Delta(x')\, \lambda_k(dx') = \varepsilon_k \int_{C_k} \Psi_k(x')\, \lambda_k(dx')\ ,$$

and we introduce the following additional assumption.

Assumption A4 For any $k \geq 1$

$$\int_{C_k} \Psi_k(x')\, \lambda_k(dx') > 0 \ .$$

Remark 5.2. Notice that

$$\inf_{x' \in C_k} \Psi_k(x') \geq \inf_{|u| \leq \Delta_k} g(u) \qquad \text{and} \qquad \lambda_k(C_k) \geq \varepsilon_k \sup_{x \in C_{k-1}} Q(x, C_k) \ ,$$

hence a sufficient condition for Assumption A4 to hold is

$$\inf_{|u| \leq \Delta_k} g(u) > 0 \qquad \text{and} \qquad \sup_{x \in C_{k-1}} Q(x, C_k) > 0 \ .$$

Finally, for any two initial conditions μ_0 and μ'_0, let $\{\mu_n^\Delta \ , \ n \geq 0\}$ and $\{\mu'^\Delta_n \ , \ n \geq 0\}$ denote the two sequences of probability distributions defined by the following recursions

$$\mu_n^\Delta = \frac{R_n^\Delta\, \mu_{n-1}^\Delta}{(R_n^\Delta\, \mu_{n-1}^\Delta)(E)} = \bar{R}_n^\Delta(\mu_{n-1}^\Delta) = \bar{R}_{n:1}^\Delta(\mu_0) \qquad \text{and} \qquad \mu'^\Delta_n = \bar{R}_{n:1}^\Delta(\mu'_0)$$

where $\bar{R}_{n:1}^\Delta = \bar{R}_n^\Delta \circ \cdots \circ \bar{R}_1^\Delta$. Notice that for any $k \geq 1$, the supports of the nonnegative measures μ_{k-1}^Δ and μ'^Δ_{k-1} are contained in the compact set C_{k-1}, hence the mixing probability measure λ_k, which is somehow arbitrary in the definition of the nonnegative kernel R_k^Δ, is not really involved in the procedure.

It follows immediately from Theorem 3.6 that

$$\|\mu_n^\Delta - \mu'^\Delta_n\| \leq \frac{2}{\log 3} \prod_{k=2}^{n} (1 - \varepsilon_k^2)\, \frac{1}{\varepsilon_1^2}\, \|\mu_0 - \mu'_0\| \ , \tag{9}$$

and

$$\|\mu_n^\Delta - \mu_n\| \leq \delta_n + \frac{2}{\log 3} \sum_{k=1}^{n-1} [\ \prod_{\ell=k+2}^{n} (1 - \varepsilon_\ell^2)\ ]\ \frac{\delta_k}{\varepsilon_{k+1}^2} \ , \tag{10}$$

where

$$\delta_k = \|\bar{R}_k(\mu_{k-1}) - \bar{R}_k^\Delta(\mu_{k-1})\| \ ,$$

for any $k \geq 1$.

Proposition 5.3. *For any* $k \geq 1$

$$\mathbb{E}[\delta_k] \leq 6\, \alpha(\Delta_{k-1}) \ ,$$

hence for k *large enough*

$$\delta_k \leq 6\, \frac{\alpha(\Delta_{k-1})}{a_k} \ , \qquad \text{a.s.}$$

where a_k *is the general term of an arbitrary converging series.*

The following result states that the sequence $\{\mu_n^\Delta \ , \ n \geq 0\}$ is exponentially stable and approximates the optimal filter $\{\mu_n \ , \ n \geq 0\}$ uniformly in time, provided that the observations are "sufficiently good", i.e. provided that Assumption A3 holds : this motivates the terminology "robust" filter used for the sequence $\{\mu_n^\Delta \ , \ n \geq 0\}$.

Theorem 5.4. *If Assumptions A1 and A2 hold, and if* $\Delta_k = \Delta$ *for any* $k \geq 1$*, then*

$$\|\mu_n^\Delta - \mu'^\Delta_n\| \leq \frac{2}{\log 3}\, (1 - \varepsilon^2(D(\Delta)))^{n-1}\, \frac{1}{\varepsilon^2(D(\Delta))}\, \|\mu_0 - \mu'_0\| \ ,$$

and

$$\mathbb{E}\|\mu_n^\Delta - \mu_n\| \leq 6\, (1 + \frac{2}{\varepsilon^4(D(\Delta)) \log 3})\, \alpha(\Delta) \ ,$$

which converges to zero as $\Delta \longrightarrow \infty$*, under Assumption A3.*

Notice that Theorem 5.4 provides an estimate of the approximation error averaged over observation sequences. To obtain an a.s. convergence result, we will rely on the following easy property.

Lemma 5.5. *If $\{\gamma_n\,,\, n \geq 1\}$ and $\{u_n\,,\, n \geq 1\}$ are two sequences of nonnegative numbers such that*

$$\lim_{n\to\infty} \sum_{k=1}^{n} \gamma_k = \infty \qquad \textit{and} \qquad \lim_{n\to\infty} u_n = 0 \;,$$

then

$$\lim_{n\to\infty} \sum_{k=1}^{n-1} [\; \prod_{\ell=k+2}^{n} (1-\gamma_\ell)\;]\;\; \gamma_{k+1}\, u_k = 0 \;.$$

Proof of Lemma 5.5. For any integer $n \geq 1$, define

$$F_n = \prod_{\ell=1}^{n} (1-\gamma_\ell) \;,$$

which converges to zero, and notice that

$$v_n = \sum_{k=1}^{n-1} [\; \prod_{\ell=k+2}^{n} (1-\gamma_\ell)\;]\;\; \gamma_{k+1}\, u_k = F_n\; \sum_{k=1}^{n-1} \frac{\gamma_{k+1}\, u_k}{F_{k+1}} \;.$$

For any $\varepsilon > 0$, there exists an integer n_ε such that $u_k \leq \varepsilon$ for any integer $k \geq n_\varepsilon$, hence

$$\sum_{k=1}^{n-1} \frac{\gamma_{k+1}\, u_k}{F_{k+1}} = \sum_{k=1}^{n_\varepsilon - 1} \frac{\gamma_{k+1}\, u_k}{F_{k+1}} + \sum_{k=n_\varepsilon}^{n-1} \frac{\gamma_{k+1}\, u_k}{F_{k+1}} \leq \sum_{k=1}^{n_\varepsilon - 1} \frac{\gamma_{k+1}\, u_k}{F_{k+1}} + \varepsilon \sum_{k=1}^{n-1} \frac{\gamma_{k+1}}{F_{k+1}} \;.$$

Finally, notice that

$$\frac{1}{F_{k+1}} - \frac{1}{F_k} = \frac{1}{F_{k+1}}\, [1 - (1-\gamma_{k+1})] = \frac{\gamma_{k+1}}{F_{k+1}} \;,$$

hence

$$\sum_{k=1}^{n-1} \frac{\gamma_{k+1}}{F_{k+1}} = \sum_{k=1}^{n-1} [\frac{1}{F_{k+1}} - \frac{1}{F_k}] = \frac{1}{F_n} - \frac{1}{F_1} \leq \frac{1}{F_n} \;,$$

and

$$v_n \leq F_n \sum_{k=1}^{n_\varepsilon - 1} \frac{\gamma_{k+1}\, u_k}{F_{k+1}} + \varepsilon \;.$$

Therefore $\limsup_{n\to\infty} v_n \leq \varepsilon$, and since $\varepsilon > 0$ is arbitrary, then v_n converges to zero. □

