LONG-TERM STABILITY OF SEQUENTIAL MONTE CARLO METHODS UNDER VERIFIABLE CONDITIONS

BY RANDEL DOUC¹, ERIC MOULINES¹ AND JIMMY OLSSON²

Institut Télécom/Télécom SudParis, Institut Télécom/Télécom ParisTech and KTH Royal Institute of Technology

This paper discusses particle filtering in general hidden Markov models (HMMs) and presents novel theoretical results on the long-term stability of bootstrap-type particle filters. More specifically, we establish that the asymptotic variance of the Monte Carlo estimates produced by the bootstrap filter is uniformly bounded in time. On the contrary to most previous results of this type, which in general presuppose that the state space of the hidden state process is compact (an assumption that is rarely satisfied in practice), our very mild assumptions are satisfied for a large class of HMMs with possibly noncompact state space. In addition, we derive a similar time uniform bound on the asymptotic $L^2$ error. Importantly, our results hold for misspecified models; that is, we do not at all assume that the data entering into the particle filter originate from the model governing the dynamics of the particles or not even from an HMM.

1. Introduction. This paper deals with estimation in general hidden Markov models (HMMs) via sequential Monte Carlo (SMC) methods (or particle filters). More specifically, we present novel results on the numerical stability of the bootstrap particle filter that hold under very general and easily verifiable assumptions. Before stating the results we provide some background.

Consider an HMM $(X_n,Y_n)_{n \in \mathbb{N}}$, where the Markov chain (or state sequence) $(X_n)_{n \in \mathbb{N}}$, taking values in some general state space $(\mathcal{X},\mathcal{X})$, is only

¹Supported by the Agence Nationale de la Recherche through the 2009–2012 project Big MC.
²Supported by the Swedish Research Council, Grant 2011-5577.

AMS 2000 subject classifications. Primary 62M09; secondary 62F12.

Key words and phrases. Asymptotic variance, general hidden Markov models, local Doeblin condition, bootstrap particle filter, sequential Monte Carlo methods, time uniform convergence.

This is an electronic reprint of the original article published by the Institute of Mathematical Statistics in The Annals of Applied Probability, 2014, Vol. 24, No. 5, 1767–1802. This reprint differs from the original in pagination and typographic detail.
partially observed through the sequence \((Y_n)_{n \in \mathbb{N}}\) of observations taking values in another general state space \((Y, \mathcal{Y})\). More specifically, conditionally on the state sequence \((X_n)_{n \in \mathbb{N}}\), the observations are assumed to be conditionally independent and such that the conditional distribution of each \(Y_n\) depends on the corresponding state \(X_n\) only; see, for example, [2] and the references therein. We denote by \(Q\) and \(\chi\) the kernel and initial distribution of \((X_n)_{n \in \mathbb{N}}\), respectively. Even though \(n\) is not necessarily a temporal index, it will in the following be referred to as “time.”

Any kind of statistical estimation in HMMs typically involves computation of the conditional distribution of one or several hidden states given a set of observations. Of particular interest are the filter distributions, where the filter distribution \(\bar{\phi}_\chi(Y_n^0)\) at time \(n\) is defined as the conditional distribution of \(X_n\) given the corresponding observation history \(Y_n^0 = (Y_0, \ldots, Y_n)\) (this will be our generic notation for vectors). The problem of computing, recursively in \(n\) and in a single sweep of the data, the sequence of filter distributions is referred to as optimal filtering. Of similar interest are the predictor distributions, where the predictor distribution \(\phi_\chi(Y_{n-1}^0)\) at time \(n\) is defined as the conditional distribution of the state \(X_n\) given the preceding observation history \(Y_{n-1}^0\) (more precise definitions of filter and predictor distributions are given in Section 2). In this paper we focus on the computation of these distributions, which can be carried through in a recursive fashion according to

\[
\begin{align*}
(1) & \quad \bar{\phi}_\chi(Y_n^0)(A) = \frac{\int 1_A(x)g(Y_n)(x)\phi_\chi(Y_{n-1}^0)(dx)}{\int g(Y_n)(x)\phi_\chi(Y_{n-1}^0)(dx)}, \quad A \in \mathcal{X}, \\
(2) & \quad \phi_\chi(Y_{n-1}^0)(A) = \int Q(x,A)\bar{\phi}_\chi(Y_n^0)(dx), \quad A \in \mathcal{X},
\end{align*}
\]

where \(g(Y_n)\) is the local likelihood of the hidden state \(X_n\) given the observation \(Y_n\). Steps (1) and (2) are typically referred to as correction and prediction, respectively. In the correction step, the predictor \(\phi_\chi(Y_{n-1}^0)\) is, as the new observation \(Y_n\) becomes available, weighted with the local likelihood, providing the filter \(\bar{\phi}_\chi(Y_n^0)\); in the prediction step, the filter \(\bar{\phi}_\chi(Y_n^0)\) is propagated through the dynamics \(Q\) of the latent Markov chain, yielding the predictor \(\phi_\chi(Y_{n-1}^0)\) at the consecutive time step. The correction and prediction steps form jointly a measure-valued mapping \(\Phi\) generating recursively the predictor distribution flow according to

\[
\phi_\chi(Y_0^0) = \Phi(Y_n)(\phi_\chi(Y_{n-1}^0))
\]

(we refer again to Section 2 for precise definitions).

Unless the HMM is either a linear Gaussian model or a model comprising only a finite number of possible states, exact numeric computation of the predictor distributions is in general infeasible. Thus, one is generally
confined to using finite-dimensional approximations of these measures, and in this paper we concentrate on the use of particle filters for this purpose. A particle filter approximates the predictor distribution at time $n$ by the empirical measure $\phi_N^X(Y_{n-1}^n)$ associated with a finite sample $(\xi_i^n)_{i=1}^N$ of particles evolving randomly and recursively in time. At each iteration of the algorithm, the particle sample is updated through a selection step and a mutation step, corresponding directly to correction and prediction, respectively. The selection operation duplicates or eliminates particles with high or low posterior probability, respectively, while the mutation operation disseminates randomly the particles in the state space for further selection at the next iteration. The most basic algorithm—proposed in [19] and referred to as the bootstrap particle filter—performs selection by resampling multinomially the predictive particles $(\xi_i^n)_{i=1}^N$ with probabilities proportional to the local likelihood $\omega_i^n = g(Y_n)(\xi_i^n)$ of the particle locations; after this, the selected particle swarm is mutated according to the dynamics $Q$ of the latent Markov chain. Here the self-normalized importance sampling estimator $\bar{\phi}_N^X(Y_0^n) \chi(Y_{n-1}^0)$ associated with the weighted particle sample $(\xi_i^n, \omega_i^n)_{i=1}^N$ provides an approximation of the filter $\bar{\phi}(Y_0^n)$. Thus, subjecting the particle sample $(\xi_i^n)_{i=1}^N$ to selection and mutation is in the case of the bootstrap particle filter equivalent to drawing, conditionally independently given $(\xi_i^n)_{i=1}^N$, new particles $(\xi_i^{n+1})_{i=1}^N$ from the distribution $\Phi(Y_n)(\phi_N^X(Y_{n-1}^n))$ obtained by plugging the empirical measure $\phi_N^X(Y_{n-1}^n)$ into the filter recursion, which we denote

$$(3) \quad (\xi_{i+1}^n)_{i=1}^N \overset{\text{i.i.d.}}{\sim} \Phi(Y_n)(\phi_N^X(Y_{n-1}^n)) \otimes N.$$  

Since the seminal paper [19], particle filters have been successfully applied to nonlinear filtering problems in many different fields; we refer to the collection [17] for an introduction to particle filtering in general and for miscellaneous examples of real-life applications.

The theory of particle filtering is an active field and there is a number of available convergence results concerning, for example, $L^p$ error bounds and weak convergence; see the monographs [1, 5] and the references therein. Most of these results establish the convergence, as the number of particles $N$ tends to infinity, of the particle filter for a fixed time step $n \in \mathbb{N}$. For infinite time horizons, that is, when $n$ tends to infinity, convergence is less obvious. Indeed, each recursive update (3) of the particles $(\xi_i^n)_{i=1}^N$ is based on the implicit assumption that the empirical measure $\phi_N^X(Y_{n-1}^n)$ associated with the ancestor sample approximates perfectly well the predictor $\phi_X^Y(Y_0^n)$ at the previous time step; however, since the ancestor sample is marred by an error itself, one may expect that the errors induced at the different updating steps accumulate and, consequently, that the total error propagated through the algorithm increases with $n$. This would make the algorithm useless in practice. Fortunately, it has been observed empirically by several authors
(see, e.g., [34], Section 1.1) that the convergence of particle filters appears to be uniform in time also for very general HMMs. Nevertheless, even though long-term stability is essential for the applicability of particle filters, most existing time uniform convergence results are obtained under assumptions that are generally not met in practice. The aim of the present paper is thus to establish the infinite time-horizon stability under mild and easy verifiable assumptions, satisfied by most models for which the particle filter has been found to be useful.

1.1. Previous work. The first time uniform convergence result for bootstrap-type particle filters was obtained by Del Moral and Guionnet [7] (see also [5] for refinements) using contraction properties of the mapping $\Phi$. We recall in some detail this technique. By writing

$$\phi^N_X(Y^n_0) - \phi_X(Y^n_0) = \underbrace{\phi^N_X(Y^n_0) - \Phi(Y_n)(\phi^{N-1}_X(Y^n_0))}_{\text{sampling error}} + \underbrace{\Phi(Y_n)(\phi^{N-1}_X(Y^n_0)) - \Phi(Y_n)(\phi^{N-1}_X(Y^n_0))}_{\text{initialization error}}$$

one may decompose the error $\phi^N_X(Y^n_0) - \phi_X(Y^n_0)$ into a first error (the sampling error) introduced by replacing $\Phi(Y_n)(\phi^{N-1}_X(Y^n_0))$ by its empirical estimate $\phi^N_X(Y^n_0)$ and a second error (the initialization error) originating from the discrepancy between empirical measure $\phi^N_X(Y^{n-1}_0)$ associated with the ancestor particles and the true predictor $\phi_X(Y^{n-1}_0)$. The sampling error is easy to control. One may, for example, use the Marcinkiewicz–Zygmund inequality to bound the $L^p$ error by $cN^{-1/2}$, where $c \in \mathbb{R}_+$ is a universal constant. Exponential deviation inequalities may also be obtained. For the initialization error, we may expect that the mapping $\Phi(Y_n)$ is in some sense contracting and thus downscales the discrepancy between $\phi^N_X(Y^{n-1}_0)$ and $\phi_X(Y^{n-1}_0)$. This is the point where the exponential forgetting of the predictor distribution becomes crucial. Assume, for instance, that there exists a constant $\rho \in [0, 1[$ such that $\|\Phi(Y^n_0)(\mu) - \Phi(Y^n_0)(\nu)\| \leq \rho^{n-m+1}\|\mu - \nu\|$ for any integers $0 \leq m \leq n$ and any probability measures $\mu$ and $\nu$, where $\| \cdot \|$ is some suitable norm on the space of probability measures and $\Phi(Y^n_0) \triangleq \Phi(Y^n_0) \circ \Phi(Y_{n-1}) \circ \cdots \circ \Phi(Y_m)$. Since $\Phi(Y^n_0)(\mu)$ is the predictor distribution $\phi^N_Y(Y^n_0)$ obtained when the hidden chain is initialized with the distribution $\mu$ at time $m$, this means that the predictor distribution forgets the initial distribution geometrically fast. In addition, the forgetting rate $\rho$ is uniform with respect to the observations. The uniformity with respect to the observations is of course the main reason why the assumptions on the model are so stringent.
Now, decomposing similarly also the initialization error and proceeding recursively yields the telescoping sum

\[
\phi_X^n(Y_0^n) - \phi_X(Y_0^n) = \phi_X^n(Y_0^n) - \Phi(Y_n)\langle \phi_X^n(Y_0^{n-1}) \rangle \\
+ \sum_{k=1}^{n-1} (\Phi(Y_{k+1}^n)(\phi_X^n(Y_k^n)) - \Phi(Y_{k+1}^n) \circ \Phi(Y_k^n)(\phi_X^n(Y_k^{k-1}))) \\
+ \Phi(Y_1^n)(\phi_X^n(Y_0)) - \Phi(Y_1^n)(\phi_X(Y_0)).
\]

Now each term of the sum above can be viewed as a downscaling (by a factor \(\rho^{n-k}\)) of the sampling error between \(\phi_X^n(Y_k^n)\) and \(\Phi(Y_k^n)(\phi_X^n(Y_k^{k-1}))\) through the contraction of \(\Phi(Y_{k+1}^n)\). Denoting by \(\delta_n\) the L^p error of \(\phi_X^n(Y_0^n)\) and assuming that the initial sample is obtained through standard importance sampling, implying that \(\delta_0 \leq cN^{-1/2}\), provides, using the contraction of \(\Phi(Y_{k+1}^n)\), the uniform L^p error bound \(\delta_n \leq cN^{-1/2} \sum_{k=0}^{n-1} \rho^{n-k} \leq cN^{-1/2}(1 - \rho)^{-1}\).

Even though this result is often used as a general guideline on particle filter stability, it relies nevertheless heavily on the assumption that the kernel \(Q\) of hidden Markov chain satisfies the following strong mixing condition, which is even more stringent than the already very strong one-step global Doeblin condition: there exist constants \(\epsilon^+ > \epsilon^- > 0\) and a probability measure \(\nu\) on \((X, \mathcal{X})\) such that for all \(x \in X\) and \(A \in \mathcal{X}\),

\[
\epsilon^+ \nu(A) \leq Q(x, A) \leq \epsilon^- \nu(A).
\]

This assumption, which in particular implies that the Markov chain is uniformly geometrically ergodic, restricts the applicability of the stability result in question to models where the state space \(X\) is small. For Markov chains on separable metric spaces, provided that the kernel is strongly Feller, condition (5) typically requires the state space to be compact. Some refinements have been obtained in, for example, [2, 4, 5, 16, 20, 25–28, 33].