Theorem 5.6. *If Assumptions A1 and A2 hold, and if the sequence $\{\Delta_k\,,\, k \geq 1\}$ is such that*

$$\lim_{n\to\infty} \sum_{k=1}^{n} \varepsilon_k^2 = \infty \;,$$

then

$$\lim_{n\to\infty} \|\mu_n^\Delta - \mu_n'^\Delta\| = 0 \;.$$

If in addition

$$\lim_{k\to\infty} \frac{\alpha(\Delta_{k-1})}{a_k\, \varepsilon_{k+1}^4} = 0 \;, \tag{11}$$

where a_k is the general term of an arbitrary converging series, then almost surely

$$\lim_{k\to\infty} \frac{\delta_k}{\varepsilon_{k+1}^4} = 0 \qquad \textit{hence} \qquad \lim_{n\to\infty} \|\mu_n^\Delta - \mu_n\| = 0 \;.$$

Remark 5.7. Notice that

$$\frac{\alpha(\Delta_{k-1})}{\varepsilon^4_{k+1}} = \frac{\alpha(\Delta_{k-1})}{\varepsilon^4(D(\Delta_{k-1}))} \, \frac{\varepsilon^4_k}{\varepsilon^4_{k+1}} \; ,$$

hence under Assumption A3, a sufficient condition for (11) to hold is

$$\limsup_{k\to\infty} \frac{\varepsilon^4_k}{a_k \, \varepsilon^4_{k+1}} < \infty \; .$$

Proof of Proposition 5.3. For any $\mu, \mu' \in \mathcal{M}^+$ and any nonnegative function Λ defined on E such that $\langle \mu, \Lambda \rangle > 0$, the following estimate holds

$$\|\Lambda \cdot \mu - \Lambda \cdot \mu'\| \leq 2 \int_E \frac{\Lambda(x)}{\langle \mu, \Lambda \rangle} \, |\mu - \mu'|(dx) \; . \tag{12}$$

Using inequality (12) with $\Lambda = \Psi_k$ and

$$\mu = \mu_{k|k-1} = Q \, \mu_{k-1} \; = \; \mathbf{1}_{C_k} \, Q(\mathbf{1}_{C_{k-1}} \, \mu_{k-1}) + \mathbf{1}_{C_k} \, Q(\mathbf{1}_{C^c_{k-1}} \, \mu_{k-1}) + \mathbf{1}_{C^c_k} \, Q \, \mu_{k-1} \; ,$$

$$\mu' \; = \; \mathbf{1}_{C_k} \, Q(\mathbf{1}_{C_{k-1}} \, \mu_{k-1}) + \mu_{k-1}(C^c_{k-1}) \, \mathbf{1}_{C_k} \, \lambda_k \; ,$$

yields the following bound of the local error

$$\begin{aligned} \delta_k \; &\leq \; 2 \int_E \frac{\Psi_k(x')}{\langle \mu_{k|k-1}, \Psi_k \rangle} \, | \, \mathbf{1}_{C_k} \, Q(\mathbf{1}_{C^c_{k-1}} \, \mu_{k-1}) + \mathbf{1}_{C^c_k} \, Q \, \mu_{k-1} - \mu_{k-1}(C^c_{k-1}) \, \mathbf{1}_{C_k} \, \lambda_k \, |(dx') \\ &\leq \; 2 \int_E \frac{\Psi_k(x')}{\langle \mu_{k|k-1}, \Psi_k \rangle} \, [\, \mathbf{1}_{C_k}(x') \, Q(\mathbf{1}_{C^c_{k-1}} \, \mu_{k-1})(dx') + \mathbf{1}_{C^c_k}(x') \, Q \, \mu_{k-1}(dx') \\ &\qquad + \mu_{k-1}(C^c_{k-1}) \, \mathbf{1}_{C_k}(x') \, \lambda_k(dx') \,] \; . \end{aligned} \tag{13}$$

It follows from Remark 2.2 in [11] that for any test function ψ defined on F

$$\mathbb{E}[\, \frac{\psi(Y_k)}{\langle \mu_{k|k-1}, \Psi_k \rangle} \mid Y_{1:k-1}] = \int_F \psi(y) \, dy \; .$$

In particular, if $\psi(y) = g(y - h(x'))$, then $\psi(Y_k) = \Psi_k(x')$ and

$$\mathbb{E}[\, \frac{\Psi_k(x')}{\langle \mu_{k|k-1}, \Psi_k \rangle} \mid Y_{1:k-1}] = \int_F g(y - h(x')) \, dy = \int_F g(u) \, du = 1 \; ,$$

and if $\psi(y) = g(y - h(x')) \, \mathbf{1}_{(|y - h(x')| > \Delta_k)}$, then $\psi(Y_k) = \Psi_k(x') \, \mathbf{1}_{C^c_k}(x')$ and

$$\begin{aligned} \mathbb{E}[\, \frac{\Psi_k(x') \, \mathbf{1}_{C^c_k}(x')}{\langle \mu_{k|k-1}, \Psi_k \rangle} \mid Y_{1:k-1}] \; &= \; \int_F g(y - h(x')) \, \mathbf{1}_{(|y - h(x')| > \Delta_k)} \, dy \\ &= \; \int_F g(u) \, \mathbf{1}_{(|u| > \Delta_k)} \, du = \alpha(\Delta_k) \; , \end{aligned}$$

for any $x' \in E$. Notice that

$$\int_E Q(\mathbf{1}_{C^c_{k-1}} \, \mu_{k-1})(dx') = \int_E \int_E \mathbf{1}_{C^c_{k-1}}(x) \, \mu_{k-1}(dx) \, Q(x, dx') = \mu_{k-1}(C^c_{k-1}) \; ,$$

and

$$\begin{aligned} \mathbb{E}[\, \mu_{k-1}(C^c_{k-1}) \,] \; &= \; \mathbb{E}[\int_E \frac{\Psi_{k-1}(x) \, \mathbf{1}_{C^c_{k-1}}(x)}{\langle \mu_{k-1|k-2}, \Psi_{k-1} \rangle} \, \mu_{k-1|k-2}(dx) \,] \\ &= \; \mathbb{E}[\int_E \mathbb{E}[\, \frac{\Psi_{k-1}(x) \, \mathbf{1}_{C^c_{k-1}}(x)}{\langle \mu_{k-1|k-2}, \Psi_{k-1} \rangle} \mid Y_{1:k-2} \,] \, \mu_{k-1|k-2}(dx) \,] = \alpha(\Delta_{k-1}) \; . \end{aligned}$$