The long-term stability of particle filters is also related to the boundedness of the asymptotic variance. The first central limit theorem (CLT) for bootstrap-type particle filters was derived by Del Moral and Guionnet [6]. More specifically, it was shown that the normalized Monte Carlo error \(\sqrt{N}(\phi_X^n(Y_0^{n-1})h - \phi_X(Y_0^{n-1})h)\) tends weakly, for a fixed \(n \in \mathbb{N}^+\) and as the particle population size \(N\) tends to infinity, to a zero mean normally distributed variable with variance \(\sigma^2_X(Y_0^{n-1})(h)\). Here we have used the notation \(\mu h \overset{\Delta}{=} \int h(x)\mu(dx)\) to denote expectations. The original proof of the CLT was later simplified and extended to more general particle filtering algorithms in [3, 14, 16, 21, 24]; in Section 2 we recall in detail the version obtained in [14] and provide an explicit expression of the asymptotic
variance $\sigma_\chi^2(Y_0^{n-1})(h)$. As shown first by [7], Theorem 3.1, it is possible, using the strong mixing assumption described above, to bound uniformly also the asymptotic variance $\sigma_\chi^2(Y_0^{n-1})(h)$ by similar forgetting-based arguments. Here a key ingredient is that the particles $(\xi_i^n)_{i=1}^N$ obtained at the different time steps become, asymptotically as $N$ tends to infinity, statistically independent. Consequently, the total asymptotic variance of $\sqrt{N}(\phi_\chi^N(Y_0^{n-1})h - \phi_\chi(Y_0^{n-1})h)$ is obtained by simply summing up the asymptotic variances of the error terms $\sqrt{N}(\Phi(Y_k^{n})\phi_\chi^N(Y_0^{k})h - \Phi(Y_k^{n})\phi_\chi(Y_0^{k-1})h)$ in the decomposition (4). Finally, applying again the contraction of the composed mapping $\Phi(Y_k^{n})$ yields a uniform bound on the total asymptotic variance in accordance with the calculation above. In [12], a similar stability result was obtained for a particle-based version of the forward-filtering backward-simulation algorithm (proposed in [18]); nevertheless, also the analysis of this work relies completely on the assumption of strong mixing of the latent Markov chain, which, as we already pointed out, does not hold for most models used in practice.

A first breakthrough toward stability results for noncompact state spaces was made in [34]. This work establishes, again for bootstrap-type particle filters, a uniform time average convergence result of form

$$\lim_{N \to \infty} \sup_{n \in \mathbb{N}} E\left(n^{-1} \sum_{k=1}^n \|\tilde{\phi}_\chi^N(Y_0^k) - \tilde{\phi}_\chi(Y_0^k)\|_{BL}\right) = 0,$$

where $\| \cdot \|_{BL}$ denotes the dual bounded-Lipschitz norm. This result, obtained as a special case of a general approximation theorem derived in the same paper, was established under very weak assumptions on the local likelihood (supposed to be bounded and continuous) and the Markov kernel (supposed to be Feller). These assumptions are, together with the basic assumption that the hidden Markov chain is positive Harris and aperiodic, satisfied for a large class of HMMs with possibly noncompact state spaces. Nevertheless, the proof is heavily based on the assumption that the particles evolve according to exactly the same model dynamics as the observations entered into the algorithm, in other words, that the model is perfectly specified. This is of course never true in practice. In addition, the convergence result (6) does not, in contrast to $L^p$ bounds and CLTs, provide a rate of convergence of the algorithm.

1.2. Approach of this paper. In this paper we return to more standard convergence modes and reconsider the asymptotic variance and $L^p$ error of bootstrap particle filters. As noticed by Johansen and Doucet [21], restricting the analysis of bootstrap-type particle filters does not imply a significant loss of generality, as the CLT for more general auxiliary particle filters [29]
can be straightforwardly obtained by applying the bootstrap filter CLT to a somewhat modified HMM incorporating the so-called adjustment multiplier weights of the auxiliary particle filter into the model dynamics. Our aim is to establish that the asymptotic variance and $L^p$ error are stochastically bounded in the noncompact case. Recall that a sequence $(\mu_n)_{n \in \mathbb{N}}$ of probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is tight if for all $\epsilon > 0$ there exists a compact interval $I = [-a, a] \subset \mathbb{R}$ such that $\mu_n(I) \leq \epsilon$ for all $n$. In addition, we call a sequence $(Z_n)_{n \in \mathbb{N}}$ of random variables, with $Z_n \sim \mu_n$, tight if the sequence $(\mu_n)_{n \in \mathbb{N}}$ of marginal distributions is tight. In this paper, we show that the sequence $(\sigma^2_n(Y_0^{n-1})(h))_{n \in \mathbb{N}^*}$ of asymptotic variances is tight for any stationary and ergodic sequence $(Y_n)_{n \in \mathbb{N}}$ of observations. In particular, we do not at all assume that the observations originate from the model governing the dynamics of the particle filter or not even from an HMM.

Our proofs are based on novel coupling techniques developed in [15] (and going back to [23] and [11]) with the purpose of establishing the convergence of the relative entropy for misspecified HMMs. In our analysis, the strong mixing assumption (5) is replaced by the considerably weaker $r$-local Doeblin condition (19). This assumption is, for instance, trivially satisfied (for $r = 1$) if there exist a measurable set $C \subseteq X$, a probability measure $\lambda_C$ on $(X, \mathcal{X})$ such that $\lambda_C(C) = 1$ and positive constants $0 < \epsilon^- < \epsilon^+_{c}$ such that for all $x \in X$ and all $A \in \mathcal{X}$,

\begin{equation}
\epsilon^-_{c} \lambda_C(A) \leq Q(x, A \cap C) \leq \epsilon^+_{c} \lambda_C(A),
\end{equation}

a condition that is easily verified for many HMMs with noncompact state space [we emphasize, however, that assumption (19) is even weaker than (7)].

Remark 1. Finally, we remark that Del Moral and Guionnet [7] studied the stability of SMC methods within the framework of a general normalized Feynman–Kac prediction model consisting of a sequence $(\mu_n)_{n \in \mathbb{N}}$ of measures-defined recursively on a sequence $(E_n, \mathcal{E}_n)_{n \in \mathbb{N}}$ of measurable spaces by

$$
\mu_{n+1}(A) = \frac{\int G_n(x)K_n(x, A)\mu_n(dx)}{\int G_n(x)\mu_n(dx)}, \quad A \in \mathcal{E}_{n+1},
$$

where $G_n$ is a positive potential function on $E_n$, and $K_n$ is a Markov transition kernel from $(E_n, \mathcal{E}_n)$ to $(E_{n+1}, \mathcal{E}_{n+1})$; see also [5], Section 2.3. Conditionally on the observations $(Y_n)_{n \in \mathbb{N}}$, the flow of predictor distributions can obviously, by (1)–(2), be formulated as a normalized Feynman–Kac prediction model by letting $(E_n, \mathcal{E}_n) \equiv (X, \mathcal{X})$, $\mu_n \equiv \phi(x)(Y_0^{n-1})$, $G_n \equiv g(Y_n)$, and $K_n \equiv Q$, $n \in \mathbb{N}$. Imposing, as in [7], the assumption that the transition kernels $(K_n)_{n \in \mathbb{N}}$ satisfy jointly the global Doeblin condition (5) provides a mixing rate $\rho$ that is uniform in the observations, and any stability result
obtained for fixed observations holds thus automatically true also when the observations are allowed to vary randomly.

Similarly, we could in the present paper have taken directly the recursion (1)–(2) as a starting point, by suppressing its connection with HMMs and describing the same as a normalized Feynman–Kac prediction model indexed by a stationary and ergodic sequence \((Y_n)_{n \in \mathbb{N}}\) of random parameters. However, as the results obtained in [15], which are fundamental in our analysis, describes the convergence of the normalized log-likelihood function for general HMMs, a quantity whose interpretation is not equally clear in the context of Feynman–Kac models, we have chosen to express our results in the language of HMMs as well.

To sum up, the contribution of the present paper is twofold, since:

- we present time uniform bounds providing also the rate of convergence in \(N\) of the particle filter for HMMs with possibly noncompact state space;
- we establish long-term stability of the particle filter also in the case of misspecification, that is, when the stationary law of the observations entering the particle filter differs from that of the HMM governing the dynamics of the particles \((\xi^i_n)_{i=1}^N\).

1.3. Outline of the paper. The paper is organized as follows. Section 2 provides the main notation and definitions. It also introduces the concepts of HMMs and bootstrap particle filters. In Section 3 our main results are stated together with the main layouts of the proofs. Section 4 treats some examples, and Section 5 and Section 5.2 provide the full details of our proofs.

2. Preliminaries.

2.1. Notation. We preface the introduction of HMMs with some notation. Let \((X, \mathcal{X})\) be a measurable space, where \(\mathcal{X}\) is a countably generated \(\sigma\)-field. Denote by \(\mathcal{F}(\mathcal{X})\) [resp., \(\mathcal{F}_+(\mathcal{X})\)] the space of bounded (resp., bounded and nonnegative) \(\mathcal{X}/\mathcal{B}(\mathbb{R})\)-measurable functions on \(X\) equipped with the supremum norm \(\|f\|_\infty \triangleq \sup_{x \in X}|f(x)|\). In addition, denote by \(\mathcal{P}(\mathcal{X})\) the set of probability measures on \((X, \mathcal{X})\). Let \(K: X \times \mathcal{X} \to \mathbb{R}_+\) be a finite kernel on \((X, \mathcal{X})\), that is, for each \(x \in X\), the mapping \(K(x, \cdot): \mathcal{X} \ni A \mapsto K(x, A)\) is a finite measure on \(\mathcal{X}\), and for each \(A \in \mathcal{X}\), the function \(K(\cdot, A): X \ni x \mapsto K(x, A)\) is \(\mathcal{X}/\mathcal{B}([0,1])\)-measurable. If \(K(x, \cdot)\) is a probability measure on \((X, \mathcal{X})\) for all \(x \in X\), then the kernel \(K\) is said to be Markov. A kernel induces two integral operators, the first acting on the space \(\mathcal{M}(\mathcal{X})\) of \(\sigma\)-finite measures on \((X, \mathcal{X})\) and the other on \(\mathcal{F}(\mathcal{X})\). More specifically, for \(\mu \in \mathcal{M}(\mathcal{X})\) and \(f \in \mathcal{F}(\mathcal{X})\), we define the measure

\[\mu K: \mathcal{X} \ni A \mapsto \int K(x, A)\mu(dx)\]
and the function

$$Kf: X \ni x \mapsto \int f(x')K(x, dx').$$

Moreover, the composition (or product) of two kernels $K$ and $M$ on $(X, \mathcal{X})$ is defined as

$$KM: X \times \mathcal{X} \ni (x, A) \mapsto \int M(x', A)K(x, dx').$$

### 2.2. Hidden Markov models.

Let $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ be two measurable spaces. We specify the HMM as follows. Let $Q: X \times \mathcal{X} \to [0, 1]$ and $G: X \times \mathcal{Y} \times [0, 1]$ be given Markov kernels, and let $\chi$ be a given initial distribution on $(X, \mathcal{X})$. In this setting, define the Markov kernel

$$T((x, y), A) \equiv \int \int \mathbb{1}_A(x', y')Q(x, dx')G(x', dy'),$$

on the product space $(X \times Y, \mathcal{X} \otimes \mathcal{Y})$. Let $(X_n, Y_n)_{n \in \mathbb{N}}$ be the canonical Markov chain induced by $T$ and the initial distribution $\mathcal{X} \otimes \mathcal{Y} \ni A \mapsto \int \mathbb{1}_A(x, y)\chi(dx)G(x, dy)$. The bivariate process $(X_n, Y_n)_{n \in \mathbb{N}}$ is what we refer to as the HMM. We shall denote by $P_\chi$ and $E_\chi$ the probability measure and corresponding expectation associated with the HMM on the canonical space $((X \otimes Y)^\mathbb{N}, (\mathcal{X} \otimes \mathcal{Y})^\otimes)$.

We assume that the observation kernel $G$ is nondegenerated in the sense that there exist a $\sigma$-finite measure $\nu$ on $(Y, \mathcal{Y})$ and a measurable function $g: X \times Y \to [0, \infty]$ such that

$$G(x, A) = \int \mathbb{1}_A(y)g(x, y)\nu(dy), \quad x \in X, A \in \mathcal{Y}.$$

For a given observation $y \in Y$, we let

$$g(y): X \ni x \mapsto g(x, y)$$

denote the local likelihood function of the state given the corresponding observation $y$.

When operating on HMMs we are in general interested in computing expectations of type $E_\chi(h(X_k^\ell)|Y_0^m)$ for integers $(k, \ell, m) \in \mathbb{N}^3$ with $k \leq \ell$ and functions $h \in \mathcal{F}(X^{\ell-k+1})$. Of particular interest are quantities of form $E_\chi(h(X_n)|Y_0^{n-1})$ or $E_\chi(h(X_n)|Y_0^n)$, and the term optimal filtering refers to problem of computing, recursively in $n$, such conditional distributions and expectations as new data becomes available. For any record $y_k^m \in Y^{m-k+1}$ of observations, let $L(y_k^m)$ be the unnormalized kernel on $(X, \mathcal{X})$ defined by

$$L(y_k^m)(x_k, A) \equiv \int \cdots \int \mathbb{1}_A(x_{m+1})\prod_{\ell=k}^m g(y_{\ell})(x_{\ell})Q(x_{\ell}, dx_{\ell+1}),$$

$$x_k \in X, A \in \mathcal{X}, \quad (8)$$
with the convention
\[ L(y_k^n)(x, A) \triangleq \delta_x(A) \quad \text{for } k > m \]
(9)
(where \( \delta_x \) denotes the Dirac mass at point \( x \)). Note that the function \( y_0^{n-1} \mapsto \chi L(y_0^{n-1}) \mathbb{1}_X \) is exactly the density of the observations \( Y_0^{n-1} \) (i.e., the likelihood function) with respect to \( \nu^{\otimes n} \). Also note that for any \( \ell \in \{k, \ldots, m-1\} \),
\[ L(y_k^m) = L(y_k^\ell)L(y_{\ell+1}^m). \]
(10)

Let \( \phi \langle y_k^m \rangle \) be the probability measure defined by
\[ \phi \langle y_k^m \rangle (A) = \frac{\chi L(y_k^m) \mathbb{1}_A}{\chi L(y_k^m) \mathbb{1}_X}, \quad A \in \mathcal{X}. \]
(11)

Note that this implies that \( \phi \langle y_k^m \rangle = \chi \) when \( k > m \). Using the notation, it can be shown (see, e.g., [2], Proposition 3.1.4) that for any \( h \in \mathcal{F}(\chi) \),
\[ \mathbb{E}_\chi(h(X_n)|Y_0^{n-1}) = \int h(x)\phi \langle Y_0^{n-1} \rangle(dx), \]
that is, \( \phi \langle Y_0^{n-1} \rangle \) is the predictor of \( X_n \) given the observations \( Y_0^{n-1} \). From definition (11) one immediately obtains the recursion
\[ \phi \langle y_0^n \rangle (A) = \frac{\phi \langle y_0^{n-1} \rangle L(y_0^n) \mathbb{1}_A}{\phi \langle y_0^{n-1} \rangle L(y_0^n) \mathbb{1}_X} = \frac{\int g(y_n)(x)Q(x, A)\phi \langle y_0^{n-1} \rangle(dx)}{\int g(y_n)(x)\phi \langle y_0^{n-1} \rangle(dx)}, \quad A \in \mathcal{X}, \]
which can be expressed in condensed form as
\[ \phi \langle y_0^n \rangle = \Phi(y_n)(\phi \langle y_0^{n-1} \rangle) \]
(12)
with \( \Phi(y_n) \) being the measure-valued transformation
\[ \Phi(y_n):\mathcal{P}(\chi) \ni \mu \mapsto \Psi(y_n)(\mu)Q \]
and \( \Psi(y_n) \) transforms a measure \( \mu \in \mathcal{P}(\chi) \) into the measure
\[ \Psi(y_n)(\mu):\mathcal{X} \ni A \mapsto \frac{\mathbb{1}_A(x)g(y_n)(x)\mu(dx)}{\int g(y_n)(x)\mu(dx)} \]
in \( \mathcal{P}(\chi) \). By introducing the filter distributions
\[ \bar{\phi} \langle y_0^n \rangle \triangleq \Psi(y_n)(\phi \langle y_0^{n-1} \rangle), \]
satisfying, for all \( h \in \mathcal{F}(\chi) \),
\[ \mathbb{E}_\chi(h(X_n)|Y_0^n) = \int h(x)\bar{\phi} \langle Y_0^n \rangle(dx) \]
(see again [2], Proposition 3.1.4), we may express one iteration of the filter recursion in terms of the two operations
\[ \phi \langle y_0^{n-1} \rangle \xrightarrow{\text{Updating}} \bar{\phi} \langle y_0^n \rangle \xrightarrow{\text{Prediction}} \phi \langle y_0^n \rangle. \]
As mentioned in the Introduction, it is in general infeasible to obtain closed-form solutions to the recursion (12). In the following section we discuss how approximate solutions to (12) can be obtained using particle filters, with focus set on the bootstrap particle filter proposed in [19].

2.3. The bootstrap particle filter. In the following we assume that all random variables are defined on a common probability space \((Ω, A, P)\). The bootstrap particle filter updates sequentially a set of weighted simulations in order to approximate online the flow of predictor and filter distributions. In order to describe precisely how this is done for a given sequence \((y_n)_{n∈N}\) of observations, we proceed inductively and assume that we are given a sample of \(X\)-valued random draws \((ξ_i^n)_{N_i=1}^N\) (the particles) such that the empirical measure

\[
φ^N_X(y_0^{n-1}) ≜ \frac{1}{N} \sum_{i=1}^N δ_{ξ_i^n}
\]

associated with these draws targets the predictor \(φ_X(y_0^{n-1})\) in the sense that \(φ^N_X(y_0^{n-1})h = \sum_{i=1}^N h(ξ_i^n)/N\) estimates \(φ_X(y_0^{n-1})h\) for any \(h ∈ F(Χ)\). In order to form a new particle sample \((ξ_i^{n+1})_{N_i=1}^N\) approximating the predictor \(φ_X(y_0^{n})\) at the subsequent time step, we replace, in (12), the true predictor \(φ_X(y_0^{n-1})\) by the particle estimate \(φ^N_X(y_0^{n-1})\), and pass the latter through the updating and prediction steps. This yields, after updating, the self-normalized approximation

\[
φ^N_X(y_0^{n}) ≜ Ψ(y_n)(φ^N_X(y_0^{n-1})) = \sum_{i=1}^N \frac{ω_i^n}{Ω_n^N} δ_{ξ_i^n}
\]

of the filter \(φ_X(y_0^{n})\), where we have introduced the importance weights \(ω_i^n ≜ g(y_n)(ξ_i^n), i ∈ \{1, 2, \ldots, N\}\) and \(Ω_n^N ≜ \sum_{i=1}^N ω_i^n\). Moreover, by propagating filter approximation (13) through prediction step one obtains the approximation

\[
\sum_{i=1}^N \frac{ω_i^n}{Ω_n^N} Q(ξ_i^n, ·)
\]

of the predictor \(φ_X(y_0^{n})\). Finally, the sample \((ξ_i^{n+1})_{N_i=1}^N\) is generated by simulating \(N\) conditionally independent draws from the mixture in (14) using the following algorithm:

\[
\text{set } Ω_n^N ← 0
\]

\[\text{for } i = 1 \rightarrow N \text{ do}
\]

\[\text{set } ω_i^n ← g(y_n)(ξ_i^n)\]
set $\Omega_n^N \leftarrow \Omega_n^N + \omega_n^i$
end for
for $i = 1 \rightarrow N$ do
\begin{align*}
&\text{draw } I_n^i \sim (\omega_n^i/\Omega_n^N)^N_{\ell=1} \\
&\text{draw } \xi_{n+1}^i \sim Q(\xi_{n+1}^i, \cdot)
\end{align*}
end for

After this, the empirical measure $\phi^N_\chi(\langle y_0^n \rangle) = \sum_{i=1}^N \delta_{\xi_{n+1}^i}/N$ is returned as an approximation of $\phi_\chi(\langle y_0^n \rangle)$. In the scheme above, the operation $\sim$ means implicitly that all draws (for different $i$'s) are conditionally independent. Moreover, the operation $I_n^i \sim (\omega_n^i/\Omega_n^N)^N_{\ell=1}$ means that each index $I_n^i$ is simulated according to the discrete probability distribution generated by the normalized importance weights $(\omega_n^i/\Omega_n^N)^N_{\ell=1}$. The procedure described above is repeated recursively in order to produce particle approximations of the predictor and filter distributions at all time steps. The algorithm is typically initialized by drawing $N$ i.i.d. particles $(\xi_{0,1}^N)_{i=1}^N$ from the initial distribution $\chi$ and letting $\sum_{i=1}^N \delta_{\xi_{0,1}^N}/N$ be an estimate of $\chi$.

As mentioned in the Introduction, the asymptotic properties, as the number $N$ of particles tends to infinity, of the bootstrap particle filter output are well investigated. When it concerns weak convergence, Del Moral and Guionnet [6] established the following CLT. Define for $h \in \mathcal{F}(X)$,

\begin{equation}
\sigma^2_\chi(\langle y_0^{n-1} \rangle(h) \triangleq \sum_{k=0}^n \phi_\chi(\langle y_0^{k-1} \rangle(h - \phi_\chi(\langle y_0^{n-1} \rangle(h \times L(\langle y_0^{n-1} \rangle)\mathbb{1}_X \right)^2.
\end{equation}

THEOREM 2 [6]. For all $h \in \mathcal{F}(X)$ and all $y_0^{n-1} \in Y^n$ such that $\|g(y_t)\|_\infty < \infty$ for all $y_t$, it holds, as $N \rightarrow \infty$,

\begin{equation}
\sqrt{N}(\phi^N_\chi(\langle y_0^{n-1} \rangle(h - \phi_\chi(\langle y_0^{n-1} \rangle)h) \xrightarrow{D} \sigma_\chi(\langle y_0^{n-1} \rangle(h)Z,
\end{equation}

where $\sigma_\chi(\langle y_0^{n-1} \rangle(h)$ is defined in (15), and $Z$ is a standard normal-distributed random variable.

The next corollary states the corresponding CLT for the particle filter. Also this result is standard and is an immediate consequence of Theorem 2, the fact that $\Omega_n^N/N$ converges, for all $y_0^n \in Y^{n+1}$ and as $N$ tends to infinity, in probability to $\phi_\chi(\langle y_0^{n-1} \rangle g(y_n)$ (see, e.g., [14], Theorem 10), and Slutsky’s lemma. Let again $\sigma^2_\chi$ be given by (15), and define for $y_0^n \in Y^{n+1}$ and $h \in \mathcal{F}(X)$,

\begin{equation}
\tilde{\sigma}^2_\chi(\langle y_0^n \rangle(h) \triangleq \frac{\sigma^2_\chi(\langle y_0^{n-1} \rangle(g(y_n)\{h - \tilde{\phi}_\chi(\langle y_0^n \rangle)h\})}{(\phi_\chi(\langle y_0^{n-1} \rangle g(y_n))^2}.
\end{equation}
Corollary 3. For all \( h \in \mathcal{F}(X) \) and \( \eta_0^n \in Y^{n+1} \) such that \( \|g(y_0^0)\|_\infty < \infty \) for all \( y_0^0 \) it holds, as \( N \to \infty \),
\[
\sqrt{N}(\hat{\sigma}^N_\chi(\eta_0^n)h - \hat{\sigma}_\chi(\eta_0^n)h) \overset{D}{\to} \tilde{\sigma}_\chi(\eta_0^n)(h)Z,
\]
where \( \tilde{\sigma}_\chi(\eta_0^n)(h) \) is defined in (17) and \( Z \) is a standard normal-distributed random variable.

When the observations \( (Y_n)_{n \in \mathbb{N}} \) entering the particle filter are random, the sequences \( (\sigma^2_\chi(Y_0^n)h)_{n \in \mathbb{N}} \) and \( (\tilde{\sigma}^2_\chi(Y_0^n)h)_{n \in \mathbb{N}} \) of asymptotic variances are \( (\mathcal{F}^Y_n)_{n \in \mathbb{N}} \)-adapted stochastic processes, where \( (\mathcal{F}^Y_n)_{n \in \mathbb{N}} \) is the natural filtration of the observation process. The aim of the next section is to establish that these sequences are tight. \textit{Importantly, we assume in the following that the observations \( (Y_n)_{n \in \mathbb{N}} \) entering the particle filter algorithm is an arbitrary \( \mathbb{P} \)-stationary sequence taking values in \( Y \).} The stationary process \( (Y_n)_{n \in \mathbb{N}} \) can be embedded into a stationary process \( (Y_n)_{n \in \mathbb{Z}} \) with doubly infinite time. In particular, we do not at all assume that the observations originate from the model governing the dynamics of the particles; indeed, in the framework we consider, we do not even assume that the observations originate from an HMM.

3. Main results and assumptions. Before listing our main assumptions, we recall the definition of a \( r \)-local Doeblin set.

Definition 4. Let \( r \in \mathbb{N}^* \). A set \( C \in \mathcal{X} \) is \( r \)-\textit{local Doeblin} with respect to \( \{Q,g\} \) if there exist positive functions \( \epsilon^-_C : Y^r \to \mathbb{R}^+ \) and \( \epsilon^+_C : Y^r \to \mathbb{R}^+ \), a family \( \{\mu_C(z) ; z \in Y^r\} \) of probability measures, and a family \( \{\varphi_C(z) ; z \in Y^r\} \) of positive functions such that for all \( z \in Y^r \), \( \mu_C(z)(C) = 1 \) and for all \( A \in \mathcal{X} \) and \( x \in C \),
\[
(19) \quad \epsilon^-_C(z)\varphi_C(z)(x)\mu_C(z)(A) \leq L(z)(x,A \cap C) \\
\leq \epsilon^+_C(z)\varphi_C(z)(x)\mu_C(z)(A).
\]

(A1) The process \( (Y_n)_{n \in \mathbb{Z}} \) is strictly stationary and ergodic. Moreover, there exist an integer \( r \in \mathbb{N}^* \) and a set \( K \in Y^{\otimes r} \) such that the following holds:

(i) The process \( (Z_n)_{n \in \mathbb{Z}} \), where \( Z_n \triangleq Y_{nr}^{(n+1)r-1} \), is ergodic and such that \( \mathbb{P}(Z_0 \in K) > 2/3 \).

(ii) For all \( \eta > 0 \) there exists an \( r \)-local Doeblin set \( C \in \mathcal{X} \) such that for all \( y_0^{r-1} \in K \),
\[
(20) \quad \sup_{x \in C^c} L(y_0^{r-1}(x,X)) \leq \eta \|L(y_0^{r-1}(\cdot,X))\|_\infty < \infty
\]
and

\[ \inf_{y_0^{-1} \in K} \frac{\epsilon^+_{\mathcal{C}}(y_0^{r-1})}{\epsilon^+_{\mathcal{C}}(y_0^{r-1})} > 0, \tag{21} \]

where the functions \( \epsilon^+_{\mathcal{C}} \) and \( \epsilon^-_{\mathcal{C}} \) are given in Definition 4.

(iii) There exists a set \( D \in \mathcal{X} \) such that

\[ \mathbb{E}\left( \ln^- \inf_{x \in D} \delta_x \mathbf{L}(Y_0^{r-1}) \mathbb{1}_D \right) < \infty. \tag{22} \]

(A2) (i) \( g(x, y) > 0 \) for all \((x, y) \in X \times Y\).
(ii) \( \mathbb{E}(\ln^+ \|g(Y_0)\|_\infty) < \infty \).

Remark 5. Assumption (A1)(i) is inherited from [15]. To get some rationale behind the constant \( 2/3 \) appearing in the assumption, note that the same is in fact equivalent to \( 1 - \mathbb{P}(Z_0 \in K) < 2\mathbb{P}(Z_0 \in K) - 1 \). In that case there exist \( 0 < \gamma^- < \gamma^+ \) such that

\[ 1 - \mathbb{P}(Z_0 \in K) < \gamma^- < \gamma^+ < 2\mathbb{P}(Z_0 \in K) - 1, \]

which is equivalent to \( \mathbb{P}(Z_0 \in K) > \max\{1 - \gamma^-, (1 + \gamma^+)/2\} \). The latter inequality is essential when applying [15], Proposition 5; see also the proof of [15], Proposition 8.