Now we can bound each term in the right–hand side of (13). The first term is bounded as follows

$$\begin{aligned}
&\mathbb{E}[\int_E \frac{\Psi_k(x')\,\mathbf{1}_{C_k}(x')}{\langle \mu_{k|k-1}, \Psi_k \rangle}\, Q(\mathbf{1}_{C^c_{k-1}}\, \mu_{k-1})(dx')\,] \\
&= \mathbb{E}[\int_E \mathbb{E}[\frac{\Psi_k(x')\,\mathbf{1}_{C_k}(x')}{\langle \mu_{k|k-1}, \Psi_k \rangle} \mid Y_{1:k-1}\,]\, Q(\mathbf{1}_{C^c_{k-1}}\, \mu_{k-1})(dx')\,] \\
&\leq \mathbb{E}[\int_E Q(\mathbf{1}_{C^c_{k-1}}\, \mu_{k-1})(dx')\,] = \alpha(\Delta_{k-1}) \;, \qquad (14)
\end{aligned}$$

the second term is estimated as follows

$$\begin{aligned}
&\mathbb{E}[\int_E \frac{\Psi_k(x')\,\mathbf{1}_{C^c_k}(x')}{\langle \mu_{k|k-1}, \Psi_k \rangle}\, \mu_{k|k-1}(dx')\,] \\
&= \mathbb{E}[\int_E \mathbb{E}[\frac{\Psi_k(x')\,\mathbf{1}_{C^c_k}(x')}{\langle \mu_{k|k-1}, \Psi_k \rangle} \mid Y_{1:k-1}\,]\, \mu_{k|k-1}(dx')\,] = \alpha(\Delta_k) \;, \qquad (15)
\end{aligned}$$

and finally, the last term is bounded as follows

$$\begin{aligned}
&\mathbb{E}[\int_E \frac{\Psi_k(x')\,\mathbf{1}_{C_k}(x')}{\langle \mu_{k|k-1}, \Psi_k \rangle}\, \mu_{k-1}(C^c_{k-1})\, \lambda_k(dx')\,] \\
&= \mathbb{E}[\int_E \mathbb{E}[\frac{\Psi_k(x')\,\mathbf{1}_{C_k}(x')}{\langle \mu_{k|k-1}, \Psi_k \rangle} \mid Y_{1:k-1}\,]\, \mu_{k-1}(C^c_{k-1})\, \lambda_k(dx')\,] \\
&\leq \mathbb{E}[\,\mu_{k-1}(C^c_{k-1})\,] = \alpha(\Delta_{k-1}) \;. \qquad (16)
\end{aligned}$$

Inserting the bounds (14), (15) and (16) into inequality (13), finally yields

$$\mathbb{E}[\delta_k] \leq 2\,\alpha(\Delta_k) + 4\,\alpha(\Delta_{k-1}) \leq 6\,\alpha(\Delta_{k-1}) \;.$$

The second part of the proof follows immediately from the Borel–Cantelli lemma. □

Example 5.8. To illustrate the results of Theorems 5.4 and 5.6, we consider the following signal / observation model, where $E = F = \mathbb{R}$ and

$$\begin{cases} X_{n+1} = f(X_n) + W_n \;, & X_0 \sim \mu_0 \;, \\ Y_n = h(X_n) + V_n \;, & \end{cases}$$

for all $n \geq 1$, where $\{W_n\,,\, n \geq 0\}$ and $\{V_n\,,\, n \geq 1\}$ are two independent sequences of i.i.d. random variables, with exponential probability densities q and g w.r.t. the Lebesgue measure, with standard deviations σ and s, respectively. More precisely, assume that

- the noise probability densities are given by

  $$q(w) = \frac{1}{2\sigma}\exp(-\frac{|w|}{\sigma}) \qquad \text{and} \qquad g(v) = \frac{1}{2s}\exp(-\frac{|v|}{s}) \;,$$

  for any $w, v \in \mathbb{R}$,

- the function f is Lipschitz continuous on $\mathbb{R}$, i.e. there exists a positive constant $a > 0$ such that

  $$|f(x) - f(x')| \leq a\,|x - x'| \;,$$

  for any $x, x' \in \mathbb{R}$,

- the function h is injective on $\mathbb{R}$ with Lipschitz inverse, i.e. there exists a positive constant $b > 0$ such that

  $$|x - x'| \leq b\,|h(x) - h(x')| \;,$$

  for any $x, x' \in \mathbb{R}$.

It follows from Example 4.3 and from Remarks 4.4 and 5.1, that Assumptions A1 and A2 are satisfied, and one can easily check that $\varepsilon^4(D(\Delta)) \geq 1/(16\,\sigma^4)\, e^{-8\,a\,b\,\Delta/\sigma}$ and $\alpha(\Delta) = \frac{1}{2}\, e^{-\Delta/s}$, hence

$$\frac{\alpha(\Delta)}{\varepsilon^4(D(\Delta))} \leq 8\,\sigma^4\, e^{-(\sigma/s-8\,a\,b)\,\Delta/\sigma} \; .$$

If $\sigma/s > 8\,a\,b$, i.e. if the observation noise variance is small enough w.r.t. the signal noise variance, then Assumption A3 is satisfied, and Theorem 5.4 provides an exponentially stable sequence that approximates the optimal filter uniformly in time.

If $\Delta_k = \sigma/(4\,a\,b)\,\log k$ for any $k \geq 1$, then $\varepsilon_k^2 \geq 1/(4\,\sigma^2)\, e^{-4\,a\,b\,\Delta_k/\sigma} = 1/(4\,\sigma^2\,k)$ and the series with general term ε_k^2 diverges. If in addition $\sigma/s > 12\,a\,b$, and $a_k = k^{-(1+\gamma)}$ for any $k \geq 1$, with $0 < \gamma < (\sigma/s - 12\,a\,b)/(4\,a\,b)$, then the series with general term a_k converges, and

$$\frac{\alpha(\Delta_{k-1})}{a_k\,\varepsilon_{k+1}^4} \leq 8\,\sigma^4\, \frac{(k+1)^2\,k^{1+\gamma}}{(k-1)^{3+\gamma}}\,(k-1)^{3+\gamma-\sigma/(4\,a\,b\,s)} \longrightarrow 0 \; ,$$

since $3+\gamma-\sigma/(4\,a\,b\,s) = \gamma - (\sigma/s - 12\,a\,b)/(4\,a\,b) < 0$. Therefore, it follows from Theorem 5.6 that the “robust” filter converges a.s. to the optimal filter, as time goes to infinity.

6 Stability of the optimal filter w.r.t. the initial condition

In this section, we use Theorems 5.4 and 5.6 to show that the optimal filter $\{\mu_n\,,\, n \geq 1\}$ inherits some of the stability properties of the “robust” filter $\{\mu_n^\Delta\,,\, n \geq 1\}$. The triangle inequality yields

$$\|\mu_n - \mu'_n\| \leq \|\mu_n - \mu_n^\Delta\| + \|\mu_n^\Delta - \mu_n'^\Delta\| + \|\mu_n'^\Delta - \mu'_n\| \; ,$$

and in addition to (9) and (10), it follows immediately from Theorem 3.6 that

$$\|\mu_n'^\Delta - \mu'_n\| \leq \delta'_n + \frac{2}{\log 3} \sum_{k=1}^{n-1} \Big[\prod_{\ell=k+2}^{n} (1-\varepsilon_\ell^2) \Big] \, \frac{\delta'_k}{\varepsilon_{k+1}^2} \; ,$$

where

$$\delta'_k = \|\bar{R}_k(\mu'_{k-1}) - \bar{R}_k^\Delta(\mu'_{k-1})\| \; ,$$

for any $k \geq 1$.

Proposition 6.1. *If μ_0 and μ'_0 are comparable, with $c_0 = h(\mu_0, \mu'_0) < \infty$, then for any $k \geq 1$*

$$\mathbb{E}[\delta'_k] \leq 6\,\exp(2\,c_0)\,\alpha(\Delta_{k-1}) \; ,$$

hence for k large enough

$$\delta'_k \leq 6\,\exp(2\,c_0)\,\frac{\alpha(\Delta_{k-1})}{a_k} \; , \qquad \text{a.s.}$$

where a_k is the general term of an arbitrary converging series.

The following result states that the optimal filter $\{\mu_n\,,\, n \geq 0\}$ is stable (but not exponentially stable) provided that the observations are “sufficiently good”, i.e. provided that Assumption A3 holds.