Remark 6. In the case \( r = 1 \) we may replace (A1) by the simpler assumption that there exists a set \( K \in \mathcal{Y} \) such that the following holds:

(i) \( \mathbb{P}(Y_0 \in K) > 2/3 \).
(ii) For all \( \eta > 0 \) there exists a local Doeblin set \( C \in \mathcal{X} \) such that for all \( y \in K \),

\[ \sup_{x \in C^c} g(x, y) \leq \eta \|g(y)\|_\infty < \infty. \tag{23} \]

(iii) There exists a set \( D \in \mathcal{X} \) satisfying

\[ \inf_{x \in D} Q(x, D) > 0 \quad \text{and} \quad \mathbb{E}\left( \ln^- \inf_{x \in D} g(x, Y_0) \right) < \infty. \]

For the integer \( r \in \mathbb{N}^* \) and the set \( D \in \mathcal{X} \) given in (A1), define

\[ \mathcal{M}(D, r) \triangleq \{ \chi \in \mathcal{P}(\mathcal{X}) : \mathbb{E}(\ln^- \chi \mathbf{L}(Y_0^{r-1}) \mathbb{1}_D) < \infty \text{ for all } \ell \in \{0, \ldots, r\} \}. \tag{24} \]

A simple sufficient condition can be proposed to ensure that \( \chi \in \mathcal{M}(D, r) \).
Proposition 7. Assume that there exist sets $D_u \in \mathcal{X}$, $u \in \{0, \ldots, r-1\}$, such that (setting $D_r = D$ for notational convenience) for some $\delta > 0$,

$$\inf_{x \in D_{u-1}} Q(x, D_u) \geq \delta, \quad u \in \{1, \ldots, r\}$$

and

$$\mathbb{E}\left(\ln^{-1} \inf_{x \in D_u} g(x, Y_0)\right) < \infty, \quad u \in \{0, \ldots, r\}.$$  \hspace{1cm} (26)

Then any initial distribution $\chi \in \mathcal{P}(\mathcal{X})$ satisfying $\chi(D_0) > 0$ belongs to $\mathcal{M}(D, r)$.

Remark 8. To check (26) we typically assume that for any given $y \in \mathcal{Y}$, $g(y)$ is continuous and that the sets $D_i$, $i \in \{0, \ldots, r-1\}$, are compact. This condition then translates into an assumption on some generalized moments of the process $(Y_n)_{n \in \mathbb{Z}}$.

Remark 9. Assume that $\mathcal{X} = \mathbb{R}^d$ for some $d \in \mathbb{N}^*$ (or more generally, $\mathcal{X}$ is a locally compact separable metric space) and that $\mathcal{X}$ is the associated Borel $\sigma$-field. Assume in addition that for any open set $O \in \mathcal{X}$, the function $x \rightarrow Q(x, O)$ is lower semi-continuous on the space $\mathcal{X}$. Then for any $\delta > 0$ and any compact set $D_0 \in \mathcal{X}$, there exist compact sets $D_u$, $u \in \{0, \ldots, r-1\}$, satisfying (25).

We are now ready to state our main results.

3.1. Tightness of asymptotic variance.

Theorem 10. Assume (A1)–(A2). Then for all $\chi \in \mathcal{M}(D, r)$ and all $h \in \mathcal{F}(\mathcal{X})$, the sequence $(\sigma^2_{\chi}(Y_n^{n-1})(h))_{n \in \mathbb{N}^*}$ [defined in (15)] is tight.

Proof. Using definition (11) of the predictive distribution and the decomposition (10) of the likelihood, we get for all $k \in \{0, \ldots, n-1\}$,

$$\phi_\chi(Y_0^{n-1})h = \frac{\chi L(Y_0^{n-1})h}{\chi L(Y_0^{n-1})1_X} = \frac{\chi L(Y_0^{k-1})L(Y_k^{n-1})h}{\chi L(Y_0^{k-1})L(Y_k^{n-1})1_X}. $$

Plugging this identity into the expression (15) of the asymptotic variance yields

$$\sigma^2_{\chi}(Y_0^{n-1})(h) = \sum_{k=0}^{n} \int \phi_\chi(Y_0^{k-1})(dx) \left[ \frac{\Delta_{\phi, \phi_\chi}(Y_0^{k-1})L(Y_k^{n-1})1_X}{(\phi_\chi(Y_0^{k-1})L(Y_k^{n-1})1_X)^2} \right]^2,$$

where for all sequences $y_k^{n-1} \in \mathcal{Y}^{n-k}$, functions $f$ and $h$ in $\mathcal{F}(\mathcal{X})$ and probability measures $\chi$ and $\chi'$ in $\mathcal{P}(\mathcal{X})$,

$$\Delta_{\chi, \chi'}(y_k^{n-1})(f, h) \triangleq \chi L(y_k^{n-1})f \times \chi' L(y_k^{n-1})h - \chi L(y_k^{n-1})h \times \chi' L(y_k^{n-1})f.$$  \hspace{1cm} (28)
Using (11), we obtain for all sequences $y_0^{n-1} \in \mathcal{Y}$,
\[
\phi_{\chi}(y_0^{k-1}) \mathcal{L}(y_k^{n-1}) \mathbf{1}_X = \frac{\chi \mathcal{L}(y_0^{n-1}) \mathbf{1}_X}{\chi \mathcal{L}(y_0^{k-1}) \mathbf{1}_X} = \prod_{\ell=k}^{n-1} \frac{\chi \mathcal{L}(y_0^{\ell-1}) \mathbf{1}_X}{\chi \mathcal{L}(y_0^{\ell-1}) \mathbf{1}_X} = \prod_{\ell=k}^{n-1} \pi_{\chi}(y_0^{\ell-1})(y_{\ell}),
\]
where $\pi_{\chi}(y_0^{\ell-1})(y_{\ell})$ is the density of the conditional distribution of $Y_\ell$ given $Y_0^{\ell-1}$ (i.e., the one-step observation predictor at time $\ell$) defined by
\[
(29) \quad \pi_{\chi}(y_0^{\ell-1})(y_{\ell}) \triangleq \int \phi_{\chi}(y_0^{\ell-1})(dx)g(x, y_{\ell}).
\]
With this notation, the likelihood function \(\chi \mathcal{L}(y_0^{n-1}) \mathbf{1}_X\) equals the product $\prod_{k=0}^{n-1} \pi_{\chi}(y_0^{k-1})(y_k)$ [where we let $\pi_{\chi}(y_0^{-1})(y_0)$ denote the marginal density of $Y_0$].

Now, using coupling results obtained in [15] one may prove that the predictor distribution forgets its initial distribution exponentially fast under the $r$-local Doeblin assumption (19). Moreover, this implies that also the log-density of the one-step observation predictor forgets its initial distribution exponentially fast; that is, for all initial distributions $\chi$ and $\chi'$ there is a deterministic constant $\beta \in [0, 1]$ and an almost surely bounded random variable $C_{\chi, \chi'}$ such that for all $(k, m) \in \mathbb{N}^* \times \mathbb{N}$ and almost all observation sequences,
\[
|\ln \pi_{\chi}(Y_{-m}^{k-1})(Y_k) - \ln \pi_{\chi'}(Y_{-m}^{k-1})(Y_k)| \leq C_{\chi, \chi'} \beta^{k+m}.
\]
Using this, it is shown in [15], Proposition 1, that:

(i) There exists a function $\pi: \mathcal{Y}^\mathbb{Z} \times \mathcal{Y} \to \mathbb{R}$ such that for all probability measures $\chi \in \mathcal{M}(\mathcal{D}, r)$,
\[
\lim_{m \to \infty} \pi_{\chi}(Y_{-m})(Y_0) = \pi(Y_{-\infty})(Y_0), \quad \mathbb{P}\text{-a.s.}
\]
Moreover,
\[
(31) \quad \mathbb{E}(|\ln \pi(Y_{-\infty})(Y_0)|) < \infty.
\]
(ii) For all probability measures $\chi \in \mathcal{M}(\mathcal{D}, r)$, the normalized log-likelihood function converges according to
\[
(32) \quad \lim_{n \to \infty} n^{-1} \ln \chi \mathcal{L}(Y_0^{n-1}) \mathbf{1}_X = \ell_\infty, \quad \mathbb{P}\text{-a.s.},
\]
where $\ell_\infty$ is the negated relative entropy, that is, the expectation of $\ln \pi(Y_{-\infty})(Y_0)$ under the stationary distribution, that is,
\[
(33) \quad \ell_\infty \triangleq \mathbb{E}(\ln \pi(Y_{-\infty})(Y_0)).
\]
As a first step, we bound the asymptotic variance $\sigma^2_X(h|Y^{n-1}_0)$ [defined in (15)] by the product of two quantities, namely $\sigma^2_X(Y^{n-1}_0|Y_0)$, where

$$A \triangleq \left( \sup_{(k,m) \in \mathbb{N}^2 : k \leq m} \prod_{\ell=k}^{m-1} \frac{\pi(Y^{\ell-1}_\infty|X)|}{\pi_X(Y^{\ell-1}_0|Y_0)} \right)^4,$$

$$B_n \triangleq \sum_{m=0}^{n} \left( \sup_{x} \left| \frac{\Delta_{\delta_x,\phi},(Y^{m-1}_0|Y^{m-1}_0)(h,1x)}{\prod_{\ell=m}^{n-1} \pi(Y^{\ell-1}_\infty|Y_\ell)} \right| \right)^2.$$

Quantity (34) can be bounded using the exponential forgetting (30) of the one-step predictor log-density. More precisely, note that

$$\pi_X(Y^{\ell-1}_m|Y_\ell) = \frac{\chi(L(Y^{\ell-1}_m)1x)}{\chi(L(Y^{\ell-1}_m)1x)};$$

thus, by applying Proposition 16(ii), we conclude that there exist $\beta \in ]0,1[$ and a $\mathbb{P}$-a.s. finite random variable $C_X$ such that for all $n \in \mathbb{N}$,

$$\prod_{\ell=k}^{m-1} \frac{\pi(Y^{\ell-1}_\infty|X)|}{\pi_X(Y^{\ell-1}_0|Y_0)} \leq \prod_{\ell=k}^{m} \prod_{m=0}^{\infty} \exp(C_X \beta^{\ell+m})$$

$$\leq \exp(C_X/(1-\beta)^2) < \infty, \quad \mathbb{P} \text{-a.s.},$$

implying that $A$ is indeed $\mathbb{P}$-a.s. finite.

Consider now the second quantity (35). Since the process $(Y_n)_{n \in \mathbb{Z}}$ is strictly stationary, $Y^{n-1}_0$ has the same distribution as $Y^{-1}_n$ for all $n \in \mathbb{N}^\ast$. Therefore, for all $n \in \mathbb{N}^\ast$, the random variable $B_n$ has the same distribution as

$$B_n \triangleq \sum_{m=0}^{n} \left( \sup_{x} \left| \frac{\Delta_{\delta_x,\phi},(Y^{m-1}_0|Y^{m-1}_0)(h,1x)}{\prod_{\ell=m}^{n-1} \pi(Y^{\ell-1}_\infty|Y_\ell)} \right| \right)^2.$$

We will show that $\sup_{n \in \mathbb{N}^\ast} B_n$ is $\mathbb{P}$-a.s. finite, which implies that the sequence $(B_n)_{n \in \mathbb{N}^\ast}$ is tight. We split each term of $B_n$ into two factors according to

$$\sup_{x} \left| \frac{\Delta_{\delta_x,\phi},(Y^{m-1}_0|Y^{m-1}_0)(h,1x)}{\prod_{\ell=m}^{n-1} \pi(Y^{\ell-1}_\infty|Y_\ell)} \right|^2 \leq \left( \frac{\|L(Y^{m-1}_m)1x\|_\infty}{\prod_{\ell=m}^{n-1} \pi(Y^{\ell-1}_\infty|Y_\ell)} \right)^2 \sup_{x} \left| \frac{\Delta_{\delta_x,\phi},(Y^{m-1}_0|Y^{m-1}_0)(h,1x)}{\|L(Y^{m-1}_m)1x\|_\infty} \right|^2.$$
and consider each factor separately.

We will show that the first factor in (38) grows at most subgeometrically fast. Indeed, note that

$$\left( \frac{\|L(Y^{-1})_{\mathbb{1}}\|_{\infty}}{\prod_{\ell=1}^{m} \pi(Y_{\infty}^{-\ell})} \right)^{2} = \exp(m \varepsilon_{m}),$$

where

$$\varepsilon_{m} \triangleq \frac{2}{m} \left( \ln \|L(Y^{-1})_{\mathbb{1}}\|_{\infty} - \sum_{\ell=1}^{m} \ln \pi(Y_{\infty}^{-\ell}) \right).$$

According to Lemma 17, \( \varepsilon_{m} \to 2(\ell_{\infty} - \ell_{\infty}) = 0 \), \( \mathbb{P} \)-a.s., as \( m \to \infty \).

The second factor in (38) is handled using Proposition 16(iii), which guarantees the existence of a constant \( \beta \in [0,1] \) and a \( \mathbb{P} \)-a.s. random variable \( C \) such that for all \((m,n) \in (\mathbb{N}^{*})^{2}\),

$$\sup_{x \in X} \frac{|\Delta_{\delta_{x},\phi_{Y}^{-m}}(Y_{Y_{\infty}^{-m}})(h,1_{x})|}{\|L(Y_{Y_{\infty}^{-m}})_{\mathbb{1}}\|_{\infty}^{2}} \leq C \beta^{m} \|h\|_{\infty}.$$

This concludes the proof. \( \square \)

**Theorem 11.** Assume (A1)-(A2). Then for all \( \chi \in \mathcal{M}(\mathbb{D},r) \) and all \( h \in \mathcal{F}(\chi) \), the sequence \( (\hat{\sigma}_{\chi}^{2}(Y_{0}^{n})(h))_{n \in \mathbb{N}^{*}} \) [defined in (17)] is tight.