Theorem 6.2. *If Assumptions A1 and A2 hold, if $\Delta_k = \Delta$ for any $k \geq 1$, and if μ_0 and μ'_0 are comparable, with $c_0 = h(\mu_0, \mu'_0) < \infty$, then*

$$\begin{aligned} \mathbb{E}\|\mu_n - \mu'_n\| \;\leq\; & 6\,(1+\exp(2\,c_0))\,(1 + \frac{2}{\varepsilon^4(D(\Delta))\,\log 3})\,\alpha(\Delta) \\ & + \frac{2}{\log 3}\,(1-\varepsilon^2(D(\Delta)))^{n-1}\,\frac{1}{\varepsilon^2(D(\Delta))}\,\|\mu_0 - \mu'_0\| \; , \end{aligned} \tag{17}$$

hence

$$\lim_{n\to\infty} \mathbb{E}\|\mu_n - \mu'_n\| = 0 \; ,$$

under Assumption A3.

Theorem 6.3. *If Assumptions A1 and A2 hold, if the sequence* $\{\Delta_k\,,\ k \geq 1\}$ *is such that*

$$\lim_{n\to\infty} \sum_{k=1}^{n} \varepsilon_k^2 = \infty \ ,$$

if μ_0 *and* μ'_0 *are comparable, and if*

$$\lim_{k\to\infty} \frac{\alpha(\Delta_{k-1})}{a_k\, \varepsilon_{k+1}^4} = 0 \ ,$$

where a_k *is the general term of an arbitrary converging series, then almost surely*

$$\lim_{k\to\infty} \frac{\delta_k}{\varepsilon_{k+1}^4} = 0 \qquad \text{and} \qquad \lim_{k\to\infty} \frac{\delta'_k}{\varepsilon_{k+1}^4} = 0 \qquad \text{hence} \qquad \lim_{n\to\infty} \|\mu_n - \mu'_n\| = 0 \ .$$

Proof of Proposition 6.1. Using inequality (12) yields the following bound for the local error

$$\begin{aligned}\delta'_k \;\leq\; 2 \int_E \frac{\Psi_k(x')}{\langle \mu'_{k|k-1}, \Psi_k\rangle}\, [\, &\mathbf{1}_{C_k}(x')\, Q(\mathbf{1}_{C_{k-1}^c}\, \mu'_{k-1})(dx') + \mathbf{1}_{C_k^c}(x')\, Q\, \mu'_{k-1}(dx') \\ &+ \mu'_{k-1}(C_k^c)\, \mathbf{1}_{C_k^c}(x')\, \lambda_k(dx')\,]\ ,\end{aligned} \tag{18}$$

which we recognize to be inequality (13) where μ_{k-1} has been replaced by μ'_{k-1}. It is now sufficient to notice the following bound

$$\frac{\Psi_k(x')}{\langle \mu'_{k|k-1}, \Psi_k\rangle} = \frac{\Psi_k(x')}{\langle \mu_{k|k-1}, \Psi_k\rangle}\, \frac{\langle \mu_{k|k-1}, \Psi_k\rangle}{\langle \mu'_{k|k-1}, \Psi_k\rangle} \leq \frac{\Psi_k(x')}{\langle \mu_{k|k-1}, \Psi_k\rangle}\ \exp(c_0)\ , \tag{19}$$

valid for any integer $k \geq 1$: indeed, it follows from Lemma 3.4 that

$$\frac{\langle \mu_{k|k-1}, \Psi_k\rangle}{\langle \mu'_{k|k-1}, \Psi_k\rangle} = \frac{\langle Q\, \mu_{k-1}, \Psi_k\rangle}{\langle Q\, \mu'_{k-1}, \Psi_k\rangle} = \frac{(R_{k:1}\, \mu_0)(E)}{(R_{k-1:1}\, \mu_0)(E)}\, \frac{(R_{k-1:1}\, \mu'_0)(E)}{(R_{k:1}\, \mu'_0)(E)} \leq \exp(c_0)\ .$$

Notice that

$$\mathbb{E}[\, \mu'_{k-1}\,(C_{k-1}^c)\,] \leq \exp(c_0)\, \alpha(\Delta_{k-1})\ .$$

Now, just as in the proof of Proposition 5.3, we can bound each term in the right–hand side of (18), making use of the bound (19). The first term is bounded as follows

$$\mathbb{E}[\int_E \frac{\Psi_k(x')\, \mathbf{1}_{C_k}(x')}{\langle \mu'_{k|k-1}, \Psi_k\rangle}\ Q(\mathbf{1}_{C_{k-1}^c}\, \mu'_{k-1})(dx')\,] \leq \exp(2\, c_0)\, \alpha(\Delta_{k-1})\ , \tag{20}$$

the second term is bounded as follows

$$\mathbb{E}[\int_E \frac{\Psi_k(x')\, \mathbf{1}_{C_k^c}(x')}{\langle \mu'_{k|k-1}, \Psi_k\rangle}\ \mu'_{k|k-1}(dx')\,] \leq \exp(c_0)\, \alpha(\Delta_k)\ , \tag{21}$$

and finally, the last term is bounded as follows

$$\mathbb{E}[\int_E \frac{\Psi_k(x')\, \mathbf{1}_{C_k}(x')}{\langle \mu'_{k|k-1}, \Psi_k\rangle}\ \mu'_{k-1}(C_{k-1}^c)\, \lambda_k(dx')\,] \leq \exp(2\, c_0)\, \alpha(\Delta_{k-1})\ . \tag{22}$$

Inserting the bounds (20), (21) and (22) into inequality (18), finally yields

$$\mathbb{E}[\delta'_k] \leq 4\ \exp(2\, c_0)\, \alpha(\Delta_{k-1}) + 2\ \exp(c_0)\, \alpha(\Delta_k) \leq 6\ \exp(2\, c_0)\, \alpha(\Delta_{k-1})\ . \qquad \Box$$

Example 6.4. (Example 5.8 continued) Since $\Delta > 0$ is arbitrary in (17), we can take $\Delta = \alpha\, \sigma / (4\, a\, b)\, \log n$ with $0 < \alpha < 1$, hence $e^{4\, a\, b\, \Delta / \sigma} = n^\alpha$. If $\sigma / s > 8\, a\, b$, then introducing $\beta = (\sigma / s - 8\, a\, b) / (4\, a\, b)$ yields

$$\frac{\alpha(\Delta)}{\varepsilon^4(D(\Delta))} \leq 8\, \sigma^4\, e^{-(\sigma/s - 8\, a\, b)\, \Delta/\sigma} \leq 8\, \sigma^4\, e^{-4\, a\, b\, \beta\, \Delta/\sigma} \leq 8\, \sigma^4\, \frac{1}{n^{\beta\, \alpha}}\ ,$$

and

$$(1 - \varepsilon^2(D(\Delta)))^{n-1} \, \frac{1}{\varepsilon^2(D(\Delta))} \leq (1 - \frac{1}{4\,\sigma^2} \, e^{-4\,a\,b\,\Delta/\sigma})^{n-1} \, 4\,\sigma^2 \, e^{4\,a\,b\,\Delta/\sigma}$$

$$\leq 4\,\sigma^2 \, \frac{1}{n^{\beta\,\alpha}} \, (1 - \frac{1}{4\,\sigma^2\,n^\alpha})^{n-1} \, n^{(1+\beta)\,\alpha} = 4\,\sigma^2 \, \frac{1}{n^{\beta\,\alpha}} \, A_n \ ,$$

where

$$\log A_n = (n-1) \, \log(1 - \frac{1}{4\,\sigma^2\,n^\alpha}) + (1+\beta)\,\alpha \, \log n \ ,$$

can be bounded by a constant independent of n. Therefore, it follows from Theorem 6.2 that

$$\mathbb{E}\|\mu_n - \mu'_n\| \leq \frac{A}{n^{\beta\,\alpha}} \ ,$$

for any $0 < \alpha < 1$, where A is a positive constant independent of n, hence if the observation noise variance is small enough w.r.t. the signal noise variance, then the optimal filter forgets its initial condition at a rate which increases with the precision of the observation.

If in addition $\sigma/s > 12\,a\,b$, then it follows immediately from the Borel–Cantelli lemma that

$$\|\mu_n - \mu'_n\| \leq \frac{A}{n^\gamma} \ , \qquad \text{a.s.}$$

for any $0 < \gamma < (\sigma/s - 12\,a\,b)/(4\,a\,b)$, and for n large enough, where A is a positive constant independent of n.

This result confirms the idea that the optimal filter can have stability properties even when the signal is not ergodic, provided that the observations are "sufficiently good", just as in linear filtering, where detectability and stabilizability of the system are sufficient conditions for the Kalman filter to be exponentially stable, see e.g. Anderson and Moore [1, Chapter 4]. Another example is given in Budhiraja and Ocone [5], where the asymptotic stability of the nonlinear filter is proved for a nonergodic signal : however their proof uses an ergodic sum to control the propagation of the initial error, and it can only provide an asymptotic result, i.e. it cannot be used to prove the stability of the optimal filter w.r.t. model errors committed at each time step in the evolution of the filter (consider for instance the case of model misspecification). One interest of Theorem 5.6 is precisely to allow such stability results, and another application of it is given in the next section.

7 Uniform particle approximations to the optimal filter

In this section, we use Theorems 5.4 and 5.6 to construct particle filters that converge uniformly in time to the optimal filter, even though the mixing assumption does not hold. We shall use the notations of [11, Section 5] : in particular $S^N(\mu)$ is a shorthand notation for the empirical probability distribution of an N–sample with probability distribution μ, i.e.

$$S^N(\mu) = \frac{1}{N} \sum_{i=1}^N \delta_{\xi^i} \qquad \text{with} \qquad (\xi^1, \cdots, \xi^N) \text{ i.i.d.} \sim \mu,$$

and we recall the following classical results.

Lemma 7.1. *For any $\mu \in \mathcal{P}(E)$*

$$\sup_{\phi \,:\, \|\phi\|=1} \mathbb{E}|\,\langle S^N(\mu) - \mu, \phi\rangle\,| \leq \frac{1}{\sqrt{N}} \ .$$

For any nonnegative bounded measurable function Λ defined on E, and for any $\delta > 0$, let T denote the stopping time

$$T = \inf\{N \,:\, \delta^2 \sum_{i=1}^N \Lambda(\xi^i) \geq \sup_{x \in E} \Lambda(x)\} \qquad \textit{with} \qquad (\xi^1, \cdots, \xi^N, \cdots) \textit{ i.i.d.} \sim \mu \ .$$

If $\langle \mu, \Lambda\rangle > 0$, then T is a.s. finite, and

$$\sup_{\phi \,:\, \|\phi\|=1} \mathbb{E}|\,\langle \Lambda \cdot S^T(\mu) - \Lambda \cdot \mu, \phi\rangle\,| \leq 2\,\delta\,\sqrt{1+\delta^2} \ .$$

Remark 7.2. If in addition ϕ and μ are $\mathcal{F}$–measurable r.v.'s, and if conditionally w.r.t. $\mathcal{F}$ the r.v.'s $(\xi^1, \cdots, \xi^N, \cdots)$ are i.i.d. with (conditional) probability distribution μ, then the same estimates hold for conditional expectations w.r.t. $\mathcal{F}$, i.e.

$$\mathbb{E}[\,|\,\langle S^N(\mu) - \mu, \phi \rangle\,|\,|\,\mathcal{F}] \leq \frac{1}{\sqrt{N}}\;\|\phi\|\;, \tag{23}$$

and

$$\mathbb{E}[\,|\langle \Lambda \cdot S^T(\mu) - \Lambda \cdot \mu, \phi \rangle\,|\,|\,\mathcal{F}] \leq 2\,\delta\,\sqrt{1+\delta^2}\;\|\phi\|\;. \tag{24}$$

We consider again the signal / observation model introduced in Section 5. We have already seen that the "robust" filter $\{\mu_n^\Delta\,,\, n \geq 0\}$ approximates the optimal filter $\{\mu_n\,,\, n \geq 0\}$ uniformly in time. We propose to approximate the optimal filter by constructing a particle approximation $\{\mu_n^{\Delta,N}\,,\, n \geq 0\}$ to the "robust" filter : the idea is that the "robust" filter will be less sensitive than the optimal filter to the local errors induced by the particle approximation. Implicitly, it is assumed that the particle approximation at time n is based on a particle system of size N_n. The triangle inequality yields

$$\sup_{\phi\,:\,\|\phi\|=1} \mathbb{E}[\,|\,\langle \mu_n - \mu_n^{\Delta,N}, \phi \rangle\,|\,|\,Y_{1:n}] \leq \|\mu_n - \mu_n^\Delta\| + \sup_{\phi\,:\,\|\phi\|=1} \mathbb{E}[\,|\,\langle \mu_n^\Delta - \mu_n^{\Delta,N}, \phi \rangle\,|\,|\,Y_{1:n}]\;,$$

and in addition to (10), it follows immediately from Theorem 3.6 that

$$\sup_{\phi\,:\,\|\phi\|=1} \mathbb{E}[\,|\,\langle \mu_n^\Delta - \mu_n^{\Delta,N}, \phi \rangle\,|\,|\,Y_{1:n}\,] \leq \delta_n^{\Delta,N} + 2\,\frac{\delta_{n-1}^{\Delta,N}}{\varepsilon_n^2} + \frac{4}{\log 3}\sum_{k=1}^{n-2}[\,\prod_{\ell=k+3}^{n}(1-\varepsilon_\ell^2)\,]\;\frac{\delta_k^{\Delta,N}}{\varepsilon_{k+2}^2\,\varepsilon_{k+1}^2}\;,$$

where

$$\delta_k^{\Delta,N} = \sup_{\phi\,:\,\|\phi\|=1} \mathbb{E}[\,|\,\langle \mu_k^{\Delta,N} - \bar{R}_k^\Delta(\mu_{k-1}^{\Delta,N}), \phi \rangle\,|\,|\,Y_{1:k}\,]\;,$$

for any $k \geq 1$.

Consider first the robust version $\{\mu_n^{\Delta,N}\,,\, n \geq 0\}$ of the interacting particle filter approximation to the optimal filter $\{\mu_n\,,\, n \geq 0\}$, which is implemented according to the usual interacting particle filter algorithm, with the only difference that for any $n \geq 1$, the likelihood function Ψ_n is replaced by the truncated likelihood function $\Psi_n^\Delta = \mathbf{1}_{C_n}\,\Psi_n$. Initially $\mu_0^{\Delta,N} = \mu_0$, and the transition from $\mu_{n-1}^{\Delta,N}$ to $\mu_n^{\Delta,N}$ is described by the following diagram

$$\mu_{n-1}^{\Delta,N} \xrightarrow[\text{prediction}]{\text{sampled}} \mu_{n|n-1}^{\Delta,N} = S^{N_n}(Q\,\mu_{n-1}^{\Delta,N}) \xrightarrow[\text{correction}]{} \mu_n^{\Delta,N} = \Psi_n^\Delta \cdot \mu_{n|n-1}^{\Delta,N}\;.$$

In practice, the particle approximation

$$\mu_{n|n-1}^{\Delta,N} = \frac{1}{N_n}\sum_{i=1}^{N_n} \delta_{\xi_{n|n-1}^i}\;,$$

is completely characterized by the particle system $(\xi_{n|n-1}^1, \cdots, \xi_{n|n-1}^{N_n})$, and the transition from $(\xi_{n|n-1}^1, \cdots, \xi_{n|n-1}^{N_n})$ to $(\xi_{n+1|n}^1, \cdots, \xi_{n+1|n}^{N_{n+1}})$ consists of the following three steps.