**Proof.** Using the expression (27) of the asymptotic variance of the predictor approximation yields for all \( y_{0} \in Y^{n+1} \), as

$$\phi_{\chi}(y_{0}^{n-1})g(y_{n}) = \pi_{\chi}(y_{0}^{n-1})(y_{n})$$

and

$$\pi_{\chi}(y_{0}^{n-1})(y_{n}) \times \phi_{\chi}(y_{0}^{k-1})L(y_{k}^{n-1})1_{X} = \phi_{\chi}(y_{0}^{k-1})L(y_{k}^{n})1_{X},$$

the identity

$$\hat{\sigma}_{\chi}^{2}(y_{0}^{n})(h)$$

$$= \sum_{k=0}^{n} \int \phi_{\chi}(y_{0}^{k-1})(dx) \left[ \frac{\Delta_{\delta_{x},\phi_{Y}^{k-1}}(y_{0}^{k-1})(g(y_{n})\{h - \phi_{\chi}(y_{0}^{n})h\},1_{X})}{\phi_{\chi}(y_{0}^{k-1})L(y_{k}^{n-1})1_{X} \times \phi_{\chi}(y_{0}^{k-1})L(y_{k}^{n})1_{X}} \right]^{2},$$

where

$$\Delta_{\delta_{x},\phi_{Y}^{k-1}}(y_{0}^{k-1})(g(y_{n})\{h - \phi_{\chi}(y_{0}^{n})h\},1_{X})$$

$$= \delta_{x}L(y_{k}^{n-1})(g(y_{n})\{h - \phi_{\chi}(y_{0}^{n})h\}) \times \phi_{\chi}(y_{0}^{k-1})L(y_{k}^{n-1})1_{X}$$

$$- \delta_{x}L(y_{k}^{n-1})1_{X} \times \phi_{\chi}(y_{0}^{k-1})L(y_{k}^{n-1})(g(y_{n})\{h - \phi_{\chi}(y_{0}^{n})h\}),$$

(40)
Here the equation
\[(41)\]
\[
\bar{\phi}_\chi(y_0^n) h = \frac{\phi_\chi(y_0^{k-1})L(y_k^{n-1})(g(y_n)h)}{\phi_\chi(y_0^{k-1})L(y_k^{n-1})g(y_n)}
\]
implies that
\[
\phi_\chi(y_0^{k-1})L(y_k^{n-1})(g(y_n)\{h - \bar{\phi}_\chi(y_0^n)h\})
\]
\[= \phi_\chi(y_0^{k-1})L(y_k^{n-1})(g(y_n)h) - \bar{\phi}_\chi(y_0^n)h \times \phi_\chi(y_0^{k-1})L(y_k^{n-1})g(y_n)
\]
\[= 0,
\]
which implies in turn that the second term on the right-hand side of (40) vanishes. Thus, developing also the first term and reusing the identity (41) yields
\[
\frac{\Delta_{\delta_x,\phi_\chi(y_0^{k-1})}(y_k^{n-1})(g(y_n)\{h - \bar{\phi}_\chi(y_0^n)h\}, 1_\mathbb{X})}{\phi_\chi(y_0^{k-1})L(y_k^{n-1})1_\mathbb{X} \times \phi_\chi(y_0^{k-1})L(y_k^{n-1})1_\mathbb{X}}
\]
\[= \frac{\delta_x L(y_k^{n-1})(g(y_n)h) - \bar{\phi}_\chi(y_0^n)h \times \delta_x L(y_k^{n-1})g(y_n)}{\phi_\chi(y_0^{k-1})L(y_k^{n-1})1_\mathbb{X}}
\]
\[= \frac{\Delta_{\delta_x,\phi_\chi(y_0^{k-1})}(y_k^{n-1})(g(y_n)h, g(y_n))}{\phi_\chi(y_0^{k-1})L(y_k^{n-1})1_\mathbb{X} \times \phi_\chi(y_0^{k-1})L(y_k^{n-1})1_\mathbb{X}}
\]
\[= \frac{\Delta_{\delta_x,\phi_\chi(y_0^{k-1})}(y_k^{n-1})(g(y_n)h, g(y_n))}{(\phi_\chi(y_0^{k-1})L(y_k^{n-1})1_\mathbb{X})^2}.
\]
Thus, to sum up,
\[
\bar{\sigma}_\chi^2(y_0^n)(h) = \sum_{k=0}^n \phi_\chi(y_0^{k-1})(\text{d}x) \left[ \frac{\Delta_{\delta_x,\phi_\chi(y_0^{k-1})}(y_k^{n-1})(g(y_n)h, g(y_n))}{(\phi_\chi(y_0^{k-1})L(y_k^{n-1})1_\mathbb{X})^2} \right]^{2}.
\]
providing an expression of the asymptotic variance of the particle filter that resembles closely the corresponding expression (27) for the particle predictor. Using this, the tightness can be established along the very same lines as Theorem 10, and we leave the details to the interested reader.

3.2. Tightness of the asymptotic L^p error. In the following we show that tightness of the asymptotic variance implies tightness of the asymptotic L^p error (when scaled with \(\sqrt{N}\)). The asymptotic L^p error given in Theorem 12 below is obtained by establishing, for fixed time indices n, using a standard exponential deviation inequality, uniform integrability (with respect to the particle sample size N) of the sequence of scaled L^p errors. After this, weak convergence implies convergence of moments, implying in turn convergence of the L^p error.
Theorem 12. Assume (A2). Then, for all functions $h \in \mathcal{F}(X)$, constants $p \in \mathbb{R}^*_+$ and all initial distributions $\chi \in \mathcal{P}(X)$, $\mathbb{P}$-a.s.,
\[
\lim_{N \to \infty} \sqrt{N} \mathbb{E}^{1/p}(|\phi^N_X(Y_{0}^{n-1})h - \phi_X(Y_{0}^{n-1})h|^p|Y_0^{n-1})
= \sqrt{2\sigma_X(Y_0^{n-1})}(\frac{\Gamma((p+1)/2)}{\sqrt{2\pi}2})^{1/p},
\]
where $\Gamma$ is the gamma function.

Proof. Recall that if $(A_N)_{N \in \mathbb{N}^*}$ is a sequence of random variables such that $A_N \rightarrow A$ as $N \rightarrow \infty$ and $(A_N^p)_{N \in \mathbb{N}^*}$ is uniformly integrable for some $p > 0$, then $\mathbb{E}(|A|^p) < \infty$, $\lim_{N \to \infty} \mathbb{E}(A_N^p) = \mathbb{E}(A^p)$ and $\lim_{N \to \infty} \mathbb{E}(|A_N|^p) = \mathbb{E}(|A|^p)$; see, for example, [31], Theorem A, page 14. Now set, for $n \in \mathbb{N}^*$,

$A_{N,\chi}(Y_0^{n-1})(h) \triangleq \sqrt{N}(\phi^N_X(Y_0^{n-1})h - \phi_X(Y_0^{n-1})h)$.

Let $q > p$ and write

$\sup_{N \in \mathbb{N}^*} \mathbb{E}(|A_{N,\chi}(Y_0^{n-1})(h)|^q|Y_0^{n-1})$

$= \sup_{N \in \mathbb{N}^*} \int_0^\infty \mathbb{P}(|A_{N,\chi}(Y_0^{n-1})(h)| \geq \epsilon|Y_0^{n-1}) \epsilon d\epsilon$

$= \sup_{N \in \mathbb{N}^*} \left( e^{q-1} \mathbb{P}(|A_{N,\chi}(Y_0^{n-1})(h)| \geq \epsilon|Y_0^{n-1}) \epsilon d\epsilon \right)$.

Now, note that (A2)(ii) implies that $\|g(Y_0)\|_\infty$ is $\mathbb{P}$-a.s. finite for all $n \in \mathbb{N}$. Thus, Assumptions 1 and 2 of (12) are fulfilled, which implies, via Remark 1 in the same work (see also [13], Lemma 2.1, [8], Theorem 3.1, [10], Theorem 3.39, and [9], Lemma 4, for similar results), that there exist, for all $n \in \mathbb{N}$ and $h \in \mathcal{F}(X)$, positive constants $B_n$ and $C_n$ (where only the latter depends on $h$) such that for all $N \in \mathbb{N}^*$ and all $\epsilon > 0$,

$\mathbb{P}(|A_{N,\chi}(Y_0^{n-1})(h)| \geq \epsilon|Y_0^{n-1}) \leq B_n \epsilon \exp(-C_n \epsilon^2)$.

This implies that for all $n \in \mathbb{N}$, $\mathbb{P}$-a.s.,

$\sup_{N \in \mathbb{N}^*} \mathbb{E}(|A_{N,\chi}(Y_0^{n-1})(h)|^q|Y_0^{n-1}) \leq qB_n \int_0^\infty \epsilon^{q-1} \exp(-C_n \epsilon^2) \epsilon d\epsilon < \infty$,

which establishes, via [32], Lemma II.6.3, page 190, as $q > p$, that the sequence $(|A_{N,\chi}(Y_0^{n-1})(h)|^p)_{N \in \mathbb{N}}$ is uniformly integrable conditionally on $Y_0^{n-1}$, that is,

$\lim_{M \to \infty} \sup_{N \in \mathbb{N}^*} \mathbb{E}(|A_{N,\chi}(Y_0^{n-1})(h)|^p \mathbb{1}_{\{|A_{N,\chi}(Y_0^{n-1})(h)| > M\}}|Y_0^{n-1}) = 0, \quad \mathbb{P}$-a.s.
We may now complete the proof by applying Theorem 2, which states that conditionally on $Y_{0}^{n-1}$, as $N \to \infty$,

$$ A_{N,\chi}(Y_{0}^{n-1})(h) \xrightarrow{D} \sigma_{\chi}(Y_{0}^{n-1})(h)Z, $$

where $Z$ is a standard normally distributed random variable. □

We next state the corresponding result for the particle filter approximation, which is obtained along the very same lines as Theorem 12.

**Theorem 13.** Assume (A2). Then, for all functions $h \in F(X)$, constants $p \in \mathbb{R}^+ \setminus \{0\}$ and all initial distributions $\chi \in \mathcal{P}(X)$, P-a.s.,

$$ \lim_{N \to \infty} \sqrt{N} \mathbb{E}^{1/p}(|\bar{\phi}^{N}_{\chi}(Y_{0}^{n})h - \bar{\phi}_{\chi}(Y_{0}^{n})h|^p |Y_{0}^{n-1}) = \sqrt{2}\sigma_{\chi}(Y_{0}^{n-1})(h) \left( \frac{\Gamma((p+1)/2)}{\sqrt{2\pi}} \right)^{1/p}, $$

where $\Gamma$ is the gamma function.

**4. Examples.** In this section, we develop two classes of examples. In Section 4.1 we consider the **linear Gaussian state–space models**, an important model class that is used routinely in time-series analysis. Recall that in the linear Gaussian case, closed-form solutions to the optimal filtering problem can be obtained using the Kalman recursions. However, as an illustration, we analyze this model class under assumptions that are very general. In Section 4.2, we consider a significantly more general class of nonlinear state–space models. In both these examples we will find that assumptions (A1)–(A2) are satisfied and straightforwardly verified.

**4.1. Linear Gaussian state–space models.** The linear Gaussian state–space models form an important class of HMMs. Let $X = \mathbb{R}^{d_x}$ and $Y = \mathbb{R}^{d_y}$ and define state and observation sequences through the linear dynamic system

$$ X_{k+1} = AX_{k} + RU_{k}, $$
$$ Y_{k} = BX_{k} + SV_{k}, $$

where $(U_{k}, V_{k})_{k \geq 0}$ is a sequence of i.i.d. Gaussian vectors with zero mean and identity covariance matrix. The noise vectors are assumed to be independent of $X_0$. Here $U_{k}$ is $d_u$-dimensional, $V_{k}$ is $d_y$-dimensional and the matrices $A$, $R$, $B$ and $S$ have the appropriate dimensions. Note that we cover also the case $d_u < d_x$, for which the prior kernel $Q$ does not admit a transition density with respect to Lebesgue measure.
For any \( n \in \mathbb{N} \), define the observability and controllability matrices \( O_n \) and \( C_n \) by
\[
O_n \triangleq \begin{bmatrix}
B \\
BA \\
BA^2 \\
\vdots \\
BA^{n-1}
\end{bmatrix}
\quad \text{and} \quad
C_n \triangleq \begin{bmatrix}
A^{n-1}R & A^{n-2}R & \cdots & R
\end{bmatrix},
\]
respectively. We assume the following.

(LGSS1) The pair \((A, B)\) is observable, and the pair \((A, R)\) is controllable, that is, there exists \( r \in \mathbb{N} \) such that the observability matrix \( O_r \) and the controllability matrix \( C_r \) have full rank.

(LGSS2) The measurement noise covariance matrix \( S \) has full rank.

(LGSS3) \( \mathbb{E}(\|Y_0\|^2) < \infty \).

We now check assumptions (A1)–(A2). The dimension \( d_u \) of the state noise vector \( U_k \) is in many situations smaller than the dimension \( d_x \) of the state vector \( X_k \) and hence \( R^tR \) may be rank deficient (here \(^t\) denotes the transpose). Some additional notation is required: for any positive matrix \( A \) and vector \( z \) of appropriate dimension, denote \( \|z\|^2_A \triangleq t_A z A^{-1} z \). In addition, define for any \( n \in \mathbb{N} \),
\[
F_n \triangleq D_n^t D_n + S_n^t S_n,
\]
where
\[
D_n \triangleq \begin{bmatrix}
0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
BR & \ddots & 0 \\
BAR & BR & \ddots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
BA^{n-2}R & BA^{n-3}R & \cdots & BR
\end{bmatrix},
\quad
S_n \triangleq \begin{bmatrix}
S & 0 & \cdots & 0 \\
0 & S & \cdots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
0 & \cdots & 0 & S
\end{bmatrix}.
\]
Under (LGSS2), the matrix \( F_n \) is positive definite for any \( n \geq r \). When the state process is initialized at \( x_0 \in \mathcal{X} \), the likelihood of the observations \( y_0^{-1} \in \mathbb{Y} \) is given by
\[
\delta_{x_0} L(y_0^{-1}) \mathbb{1}_\mathcal{X} = (2\pi)^{-nd_x} \det^{-1/2}(F_n) \exp(-\frac{1}{2} \|y_{n-1} - O_n x_0\|^2_{F_n}),
\]
where \( y_{n-1} \triangleq [t_0, t_1, \ldots, t_{n-1}] \) and \( O_n \) is defined in (43).