(i) Correction : for all $i = 1, \cdots, N_n$, compute the weight

$$\omega_n^i = \frac{1}{c_n}\,\Psi_n^\Delta(\xi_{n|n-1}^i)\;,$$

where the normalization constant

$$c_n = \sum_{i=1}^{N_n} \Psi_n^\Delta(\xi_{n|n-1}^i)\;,$$

should be positive. Then set

$$\mu_n^{\Delta,N} = \Psi_n^\Delta \cdot \mu_{n|n-1}^{\Delta,N} = \sum_{i=1}^{N_n} \omega_n^i\,\delta_{\xi_{n|n-1}^i}\;.$$

(ii) Prediction : set

$$(Q\,\mu_n^{\Delta,N})(dx') = \sum_{i=1}^{N_n} \omega_n^i\, Q(\xi_{n|n-1}^i, dx')\ .$$

(iii) Resampling : independently for all $i = 1, \cdots, N_{n+1}$, generate a r.v. $\xi_{n+1|n}^i \sim Q\,\mu_n^{\Delta,N}$. Then set

$$\mu_{n+1|n}^{\Delta,N} = S^{N_{n+1}}(Q\,\mu_n^{\Delta,N}) = \frac{1}{N_{n+1}} \sum_{i=1}^{N_{n+1}} \delta_{\xi_{n+1|n}^i}\ .$$

In the correction step, particles are weighted according to their likelihood, i.e. to their adequation with the observation, and in the resampling step, particles with large weights are more likely to be selected than particles with small weights, hence the particle system concentrates automatically in regions of interest of the state space. To generate a sequence of independent r.v.'s according to the finite mixture probability distribution $Q\,\mu_n^{\Delta,N}$ is rather easy. This algorithm is not practical however, because the resampling mechanism is *blind*, i.e. does not use the next observation Y_{n+1} : as a result it could very well happen that *every* individual particle in the newly generated particle system $(\xi_{n+1|n}^1, \cdots, \xi_{n+1|n}^{N_{n+1}})$ falls outside the compact set C_{n+1}. To prevent this dramatic situation to happen, two alternative algorithms are proposed below : a robust version of the *adapted* particle filter, where the resampling mechanism takes the next observation into account, and a robust version of the sequential particle filter introduced in [11], respectively.

□ Robust adapted particle filter

Let $\{\mu_n^{\Delta,N}\,,\, n \geq 0\}$ denote the robust version of the adapted particle filter approximation to the optimal filter $\{\mu_n\,,\, n \geq 0\}$. Initially $\mu_0^{\Delta,N} = \mu_0$, and the transition from $\mu_{n-1}^{\Delta,N}$ to $\mu_n^{\Delta,N}$ is described by the following diagram

$$\mu_{n-1}^{\Delta,N} \xrightarrow[\text{prediction}]{\hspace{3em}} \mu_{n|n-1}^{\Delta,N} = Q\,\mu_{n-1}^{\Delta,N} \xrightarrow[\substack{\text{sampled}\\ \text{correction}}]{\hspace{3em}} \mu_n^{\Delta,N} = S^{N_n}(\Psi_n^\Delta \cdot \mu_{n|n-1}^{\Delta,N})\ .$$

In practice, the particle approximation

$$\mu_{n-1}^{\Delta,N} = \frac{1}{N_{n-1}} \sum_{i=1}^{N_{n-1}} \delta_{\xi_{n-1}^i}\ ,$$

is completely characterized by the particle system $(\xi_{n-1}^1, \cdots, \xi_{n-1}^{N_{n-1}})$ (which is contained in the compact set C_{n-1} by construction), and the transition from $(\xi_{n-1}^1, \cdots, \xi_{n-1}^{N_{n-1}})$ to $(\xi_n^1, \cdots, \xi_n^{N_n})$ consists of the following three steps.

(i) Prediction : set

$$\mu_{n|n-1}^{\Delta,N}(dx') = (Q\,\mu_{n-1}^{\Delta,N})(dx') = \frac{1}{N_{n-1}} \sum_{i=1}^{N_{n-1}} Q(\xi_{n-1}^i, dx')\ .$$

(ii) Correction : set

$$(\Psi_n^\Delta \cdot \mu_{n|n-1}^{\Delta,N})(dx') = \frac{1}{c_n} \sum_{i=1}^{N_{n-1}} Q(\xi_{n-1}^i, dx')\, \Psi_n^\Delta(x')\ ,$$

with the normalization constant

$$c_n = \int_E \sum_{i=1}^{N_{n-1}} Q(\xi_{n-1}^i, dx')\, \Psi_n^\Delta(x') = \sum_{i=1}^{N_{n-1}} \int_{C_n} Q(\xi_{n-1}^i, dx')\, \Psi_n(x')\ .$$

(iii) Resampling : independently for all $i = 1, \cdots, N_n$, generate a r.v. $\xi_n^i \sim \Psi_n^\Delta \cdot \mu_{n|n-1}^{\Delta,N}$, which automatically falls inside the compact set C_n. Then set

$$\mu_n^{\Delta,N} = \frac{1}{N_n} \sum_{i=1}^{N_n} \delta_{\xi_n^i} \; .$$

Under Assumptions A1 and A4, the normalization constant c_n is positive, hence $\Psi_n^\Delta \cdot \mu_{n|n-1}^{\Delta,N}$ is a well–defined probability distribution. To generate a sequence of independent (or dependent) r.v.'s according to this probability distribution can be done exactly (if not efficiently) using simple rejection, or approximately using importance resampling or Metropolis–Hastings importance resampling [3]. More efficient algorithms have been proposed in the literature for this purpose, using auxiliary variables [16] or local Monte Carlo methods [12], or introducing a regularization step [9, 13].

Proposition 7.3. *For any $k \geq 1$*

$$\delta_k^{\Delta,N} = \sup_{\phi \,:\, \|\phi\|=1} \mathbb{E}[\,|\,\langle \mu_k^{\Delta,N} - \bar{R}_k^\Delta(\mu_{k-1}^{\Delta,N}), \phi\rangle\,|\,|\,Y_{1:k}\,] \leq \frac{1}{\sqrt{N_k}} \; .$$

The following result states that the robust adapted particle filter $\{\mu_n^{\Delta,N}\,,\, n \geq 0\}$ converges uniformly in time to the optimal filter, when the parameter Δ grows to infinity.

Theorem 7.4. *If Assumptions A1, A2 and A4 hold, if $\Delta_k = \Delta$ and $\dfrac{1}{\sqrt{N_k}} \leq \varepsilon^2(D(\Delta))\,\alpha(\Delta)$ for any $k \geq 1$, then*

$$\sup_{\phi \,:\, \|\phi\|=1} \mathbb{E}[\,|\,\langle \mu_n - \mu_n^{\Delta,N}, \phi\rangle\,|\,] \;\leq\; 6\,(1 + \frac{2}{\varepsilon^4(D(\Delta))\log 3})\,\alpha(\Delta)$$
$$+\,(\,1 + \frac{2}{\varepsilon^2(D(\Delta))} + \frac{4}{\varepsilon^6(D(\Delta))\,\log 3}\,)\,\varepsilon^2(D(\Delta))\,\alpha(\Delta) \; ,$$

which converges to zero as $\Delta \to \infty$, under Assumption A3.