We first consider (A1). Under (LGSS1), the observability matrix \( O_r \) is full rank, and we have for any compact subset \( K \subset \mathbb{Y}^r \),
\[
\lim_{\|x_0\| \to \infty} \inf_{y_0^{-1} \in K} \|y_{r-1} - O_r x_0\|_{F_r} = \infty,
\]
showing that for all \( \eta > 0 \), we may choose a compact set \( C \subset \mathbb{R}^{d_x} \) such that (20) is satisfied. It remains to prove that any compact set \( C \) is an \( r \)-local Doeblin set satisfying condition (21). For any \( y_0^{r-1} \in \mathbb{Y}^r \) and \( x_0 \in \mathcal{X} \), the measure \( \delta_{x_0} L(y_0^{r-1}) \) is absolutely continuous with respect to the Lebesgue...
measure on \((X, \mathcal{X})\) with Radon–Nikodym derivative \(\ell(y_0^{r-1})(x_0, x_r)\) given (up to an irrelevant multiplicative factor) by
\[
(45) \quad \ell(y_0^{r-1})(x_0, x_r) \propto \det^{-1/2}(G_r) \exp \left( -\frac{1}{2} \left\| \begin{bmatrix} y_{r-1} \\ x_r \end{bmatrix} - \begin{bmatrix} O_r \\ A_r \end{bmatrix} x_0 \right\|_G^2 \right),
\]
where the covariance matrix \(G_r\) is
\[
G_r \triangleq \left[ \begin{array}{cc} D_r \\ C_r \end{array} \right] \left[ \begin{array}{c} \text{tr} D_r & \text{tr} C_r \\ \text{tr} C_r & 0 \end{array} \right] \left[ \begin{array}{c} \text{tr} S_r \\ 0 \end{array} \right].
\]
The proof of \((45)\) relies on the positivity of \(G_r\), which requires further discussion. By construction, the matrix \(G_r\) is nonnegative. For all \(y_{r-1} \in Y^r\) and \(x \in X\), the equation
\[
\left[ \begin{array}{cc} \text{tr} y_{r-1} \\ x \end{array} \right] G_r \left[ \begin{array}{cc} y_{r-1} \\ x_r \end{array} \right] = \| \text{tr} D_r y_{r-1} + \text{tr} C_r x \|^2 + \| \text{tr} S_r y_{r-1} \|^2 = 0
\]
implies that \(\| \text{tr} D_r y_{r-1} + \text{tr} C_r x \|^2 = 0\) and \(\| \text{tr} S_r y_{r-1} \|^2 = 0\). Since the matrix \(S_r\) has full rank, this implies that \(y_{r-1} = 0\). Since also \(C_r\) has full rank [the pair \((A, R)\) is controllable], this implies in turn that \(x = 0\). Therefore, the matrix \(G_r\) is positive definite and the function
\[
(x_0, x_r) \mapsto \left\| \begin{bmatrix} y_{r-1} \\ x_r \end{bmatrix} - \begin{bmatrix} O_r \\ A_r \end{bmatrix} x_0 \right\|_G^2
\]
is continuous for all \(y_{r-1}\). It is therefore bounded on any compact subset of \(X^2\). This implies that every nonempty compact set \(C \subset \mathbb{R}^{d_x}\) is an \(r\)-local Doeblin set, with \(\lambda_C(\cdot) = \lambda^{\text{Leb}}(\cdot)/\lambda^{\text{Leb}}(C)\) and
\[
\epsilon^-_C(y_0^{r-1}) = \left( \lambda^{\text{Leb}}(C) \right)^{-1} \inf_{(x_0, x_r) \in C^2} \ell(y_0^{r-1})(x_0, x_r),
\]
\[
\epsilon^+_C(y_0^{r-1}) = \left( \lambda^{\text{Leb}}(C) \right)^{-1} \sup_{(x_0, x_r) \in C^2} \ell(y_0^{r-1})(x_0, x_r).
\]
Consequently, condition \((21)\) is satisfied for any compact set \(K \subset Y^{r-1}\). It remains to verify \((A1)(iii)\). Under \((\text{LGSS1})\), the measure \(\delta_{x_0}L(y_0^{r-1})\) is absolutely continuous with respect to the Lebesgue measure \(\lambda^{\text{Leb}}\); therefore, for any set \(D \subset \mathbb{R}^{d_x}\),
\[
\inf_{x_0 \in D} \delta_{x_0}L(y_0^{r-1})(D) \geq \inf_{(x_0, x_r) \in D^2} \ell(y_0^{r-1})(x_0, x_r) \lambda^{\text{Leb}}(D).
\]
Take \(D\) to be any compact set with positive Lebesgue measure. Now,
\[
\sup_{(x_0, x_r) \in D^2} \left\| \begin{bmatrix} y_{r-1} \\ x_r \end{bmatrix} - \begin{bmatrix} O_r \\ A_r \end{bmatrix} x_0 \right\|_G^2 \leq 2 \lambda_{\text{max}}(G_r) \left\{ \| y_{r-1} \|^2 + \max_{x \in D} \| x \|^2 [1 + \lambda_{\text{max}}(\text{tr} O_r A_r + \text{tr} A_r A_r)] \right\},
\]
where \( \lambda_{\text{max}}(A) \) is the largest eigenvalue of \( A \). Under (LGSS3), \( \mathbb{E}(\|Y_0\|^2) < \infty \), implying that (A1)(iii) is satisfied for any compact set.

We now consider (A2). Under (LGSS2), \( S \) has full rank, and taking the reference measure \( \lambda^{\text{Leb}} \) as the Lebesgue measure on \( Y \), \( g(x,y) \) is, for each \( x \in X \), a Gaussian density with covariance matrix \( S^tS \). We therefore have

\[
\|g(y)\|_\infty = (2\pi)^{-d_y/2} \det^{-1/2}(S^tS) < \infty
\]

for all \( y \in Y \), which verifies (A2)(i)-(ii).

To conclude this discussion, we need to specify more explicitly the set \( \mathcal{M}(D,r) \) [see (24)] of possible initial distributions. Using Proposition 7, we verify sufficient conditions (25) and (26). To check (25), we use Remark 9: for any open subset \( O \subset \mathbb{R}^d \) and \( x \in X \), \( Q(x,O) = \mathbb{E}(\mathbb{1}_O(Ax + RU)) \), where the expectation is taken with respect to the \( d_u \)-dimensional standard normal distribution. Let \( (x_n)_{n \in \mathbb{N}} \) be a sequence in \( X \) converging to \( x \). By using that the function \( \mathbb{1}_O \) is lower semi-continuous we obtain, via Fatou’s lemma,

\[
\liminf_{n \to \infty} Q(x_n, O) \geq \mathbb{E}\left(\liminf_{n \to \infty} \mathbb{1}_O(Ax_n + RU) \right) \geq Q(x, O),
\]

showing that the function \( x \mapsto Q(x, O) \) is lower semi-continuous for any open subset \( O \).

Assumption (LGSS2) implies that for all \( (x,y) \in X \times Y \),

\[
\ln g(x,y) \geq -\frac{d_y}{2} \ln(2\pi) - \frac{1}{2} \ln \det^{-1/2}(S^tS)
\]

\[
- [\lambda_{\text{min}}(S^tS)]^{-1}(\|y\|^2 + \|Bx\|^2),
\]

where \( \lambda_{\text{min}}(S^tS) \) is the minimal eigenvalue of \( S^tS \). Therefore (26) is satisfied under (LGSS3). Consequently, we may apply Theorems 10 and 11 to establish tightness of the asymptotic variances of the particle predictor and filter approximations for any initial distribution \( \chi \in \mathcal{P}(X) \) as soon as the process \( (Y_k)_{k \in \mathbb{Z}} \) is strictly stationary ergodic and \( \mathbb{E}(\|Y_0\|^2) < \infty \).

4.2. Nonlinear state–space models. We now turn to a very general class of nonlinear state–space models. Let \( X = \mathbb{R}^d \), \( Y = \mathbb{R}^\ell \) and \( \mathcal{X} \) and \( \mathcal{Y} \) be the associated Borel \( \sigma \)-fields. In the following we assume that for each \( x \in X \), the probability measure \( Q(x,\cdot) \) has a density \( q(x,\cdot) \) with respect to the Lebesgue measure \( \lambda^{\text{Leb}} \) on \( \mathbb{R}^d \). For instance, the state sequence \( (X_k)_{k \in \mathbb{N}} \) could be defined through some nonlinear recursion

\[
X_k = T(X_{k-1}) + \Sigma(X_{k-1})\zeta_k,
\]

where \( (\zeta_k)_{k \in \mathbb{N}} \) is an i.i.d. sequence of \( d \)-dimensional random vectors with density \( \rho_\zeta \) with respect to the Lebesgue measure \( \lambda^{\text{Leb}} \) on \( \mathbb{R}^d \). Here \( T: \mathbb{R}^d \to \mathbb{R}^d \) and \( \Sigma: \mathbb{R}^d \to \mathbb{R}^{d \times d} \) are given (measurable) matrix-valued functions such
that \( \Sigma(x) \) is full rank for each \( x \in X \). Models of form (46), typically referred to as vector autoregressive conditional heteroscedasticity (ARCH) models, are often of interest in time series analysis and financial econometrics. In this context, we let the observations \((Y_k)_{k \in \mathbb{N}}\) be generated through a given measurement density \( g(x,y) \) (again with respect to the Lebesgue measure).

We now introduce the basic assumptions of this section.

(NL1) The function \((x, x') \mapsto q(x, x') \) on \( X^2 \) is positive and continuous. In addition, \( \sup_{(x,x') \in X^2} q(x, x') < \infty \).

(NL2) For any compact subset \( K \subset Y \), \( \lim_{\|x\| \to \infty} \sup_{y \in K} \|g(x,y)\|_\infty = 0 \).

(NL3) For all \((x,y) \in X \times Y \), \( g(x,y) > 0 \) and \( \mathbb{E}(\ln^+ \|g(Y_0)\|_\infty) < \infty \).

(NL4) There exists a compact subset \( D \subset Y \) such that \( \mathbb{E}(\ln^- \inf_{x \in D} g(x,Y_0)) < \infty \).

Under (NL1), every compact set \( C \subset X = \mathbb{R}^d \) with positive Lebesgue measure is 1-small and therefore local Doeblin with \( \lambda_C(\cdot) = \lambda^{\text{Leb}}(\cdot \cap C)/\lambda^{\text{Leb}}(C) \), \( \varphi_C(y_0) = \lambda^{\text{Leb}}(C) \) and

\[
\epsilon_C^- = \inf_{(x,x') \in C^2} q(x,x'), \\
\epsilon_C^+ = \sup_{(x,x') \in C^2} q(x,x').
\]

Under (NL1) and (NL2), conditions (23) and (21) are satisfied with \( r = 1 \). In addition, (22) is implied by (NL1) and (NL4). Consequently, assumption (A1) holds. Moreover, (A2) follows directly from (NL3). So, finally, under (NL1)–(NL4) we conclude, using Proposition 7, Theorems 10 and 11, that the asymptotic variances of the bootstrap particle predictor and filter approximations are tight for any initial distribution \( \chi \) such that \( \chi(D) > 0 \).

5. Proofs.

5.1. Forgetting of the initial distribution.

**Lemma 14.** Assume (A1)–(A2). Then for all \( \gamma > 2/3 \) there exist functions \( \rho_\gamma : [0,1[ \to [0,1[ \) and \( C_\gamma : [0,1[ \to \mathbb{R}_+ \) such that for all \( n \in \mathbb{N} \) and all \( z_0^{r-1} \in Y^{nr} \), where \( r \in \mathbb{N}^* \) is as in (A1) and \( z_i = y^{(i+1)r-1}_{ir} \), satisfying

\[
n^{-1} \sum_{i=0}^{n-1} \mathbb{1}_K(z_i) \geq \gamma,
\]
all functions $f$ and $h$ in $\mathcal{F}_+(\mathcal{X})$, all finite measures $\chi$ and $\chi'$ in $\mathcal{M}(\mathcal{X})$, and all $\eta \in [0,1]$,  
\begin{equation}
|\Delta_{\chi,\chi'}(z_0^{n-1})(f,h)|
\end{equation}
\begin{equation}
\leq \rho_{\gamma}(\eta)\chi L(z_0^{n-1})f \times \chi' L(z_0^{n-1})h + \chi' L(z_0^{n-1})f \times \chi L(z_0^{n-1})h
\end{equation}
\begin{equation}
+ C_{\gamma}(\eta)\eta^n \|f\|_{\infty}\|h\|_{\infty} \prod_{i=0}^{n-1} \|L(z_i)\|_{\infty}^2 \chi(X)\chi'(X),
\end{equation}
\begin{equation}
\left| \ln\left(\frac{\chi L(z_0^{n-1})f}{\chi L(z_0^{n-1})}\right) - \ln\left(\frac{\chi' L(z_0^{n-1})f}{\chi' L(z_0^{n-1})}\right) \right|
\end{equation}
\begin{equation}
\leq (1 - \rho_{\gamma}(\eta))^{-1}
\end{equation}
\begin{equation}
\times \left(2\rho_{\gamma}(\eta) + \frac{C_{\gamma}(\eta)\eta^n \|f\|_{\infty}\|h\|_{\infty} \prod_{i=0}^{n-1} \|L(z_i)\|_{\infty} \chi(X)\chi'(X)}{\chi L(z_0^{n-1})f \times \chi' L(z_0^{n-1})h} \right),
\end{equation}
\begin{equation}
\left| \frac{\chi L(z_0^{n-1})h}{\chi L(z_0^{n-1})f} - \frac{\chi' L(z_0^{n-1})h}{\chi' L(z_0^{n-1})f} \right|
\end{equation}
\begin{equation}
\leq \rho_{\gamma}(\eta)\left(\frac{\chi L(z_0^{n-1})h}{\chi L(z_0^{n-1})f} + \frac{\chi' L(z_0^{n-1})h}{\chi' L(z_0^{n-1})f} \right)
\end{equation}
\begin{equation}
+ \frac{C_{\gamma}(\eta)\eta^n \|f\|_{\infty}\|h\|_{\infty} \prod_{i=0}^{n-1} \|L(z_i)\|_{\infty}^2 \chi(X)\chi'(X)}{\chi L(z_0^{n-1})f \times \chi' L(z_0^{n-1})f}.
\end{equation}

**Proof.** The proof is adapted straightforwardly from [15], Proposition 5. □

**Lemma 15.** Assume (A1). Then there exists a constant $\kappa > 0$ such that for all $\chi \in \mathcal{M}(\mathcal{D},r)$ [where $\mathcal{M}(\mathcal{D},r)$ is defined in (24)],
\begin{equation}
\inf_{(k,m) \in \mathbb{N}^* \times \mathbb{N}} \kappa_{(k,m)} L(Y_{r-1}^{k-1})1_{\mathcal{X}} > 0, \quad \mathbb{P}\text{-a.s.,}
\end{equation}
and
\begin{equation}
\inf_{(k,m) \in \mathbb{N}^* \times \mathbb{N}} \kappa_{(k,m)} L(Y_{r-1}^{k-1})1_{\mathcal{X}} > 0, \quad \mathbb{P}\text{-a.s.}
\end{equation}