Theorem 7.5. *If Assumptions A1, A2 and A4 hold, if the sequence $\{\Delta_k\,,\, k \geq 1\}$ is such that*

$$\lim_{n\to\infty} \sum_{k=1}^{n} \varepsilon_k^2 = \infty \; ,$$

and if

$$\lim_{k\to\infty} \frac{1}{\sqrt{N_k}\,\varepsilon_{k+2}^2\,\varepsilon_{k+1}^4} = 0 \qquad \textit{and} \qquad \lim_{k\to\infty} \frac{\alpha(\Delta_{k-1})}{a_k\,\varepsilon_{k+1}^4} = 0 \; ,$$

where a_k is the general term of an arbitrary converging series, then almost surely

$$\lim_{n\to\infty} \sup_{\phi \,:\, \|\phi\|=1} \mathbb{E}[\,|\,\langle \mu_n - \mu_n^{\Delta,N}, \phi\rangle\,|\,|\,Y_{1:n}\,] = 0 \; .$$

Proof of Proposition 7.3. Using estimate (23) with $\mu = \bar{R}_k^\Delta(\mu_{k-1}^{\Delta,N})$ and $\mathcal{F} = \sigma(Y_{1:k}\,,\, \mu_{k-1}^{\Delta,N})$ yields

$$\mathbb{E}[\,|\,\langle \mu_k^{\Delta,N} - \bar{R}_k^\Delta(\mu_{k-1}^{\Delta,N}), \phi\rangle\,|\,|\,Y_{1:k}\,,\, \mu_{k-1}^{\Delta,N}]$$
$$= \mathbb{E}[\,|\,\langle S^{N_k}(\bar{R}_k^\Delta(\mu_{k-1}^{\Delta,N})) - \bar{R}_k^\Delta(\mu_{k-1}^{\Delta,N}), \phi\rangle\,|\,|\,Y_{1:k}\,,\, \mu_{k-1}^{\Delta,N}]$$
$$\leq \frac{1}{\sqrt{N_k}}\,\|\phi\| \; . \qquad \Box$$

□ Robust sequential particle filter

The sequential particle filter $\{\mu_n^{\Delta,N}\ ,\ n \geq 0\}$ associated with the "robust" filter is implemented according to the sequential particle filter algorithm described in [11], with the only difference that for any $n \geq 1$, the likelihood function Ψ_n is replaced by the truncated likelihood function $\Psi_n^\Delta = \mathbf{1}_{C_n}\, \Psi_n$. The result is that in the robust version, particles with small weights (i.e. particles that are located outside C_n) are systematically eliminated because of the truncation of the likelihood function. At time n, the robust sequential particle filter $\mu_n^{\Delta,N}$ uses a random number N_n of particles, and the sequential procedure insures that the number of particles in the region of interest, i.e. in C_n, is not zero. Initially $\mu_0^{\Delta,N} = \mu_0$, and the transition from $\mu_{n-1}^{\Delta,N}$ to $\mu_n^{\Delta,N}$ is described by the following diagram

$$\mu_{n-1}^{\Delta,N} \xrightarrow[\substack{\text{sequential}\\ \text{sampled}\\ \text{prediction}}]{} \mu_{n|n-1}^{\Delta,N} = S^{N_n}(Q\,\mu_{n-1}^{\Delta,N}) \xrightarrow[\text{correction}]{} \mu_n^{\Delta,N} = \Psi_n^\Delta \cdot \mu_{n|n-1}^{\Delta,N}\ .$$

In practice, the particle approximation

$$\mu_{n|n-1}^{\Delta,N} = \frac{1}{N_n} \sum_{i=1}^{N_n} \delta_{\xi_{n|n-1}^i}\ ,$$

is completely characterized by the particle system $(\xi_{n|n-1}^1, \cdots, \xi_{n|n-1}^{N_n})$, and the transition from $(\xi_{n|n-1}^1, \cdots, \xi_{n|n-1}^{N_n})$ to $(\xi_{n+1|n}^1, \cdots, \xi_{n+1|n}^{N_{n+1}})$ consists of the following three steps.

(i) Correction : for all $i = 1, \cdots, N_n$, compute the weight

$$\omega_n^i = \frac{1}{c_n}\, \Psi_n^\Delta(\xi_{n|n-1}^i)\ ,$$

with the normalization constant

$$c_n = \sum_{i=1}^{N_n} \Psi_n^\Delta(\xi_{n|n-1}^i)\ .$$

Then set

$$\mu_n^{\Delta,N} = \Psi_n^\Delta \cdot \mu_{n|n-1}^{\Delta,N} = \sum_{i=1}^{N_n} \omega_n^i\, \delta_{\xi_{n|n-1}^i}\ .$$

(ii) Prediction : set

$$(Q\,\mu_n^{\Delta,N})(dx') = \sum_{i=1}^{N_n} \omega_n^i\, Q(\xi_{n|n-1}^i, dx')\ .$$

(iii) Sequential resampling : independently for all $i = 1, \cdots, N_{n+1}$, generate a r.v. $\xi_{n+1|n}^i \sim Q\,\mu_n^{\Delta,N}$, where the random number N_{n+1} of particles is defined as the stopping time

$$N_{n+1} = \inf\{N\ :\ \delta_{n+1}^2 \sum_{i=1}^{N} \Psi_{n+1}^\Delta(\xi_{n+1|n}^i) \geq \sup_{x \in E} \Psi_{n+1}^\Delta(x)\}\ ,$$

and set

$$\mu_{n+1|n}^{\Delta,N} = S^{N_{n+1}}(Q\,\mu_n^{\Delta,N}) = \frac{1}{N_{n+1}} \sum_{i=1}^{N_{n+1}} \delta_{\xi_{n+1|n}^i}\ .$$

Remark 7.6. Notice that we could alternatively use

$$\bar{\mu}^{\Delta,N}_{n+1|n} = \frac{\displaystyle\sum_{i=1}^{N_{n+1}} \mathbf{1}_{C_{n+1}}(\xi^i_{n+1|n})\, \delta_{\xi^i_{n+1|n}}}{\displaystyle\sum_{i=1}^{N_{n+1}} \mathbf{1}_{C_{n+1}}(\xi^i_{n+1|n})} \; ,$$

i.e. discard immediately those particles which fall outside the compact set C_{n+1}, because in the next step

$$\Psi^\Delta_{n+1} \cdot \bar{\mu}^{\Delta,N}_{n+1|n} = \Psi^\Delta_{n+1} \cdot \mu^{\Delta,N}_{n+1|n} \; .$$

Proposition 7.7. *If Assumptions A1 and A4 hold, then for any $k \geq 1$ the random number N_k of particles is a.s. finite, and*

$$\delta^{\Delta,N}_k = \sup_{\phi \,:\, \|\phi\|=1} \mathbb{E}[\, | \langle \mu^{\Delta,N}_k - \bar{R}^\Delta_k(\mu^{\Delta,N}_{k-1}), \phi \rangle | \mid Y_{1:k} \,] \leq 2\, \delta_k \sqrt{1+\delta_k^2} \; .$$

The following result states that the robust sequential particle filter $\{\mu^{\Delta,N}_n \, , \, n \geq 0\}$ converges uniformly in time to the optimal filter, when the parameter Δ grows to infinity.

Theorem 7.8. *If Assumptions A1, A2 and A4 hold, if $\Delta_k = \Delta$ and $2\, \delta_k \sqrt{1+\delta_k^2} \leq \varepsilon^2(D(\Delta))\, \alpha(\Delta)$ for any $k \geq 1$, then*

$$\begin{aligned} \sup_{\phi \,:\, \|\phi\|=1} \mathbb{E}[\, | \langle \mu_n - \mu^{\Delta,N}_n, \phi \rangle | \,] \; &\leq \; 6\, (1 + \frac{2}{\varepsilon^4(D(\Delta)) \log 3})\, \alpha(\Delta) \\ &\quad + (\, 1 + \frac{2}{\varepsilon^2(D(\Delta))} + \frac{4}{\varepsilon^6(D(\Delta)) \, \log 3}\,)\, \varepsilon^2(D(\Delta))\, \alpha(\Delta) \; , \end{aligned}$$

which converges to zero as $\Delta \to \infty$, under Assumption A3.

Theorem 7.9. *If Assumptions A1, A2 and A4 hold, if the sequence $\{\Delta_k \, , \, k \geq 1\}$ is such that*

$$\lim_{n \to \infty} \sum_{k=1}^n \varepsilon_k^2 = \infty \; ,$$

and if

$$\lim_{k \to \infty} \frac{\delta_k}{\varepsilon^2_{k+2}\, \varepsilon^4_{k+1}} = 0 \qquad \textit{and} \qquad \lim_{k \to \infty} \frac{\alpha(\Delta_{k-1})}{a_k\, \varepsilon^4_{k+1}} = 0 \; ,$$

where a_k is the general term of an arbitrary converging series, then almost surely

$$\lim_{n \to \infty} \sup_{\phi \,:\, \|\phi\|=1} \mathbb{E}[\, | \langle \mu_n - \mu^{\Delta,N}_n, \phi \rangle | \mid Y_{1:n} \,] = 0 \; .$$

Proof of Proposition 7.7. Under Assumptions A1 and A4

$$\langle Q\, \mu^{\Delta,N}_{k-1}, \Psi^\Delta_k \rangle \geq \varepsilon_k \int_{C_k} \Psi_k(x')\, \lambda_k(dx') > 0 \; ,$$

hence the number N_k of particles at time k is a.s. finite. Using estimate (24) with $\Lambda = \Psi^\Delta_k$, $\mu = Q\, \mu^{\Delta,N}_{k-1}$ and $\mathcal{F} = \sigma(Y_{1:k} \, , \, \mu^{\Delta,N}_{k-1})$ yields

$$\begin{aligned} &\mathbb{E}[\, | \langle \mu^{\Delta,N}_k - \bar{R}^\Delta_k(\mu^{\Delta,N}_{k-1}), \phi \rangle | \mid Y_{1:k} \, , \, \mu^{\Delta,N}_{k-1}] \\ &= \mathbb{E}[\, | \langle \Psi^\Delta_k \cdot S^{N_k}(Q\, \mu^{\Delta,N}_{k-1}) - \Psi^\Delta_k \cdot (Q\, \mu^{\Delta,N}_{k-1}), \phi \rangle | \mid Y_{1:k} \, , \, \mu^{\Delta,N}_{k-1}] \\ &\leq 2\, \delta_k \sqrt{1+\delta_k^2} \; . \qquad \Box \end{aligned}$$

References

[1] B. D. O. Anderson and J. B. Moore. *Optimal Filtering*. Prentice–Hall, Englewood Cliffs, NJ, 1979.

[2] R. Atar and O. Zeitouni. Exponential stability for nonlinear filtering. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 33(6):697–725, 1997.

[3] C. Berzuini, N. G. Best, W. R. Gilks, and C. Larizza. Dynamic conditional independence models and Markov chain Monte Carlo methods. *Journal of the American Statistical Association*, 92(440):1403–1412, Dec. 1997.

[4] A. Budhiraja and D. L. Ocone. Exponential stability of discrete–time filters for bounded observation noise. *Systems and Control Letters*, 30(4):185–193, 1997.

[5] A. Budhiraja and D. L. Ocone. Exponential stability of discrete–time filters for non–ergodic signals. *Stochastic Processes and their Applications*, 82(2):245–257, Aug. 1999.

[6] G. Da Prato, M. Fuhrman, and P. Malliavin. Asymptotic ergodicity for the Zakai filtering equation. *Comptes Rendus de l'Académie des Sciences, Série I, Mathématique*, 321(5):613–616, 1995.

[7] G. Da Prato, M. Fuhrman, and P. Malliavin. Asymptotic ergodicity of the process of conditional law in some problem of nonlinear filtering. *Journal of Functional Analysis*, 164(2):356–377, 1999.

[8] P. Del Moral and A. Guionnet. On the stability of interacting processes with applications to filtering and genetic algorithms. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 37(2):155–194, 2001.

[9] M. Hürzeler and H. R. Künsch. Monte Carlo approximations for general state space models. *Journal of Computational and Graphical Statistics*, 7(2):175–193, June 1998.

[10] H. Kunita. Ergodic properties of nonlinear filtering processes. In K. Alexander and J. Watkins, editors, *Spatial Stochastic Processes : Festschrift in honor of T.E. Harris*, volume 19 of *Progress in Probability*, pages 233–256. Birkhäuser, Boston, 1991.

[11] F. Le Gland and N. Oudjane. Stability and uniform approximation of nonlinear filters using the Hilbert metric, and application to particle filters. Rapport de Recherche 4215, INRIA, June 2001. `ftp://ftp.inria.fr/INRIA/publication/publi-ps-gz/RR/RR-4215.ps.gz`.

[12] J. S. Liu and R. Chen. Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, 93(443):1032–1044, Sept. 1997.

[13] C. Musso, N. Oudjane, and F. Le Gland. Improving regularized particle filters. In A. Doucet, N. de Freitas, and N. Gordon, editors, *Sequential Monte Carlo Methods in Practice*, Statistics for Engineering and Information Science, chapter 12, pages 247–271. Springer–Verlag, New York, 2001.

[14] D. L. Ocone and E. Pardoux. Asymptotic stability of the optimal filter with respect to its initial condition. *SIAM Journal on Control and Optimization*, 34(1):226–243, Jan. 1996.

[15] N. Oudjane and S. Rubenthaler. Stability and uniform particle approximation of nonlinear filters in case of nonergodic signal. (in preparation).

[16] M. K. Pitt and N. Shephard. Filtering via simulations : auxiliary particle filter. *Journal of the American Statistical Association*, 94(446):590–599, June 1999.

[17] Ł. Stettner. On invariant measures of filtering processes. In N. Christopeit, K. Helmes, and M. Kohlmann, editors, *Stochastic Differential Systems, Bad Honnef 1988*, volume 126 of *Lecture Notes in Control and Information Sciences*, pages 279–292. Springer–Verlag, Berlin, 1989.

Contents

1 Introduction **1**

2 Signal and observation model **1**

3 Stability and robustification **3**

4 Pseudo–mixing signals **6**

5 Approximation of the optimal filter by a “robust” filter **8**

6 Stability of the optimal filter w.r.t. the initial condition **15**

7 Uniform particle approximations to the optimal filter **17**

Robust adapted particle filter (R–APF) 19

Robust sequential particle filter (R–SPF) 21

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
http://www.inria.fr
ISSN 0249-6399