**Proof.** To derive (50) we first establish that
\begin{equation}
\liminf_{k+m \to \infty} (k + m)^{-1} \ln \chi L(Y_{r-1}^{k-1})1_{\mathcal{X}}
\end{equation}
\begin{equation}
\geq -r\mathbb{E}\left(\ln \inf_{x \in \mathcal{D}} \delta_x L(Y_{r-1}^{r-1})1_{\mathcal{D}}\right) > -\infty, \quad \mathbb{P}\text{-a.s.,}
\end{equation}
where the last inequality follows from (A1)(iii). We now establish the first inequality in (52). Set $a_{k,m} \triangleq -k + [(k + m)/r]r$ and note that $-a_{k,m} \in \{-m, \ldots, -m + r - 1\}$. Then write
\[
\ln \chi L(Y_{-m}^{k-1}) \mathbb{1}_X \\
\geq \ln \chi L(Y_{-m}^{-a_{k,m}}) \mathbb{1}_D + \sum_{i=0}^{[(k+m)/r]-1} \ln \inf_{x \in D} \delta_x L(Y_{-a_{k,m}+(i+1)r-1}) \mathbb{1}_D \\
\geq -\sum_{i=0}^{r-1} \ln \chi L(Y_{-m+i}) \mathbb{1}_D - \sum_{i=0}^{[(k+m)/r]-1} \ln \inf_{x \in D} \delta_x L(Y_{-a_{k,m}+(i+1)r-1}) \mathbb{1}_D.
\]
(53)

For $i \in \mathbb{N}$, set $[i]_r \triangleq i - |i/r| r$. With this notation, $a_{k,m} = [a_{k,m}]_r + [a_{k,m}/r] r$. Then, since $[i]_r \in \{0, \ldots, r-1\}$,
\[
\sum_{i=0}^{[(k+m)/r]-1} \ln \inf_{x \in D} \delta_x L(Y_{-a_{k,m}+(i+1)r-1}) \mathbb{1}_D = -\sum_{i=0}^{r-1} \ln \inf_{x \in D} \delta_x L(Y_{-a_{k,m}+(i+1)r-1}) \mathbb{1}_D
\]
(54)

where the last identity follows by reindexing the summation. We now plug (54) into (53); the ergodicity of the process $(Z_n)_{n \in \mathbb{Z}}$ [assumption (A1)(i)] then implies, via Lemma 18, $\mathbb{P}$-a.s.,
\[
\liminf_{k+m \to \infty} (k+m)^{-1} \ln \chi L(Y_{-m}^{k-1}) \mathbb{1}_X \\
= -\sum_{j=0}^{r-1} \mathbb{E} \left( \ln \inf_{x \in D} \delta_x L(Y_{-j+r-1}) \mathbb{1}_D \right) = -r \mathbb{E} \left( \ln \inf_{x \in D} \delta_x L(Y_0^{r-1}) \mathbb{1}_D \right),
\]

which shows (52). Now, choose a constant $\kappa$ such that
\[
-r \mathbb{E} \left( \ln \inf_{x \in D} \delta_x L(Y_0^{r-1}) \mathbb{1}_D \right) > -\ln \kappa > -\infty.
\]
According to (52), there exists a $\mathbb{P}$-a.s. finite $\mathbb{N}^*$-valued random variable $N$ such that if $k + m \geq N$,

$$\ln \chi L(Y_0^{r-1})1_X \geq (-\ln \kappa)(k + m),$$

which implies that

$$\inf_{k + m \geq N} \kappa^{k + m} \chi L(Y_0^{r-1})1_X \geq 1.$$

On the other hand, assumption (A2) implies that for all $(k, m) \in \mathbb{N}^* \times \mathbb{N}$, $\chi L(Y_0^{r-1})1_X > 0$, $\mathbb{P}$-a.s. This completes the proof of (50). Finally, the proof of (51) follows by combining

$$\|L(Y_m^{r-1})1_X\|_{\infty} \geq \chi L(Y_m^{r-1})1_X$$

and (50). \qed

For all probability measures $\chi \in \mathcal{P}(\mathcal{X})$, all $(k, m) \in \mathbb{N}^* \times \mathbb{N}$, and all sequences $y_{-m}^k \in Y^{m+k+1}$, define the set

$$\mathcal{M}(y_{-m}^k)(\chi)$$

\[\triangleq \{ \tilde{\chi} \in \mathcal{P}(\mathcal{X}) : \|g(y_k)\|_{\infty} \times \tilde{\chi} L(y_{-m}^{k-1})1_X \geq (1/2) \chi L(y_{-m}^{k-1})1_X \}\]

of probability measures on $(\mathcal{X}, \mathcal{X})$ and note that this set is nonempty since $\chi \in \mathcal{M}(y_{-m}^k)(\chi)$. The choice of $1/2$ in the definition of $\mathcal{M}(y_{-m}^k)(\chi)$ is irrelevant, and this factor can be replaced by any constant strictly less than 1.

**Proposition 16.** Assume (A1)--(A2). Then there exists a constant $\beta \in [0, 1]$ such that the following hold:

(i) For all probability measures $\chi$ and $\chi'$ in $\mathcal{M}(D, r)$ there exists a $\mathbb{P}$-a.s. finite random variable $C_{\chi, \chi'}$ such that for all $(k, m) \in \mathbb{N}^* \times \mathbb{N}$ and all $\tilde{\chi} \in \mathcal{M}(y_{-m}^k)(\chi)$,

$$\ln \left( \frac{\tilde{\chi} L(y_{-m}^k)1_X}{\chi L(y_{-m}^k)1_X} \right) - \ln \left( \frac{\chi' L(y_{-m}^k)1_X}{\chi L(y_{-m}^k)1_X} \right) \leq C_{\chi, \chi'} \beta^{k+m}, \quad \mathbb{P}\text{-a.s.}$$

(ii) For all probability measures $\chi$ in $\mathcal{M}(D, r)$ there exists a $\mathbb{P}$-a.s. finite random variable $C_{\chi}$ such that for all $(k, m) \in \mathbb{N}^* \times \mathbb{N}$,

$$\ln \left( \frac{\tilde{\chi} L(y_{-m}^k)1_X}{\chi L(y_{-m}^k)1_X} \right) - \ln \left( \frac{\chi L(y_{-m}^{k-1})1_X}{\chi L(y_{-m}^{k-1})1_X} \right) \leq C_{\chi} \beta^{k+m}, \quad \mathbb{P}\text{-a.s.}$$

(iii) There exists a $\mathbb{P}$-a.s. finite random variable $C$ such that for $m \in \mathbb{N}^*$, all probability measures $\chi$ and $\chi'$ in $\mathcal{P}(\mathcal{X})$ and all $h \in \mathcal{F}(\mathcal{X})$,

$$\frac{|\Delta_{\chi, \chi'}(Y_{-m}^r)(h, 1_X)|}{\|L(Y_{-m}^r)1_X\|_{\infty}^2} \leq C \beta^m \|h\|_{\infty}, \quad \mathbb{P}\text{-a.s.}$$
Proof of Proposition 16(i) and (ii). Let $\check{\chi} \in \mathcal{M}(Y_{-m}^k)(\chi)$. Recall the notation $Z_i = Y_{ir(i+1)r^{-1}}$ and consider the decompositions
\[
\chi \mathbf{L}(Y_{-m}^k) \mathbb{1}_X = \chi \mathbf{L}(Y_{-m}^{[m/r]r^{-1}}) \mathbf{L}(Z_{[k/r]-1}) \mathbf{L}(Y_{-[k/r]}) \mathbf{L}(Y_{k-1}) \mathbb{1}_X,
\]
\[
\chi \mathbf{L}(Y_{-m}^{k-1}) \mathbb{1}_X = \chi \mathbf{L}(Y_{-m}^{[m/r]r^{-1}}) \mathbf{L}(Z_{[k/r]-1}) \mathbf{L}(Y_{k-1}) \mathbb{1}_X,
\]
where we make use of convention (9) if necessary.

Choose $\gamma$ such that $2/3 < \gamma < \mathbb{P}(Z_0 \in K)$, where $K$ is defined in (A1)(i). Assume that $(k, m) \in \mathbb{N}^* \times \mathbb{N}$ are both larger than $r$ and denote by $b_{k,m} \triangleq \lfloor k/r \rfloor + \lfloor m/r \rfloor$. In addition, define the event
\[
\Omega_{k,m} \triangleq \left\{ \left( \left\lfloor \frac{k}{r} \right\rfloor + \left\lfloor \frac{m}{r} \right\rfloor \right)^{-1} \sum_{\ell = \lfloor m/r \rfloor}^{\lfloor k/r \rfloor} \mathbb{1}_K(Z_\ell) \geq \gamma \right\}.
\]
By Lemma 14 [equation (48)] it holds for all $\eta \in ]0, 1[$, on the event $\Omega_{k,m}$,
\[
(1 - \rho_\gamma(\eta)) \left( \ln \left( \frac{\check{\chi} \mathbf{L}(Y_{-m}^k) \mathbb{1}_X}{\check{\chi} \mathbf{L}(Y_{-m}^{k-1}) \mathbb{1}_X} \right) - \ln \left( \frac{\chi \mathbf{L}(Y_{-m}^k) \mathbb{1}_X}{\chi \mathbf{L}(Y_{-m}^{k-1}) \mathbb{1}_X} \right) \right) \leq 2 \rho_{\gamma} b_{k,m}(\eta) + \frac{C_{\gamma}(\eta) \eta b_{k,m} \|g(Y_i)\|_\infty \prod_{i=-m}^{k-1} \|g(Y_i)\|_\infty^2}{\check{\chi} \mathbf{L}(Y_{-m}^k) \mathbb{1}_X \times \chi \mathbf{L}(Y_{-m}^{k-1}) \mathbb{1}_X},
\]
where:
\[(a)\] follows from (48) and the bound $\delta_x \mathbf{L}(Y_{-m}^k) \mathbb{1}_X \leq \prod_{\ell = u}^v \|g(Y_\ell)\|_\infty$, valid for $u \leq v$, and
\[(b)\] follows from the fact that $\check{\chi} \in \mathcal{M}(Y_{-m}^k)(\chi)$.

Since, under (A1)(i), the sequence $(Z_n)_{n \in \mathbb{Z}}$ is ergodic and $\mathbb{P}(Z_0 \in K) > \gamma$, Lemma 18 implies that
\[
\mathbb{P} \left( \bigcup_{j \geq 0} \bigcap_{(k, m) \in \mathbb{N}^* \times \mathbb{N}} \Omega_{k,m} \right) = 1.
\]
Hence, there exists a $\mathbb{P}$-a.s. finite integer-valued random variable $U$ such that (56) is satisfied for all $(k, m) \in \mathbb{N}^* \times \mathbb{N}$ such that $k + m \geq U$.

The lower bound obtained in Lemma 15 implies that there exists a constant $\kappa > 0$ such that for all probability measures $\chi$ and $\chi'$ in $\mathcal{M}(D, r)$ and all $(k, m) \in \mathbb{N}^* \times \mathbb{N}$, $\mathbb{P}$-a.s.,
\[
\chi \mathbf{L}(Y_{-m}^k) \mathbb{1}_X \geq C_{\chi, \chi'} \kappa^{-(k+m+1)},
\]
\[
\chi' \mathbf{L}(Y_{-m}^k) \mathbb{1}_X \geq C_{\chi, \chi'} \kappa^{-(k+m+1)},
\]
where \( C_{\chi,\chi'} \) is a \( \mathbb{P} \)-a.s. finite constant.

By plugging these bounds into (56) and using Lemma 19 with \( \eta \) sufficiently small (note that (56) is satisfied for all \( \eta \in [0,1] \)), we conclude that there exist a \( \mathbb{P} \)-a.s. finite random variable \( C_{\chi,\chi'} \) and a constant \( \beta < 1 \) such that for all \((k,m) \in \mathbb{N}^* \times \mathbb{N} \), \( \mathbb{P} \)-a.s.,
\[
\ln \left( \frac{\hat{\chi} \mathcal{L}(Y_m^k) \mathbb{1}_X}{\hat{\chi'} \mathcal{L}(Y_m^{k-1}) \mathbb{1}_X} \right) - \ln \left( \frac{\chi \mathcal{L}(Y_m^k) \mathbb{1}_X}{\chi' \mathcal{L}(Y_m^{k-1}) \mathbb{1}_X} \right) \leq C_{\chi,\chi'} \beta^{k+m},
\]
which completes the proof of Proposition 16(i). Note that \( \chi \in \mathcal{M}(Y_m^k) \) implies that the previous relation is satisfied with \( \check{\chi} = \chi \).

The proof of Proposition 16(ii) follows the same lines as the proof of Proposition 16(i) and is omitted for brevity. \( \square \)

**Proof of Proposition 16(iii).** We may assume that the function \( h \) is nonnegative (otherwise the positive and negative parts of \( h \) can be treated separately). As in the proof of Proposition 16(i), write
\[
\chi \mathcal{L}(Y_m^{-1})h = \chi \mathcal{L}(Y_m^{-1})^{r-1} \mathcal{L}(Z_m^{-1})h
\]
and define the event
\[
\Omega_m \triangleq \left\{ \left[ \frac{m}{r} \right] - \sum_{\ell=-\left[ \frac{m}{r} \right]}^{1} \mathbb{1}_K(Z_\ell) \geq \gamma \right\}.
\]
By Lemma 14 [equation (49)] it holds, on the event \( \Omega_m \),
\[
\left| \frac{\chi \mathcal{L}(Y_m^{-1})h}{\chi \mathcal{L}(Y_m^{-1})^2} - \frac{\chi' \mathcal{L}(Y_m^{-1})h}{\chi' \mathcal{L}(Y_m^{-1})^2} \right| \leq 2\|h\|_{\infty} \rho_{\gamma}^{\left[ m/r \right]}(\eta) + \frac{C_{\gamma}(\eta)\eta^{\left[ m/r \right]} \|h\|_{\infty} \prod_{i=m}^{-1} \|g(Y_i)\|_{\infty}^2}{\chi \mathcal{L}(Y_m^{-1})^2 \times \chi' \mathcal{L}(Y_m^{-1})^2},
\]
where we used that for \( u \leq v \), \( \delta_x \mathcal{L}(Y_u^v) \mathbb{1}_X \leq \prod_{\ell=u}^{v} \|g(Y_\ell)\|_{\infty}. \) Under (A1)(i), Birkhoff’s ergodic theorem (see, e.g., [32]) ensures that \( \mathbb{P}(\liminf_{m \to \infty} \Omega_m) = 1; \) therefore, there exists a \( \mathbb{P} \)-a.s. finite random variable \( U \) such that (57) is satisfied for \( m \geq U \). Then, for \( m \geq U \),
\[
\frac{|\Delta_{\chi,\chi'}(Y_m^{-1})h, \mathbb{1}_X|}{\|\mathcal{L}(Y_m^{-1})\mathbb{1}_X\|^2} = \frac{\chi \mathcal{L}(Y_m^{-1})^2 \times \chi' \mathcal{L}(Y_m^{-1})^2}{\left| \chi \mathcal{L}(Y_m^{-1})^2 \times \chi' \mathcal{L}(Y_m^{-1})^2 \right|^2} \leq 2\|h\|_{\infty} \rho_{\gamma}^{\left[ m/r \right]}(\eta) + \frac{C_{\gamma}(\eta)\eta^{\left[ m/r \right]} \|h\|_{\infty} \prod_{i=m}^{-1} \|g(Y_i)\|_{\infty}^2}{\|\mathcal{L}(Y_m^{-1})\mathbb{1}_X\|^2},
\]

30

R. DOUC, E. MOULINES AND J. OLSSON
we have used that $\chi L(Y_{-m})\mathbb{1}_X \leq \|L(Y_{-m})\mathbb{1}_X\|_\infty$. By Lemma 15, equation (51), there exist a constant $\kappa > 0$ and a $\mathbb{P}$-a.s. finite random variable $\mathbb{C}$ such that

$$\|L(Y_{-m})\mathbb{1}_X\|_\infty \geq \mathbb{C} \kappa^{-m}, \quad \mathbb{P}\text{-a.s.}$$

Finally, we complete the proof by inserting this bound into (58) and applying Lemma 19 to the right-hand side of the resulting inequality. \hfill \Box

5.2. Convergence of the log-likelihood.

**Lemma 17.** Assume (A1)–(A2). Then, $\mathbb{P}$-a.s.,

$$\lim_{n \to \infty} n^{-1} \ln \|L(Y_0^n)\mathbb{1}_X\|_\infty = \ell_\infty,$$

$$\lim_{n \to \infty} n^{-1} \ln \|L(Y_0^n)\mathbb{1}_X\|_\infty = \ell_\infty,$$

$$\lim_{n \to \infty} n^{-1} \sum_{k=1}^n \ln \pi(Y_{-k}) (Y_{-k}) = \ell_\infty,$$

where $\ell_\infty$ is defined in (33).

**Proof of (59).** Let $(\alpha_n)_{n \in \mathbb{N}^*}$ be a nondecreasing sequence such that $\lim_{n \to \infty} \alpha_n = 1$ and for any $n \in \mathbb{N}^*$, $\alpha_n \geq 1/2$. For all $n \in \mathbb{N}$, choose $\tilde{x}_n \in X$ such that

$$\alpha_n \|L(Y_0^n)\mathbb{1}_X\|_\infty \leq \delta_{\tilde{x}_n} L(Y_0^n) \mathbb{1}_X \leq \|L(Y_0^n)\mathbb{1}_X\|_\infty.$$

Note that for all $k \in \mathbb{N}^*$,

$$\delta_{\tilde{x}_{k-1}} L(Y_{0}^{k-1}) \mathbb{1}_X \geq \alpha_{k-1} \|L(Y_0^{k-1}) \mathbb{1}_X\|_\infty \geq \alpha_{k-1} \delta_{\tilde{x}_{k}} L(Y_0^{k-1}) \mathbb{1}_X.$$

On the other hand, for all probability measures $\chi \in \mathcal{P}(\mathcal{X})$ it holds that

$$\delta_{\tilde{x}_k} L(Y_0^{k-1}) \mathbb{1}_X \geq \frac{\delta_{\tilde{x}_k} L(Y_0^{k-1}) \mathbb{1}_X}{\|L(Y_0^{k-1}) \mathbb{1}_X\|_\infty} \geq \alpha_k \frac{\|L(Y_0) \mathbb{1}_X\|_\infty}{\|L(Y_0) \mathbb{1}_X\|_\infty} \geq \alpha_k \frac{\chi L(Y_0^{k}) \mathbb{1}_X}{\|L(Y_0^{k}) \mathbb{1}_X\|_\infty},$$

where (a) follows from the bound $\delta_{\tilde{x}_k} L(Y_0^{k-1}) \mathbb{1}_X \leq \|L(Y_0) \mathbb{1}_X\|_\infty \delta_{\tilde{x}_k} L(Y_0^{k-1}) \mathbb{1}_X$ and (b) stems from the definition (62) of $\alpha_n$. Then

$$0 \leq -n^{-1} \ln \|L(Y_0^n)\mathbb{1}_X\|_\infty - \ln \chi L(Y_0^n) \mathbb{1}_X$$

$$\leq -n^{-1} \ln \alpha_n + n^{-1} \ln(\alpha_n \|L(Y_0^n)\mathbb{1}_X\|_\infty) - \ln \chi L(Y_0^n) \mathbb{1}_X$$

$$\leq -n^{-1} \ln \alpha_n + n^{-1} \ln(\delta_{\tilde{x}_n} L(Y_0^n) \mathbb{1}_X) - \ln \chi L(Y_0^n) \mathbb{1}_X$$

$$= -n^{-1} \ln \alpha_n + n^{-1} \ln(\delta_{\tilde{x}_n} L(Y_0^{k-1}) \mathbb{1}_X) - \ln \chi L(Y_0^n) \mathbb{1}_X$$

$$+ n^{-1} \sum_{k=1}^n \left[ \ln \left( \frac{\delta_{\tilde{x}_n} L(Y_0^{k}) \mathbb{1}_X}{\delta_{\tilde{x}_{k-1}} L(Y_0^{k-1}) \mathbb{1}_X} \right) - \ln \left( \frac{\chi L(Y_0^{k}) \mathbb{1}_X}{\chi L(Y_0^{k-1}) \mathbb{1}_X} \right) \right].$$
For each term in the sum it holds, by (63),
\[
\ln \left( \frac{\delta \xi_k L(Y_0^{k-1})_{\|X}}{\delta \xi_{k-1}^{-1} L(Y_0^{k-1})_{\|X}} \right) - \ln \left( \frac{\chi L(Y_0^k)_{\|X}}{\chi L(Y_0^{k-1})_{\|X}} \right) \\
\leq - \ln \alpha_k - 1 + \ln \left( \frac{\delta \xi_k L(Y_0^k)_{\|X}}{\delta \xi_{k-1}^{-1} L(Y_0^{k-1})_{\|X}} \right) - \ln \left( \frac{\chi L(Y_0^k)_{\|X}}{\chi L(Y_0^{k-1})_{\|X}} \right).
\]

For all \( k \in \mathbb{N}^* \), (64) implies that
\[
\delta \xi_k L(Y_0^{k-1})_{\|X} \geq \frac{1}{2} \chi L(Y_0^k)_{\|X},
\]
so that \( \delta \xi_k \) belongs to the set \( \mathcal{M}(Y_0^{k-1})(\chi) \) [defined in (55)]. Proposition 16(i) then provides a constant \( \beta \in ]0,1[ \) and a \( \mathbb{P} \)-a.s. finite random variable \( C_\chi \) such that
\[
\ln \left( \frac{\delta \xi_k L(Y_0^k)_{\|X}}{\delta \xi_k L(Y_0^{k-1})_{\|X}} \right) - \ln \left( \frac{\chi L(Y_0^k)_{\|X}}{\chi L(Y_0^{k-1})_{\|X}} \right) \leq C_\chi \beta^k.
\]

Finally, statement (59) follows by plugging the bound (66) into (65), letting \( n \) tend to infinity and using (32). \( \square \)

**Proof of (60).** For all \((p,n) \in \mathbb{N}^2 \) such that \( p \leq n \), define \( W_{p,n} \triangleq \ln \|L(Y_p^{n-1})_{\|X}\|_\infty \) and \( \tilde{W}_{p,n} \triangleq \ln \|L(Y_{n+1}^{-p})_{\|X}\|_\infty \). Note that these two sequences are subadditive in the sense that for all \((p,n) \in \mathbb{N}^2 \) such that \( p \leq n \),
\[
W_{0,n} \leq W_{0,p} + W_{p,n}, \\
\tilde{W}_{0,n} \leq \tilde{W}_{0,p} + \tilde{W}_{p,n}.
\]

Finally, for all \( x \in D \), \( m \in \mathbb{N} \) and \( y_0^{m-1} \in Y^m \), it holds that
\[
\|L(y_0^{m-1})_{\|X}\|_\infty \geq \delta_x L(y_0^{m-1})_{\|X} \geq \prod_{\ell=0}^{m-1} \inf_{x \in D} \delta_x L(y_{kr}^{(k+1)r-1})_{\|D}.
\]

Using the stationarity of the observation process \((Y_k)_{k \in \mathbb{Z}} \), we get, via assumption (A1)(iii), for all \( m \in \mathbb{N}^* \),
\[
(mr)^{-1} \mathbb{E}(W_{0,mr}) \geq (mr)^{-1} \mathbb{E}(\tilde{W}_{0,mr}) \geq (mr)^{-1} \mathbb{E}(\ln \|L(y_0^{m-1})_{\|X}\|_\infty)
\geq r^{-1} \mathbb{E}(\ln \inf_{x \in D} \delta_x L(y_{kr}^{(k+1)r-1})_{\|D} > -\infty.
\]

The sequences \((\mathbb{E}(W_{0,n}))_{n \in \mathbb{N}^*} \) and \((\mathbb{E}(\tilde{W}_{0,n}))_{n \in \mathbb{N}^*} \) are subadditive; Fekete’s lemma (see [30]) thus implies that the sequences \((n^{-1} \mathbb{E}(W_{0,n}))_{n \in \mathbb{N}^*} \) and
\((n^{-1}E(\widetilde{W}_{0,n}))_{n \in \mathbb{N}^*}\) have limits in \([-\infty, \infty]\) and that
\[
\lim_{n \to \infty} n^{-1}E(W_{0,n}) = \lim_{n \to \infty} n^{-1}E(\widetilde{W}_{0,n}) = \inf_{n \in \mathbb{N}^*} n^{-1}E(W_{0,n}) = \inf_{n \in \mathbb{N}^*} n^{-1}E(\widetilde{W}_{0,n}).
\]

However, by (68) there exists a subsequence that is bounded away from \(-\infty\), showing that
\[
\inf_{n \in \mathbb{N}^*} n^{-1}E(W_{0,n}) = \lim_{n \to \infty} n^{-1}E(W_{0,n}) > -\infty,
\]
\[
\inf_{n \in \mathbb{N}^*} n^{-1}E(\widetilde{W}_{0,n}) = \lim_{n \to \infty} n^{-1}E(\widetilde{W}_{0,n}) > -\infty.
\]

Now, by applying Kingman’s subadditive ergodic theorem (see \cite{22}) and using again that \(E(\widetilde{W}_{0,k}) = E(W_{0,k})\) under stationarity, we obtain
\[
\lim_{n \to \infty} n^{-1}\tilde{W}_{0,n} = \inf_{n \in \mathbb{N}^*} n^{-1}E(\widetilde{W}_{0,n}) = \inf_{n \in \mathbb{N}^*} n^{-1}E(W_{0,n}) = \lim_{n \to \infty} n^{-1}W_{0,n} = \ell_{\infty}, \quad \mathbb{P}\text{-a.s.,}
\]
where the last limit follows from (59). This completes the proof of statement (60). □

**Proof of (61).** Since \(E(|\ln \pi (Y_{-\infty}^{-1}(Y_0))|) < \infty\) and the process \((Y_k)_{k \in \mathbb{Z}}\) is stationary and ergodic, (61) follows from Birkhoff’s ergodic theorem. □

**APPENDIX: TECHNICAL LEMMAS**

**Lemma 18.** If \((U_n)_{n \in \mathbb{Z}}\) is a stationary and ergodic sequence of random variables such that \(E(|U_0|) < \infty\), then
\[
\lim_{k+m \to \infty} (k+m)^{-1}\left(\sum_{\ell=-m}^{k-1} U_{\ell}\right) = E(U_0), \quad \mathbb{P}\text{-a.s.}
\]

**Proof.** Denote
\[
\Omega_1 \triangleq \left\{ \omega \in \Omega; \lim_{k+m \to \infty} (k+m)^{-1}\left(\sum_{\ell=-m}^{k-1} U_{\ell}(\omega)\right) = E(U_0) \right\},
\]
\[
\Omega_2 \triangleq \left\{ \omega \in \Omega; \lim_{m \to \infty} \frac{\sum_{\ell=-m}^{k-1} U_{\ell}(\omega)}{m} = \lim_{k \to \infty} \frac{\sum_{\ell=0}^{k-1} U_{\ell}(\omega)}{k} = E(U_0) \right\}.
\]
By Birkhoff’s ergodic theorem, \( \mathbb{P}(\Omega_2) = 1 \). To obtain (69), it is thus sufficient to show that \( \Omega_1 \cap \Omega_2 = \emptyset \). The proof is by contradiction. Assume \( \Omega_1 \cap \Omega_2 \neq \emptyset \), so that there exists \( \omega \in \Omega_1 \cap \Omega_2 \). For such \( \omega \), the fact that \( \omega \notin \Omega_1 \) implies that there exist a positive number \( \epsilon(\omega) > 0 \) and integer-valued sequences \( (k_n(\omega))_{n \in \mathbb{N}} \) and \( (m_n(\omega))_{n \in \mathbb{N}} \) such that \( k_n(\omega) + m_n(\omega) \geq n \) and for all \( n \geq 0 \),

\[
\left| \frac{\sum_{\ell=-m_n(\omega)}^{k_n(\omega)-1} U_\ell(\omega)}{k_n(\omega) + m_n(\omega)} - \mathbb{E}(U_0) \right| \geq \epsilon(\omega).
\]

Consider the following decomposition:

\[
\frac{\sum_{\ell=-m_n(\omega)}^{k_n(\omega)-1} U_\ell(\omega)}{k_n(\omega) + m_n(\omega)} = \frac{m_n(\omega)}{k_n(\omega) + m_n(\omega)} \cdot \frac{\sum_{\ell=-m_n(\omega)}^{-1} U_\ell(\omega)}{m_n(\omega)} + \frac{k_n(\omega)}{k_n(\omega) + m_n(\omega)} \cdot \sum_{\ell=0}^{k_n(\omega)-1} U_\ell(\omega).
\]

First, assume that \( (k_n(\omega))_{n \in \mathbb{N}} \) is bounded. Since \( k_n(\omega) + m_n(\omega) \geq n \), it follows that \( m_n(\omega) \) tends to infinity, implying that

\[
\lim_{n \to \infty} \frac{m_n(\omega)}{k_n(\omega) + m_n(\omega)} = 1,
\]

\[
\lim_{n \to \infty} \frac{k_n(\omega)}{k_n(\omega) + m_n(\omega)} = 0,
\]

whereas \( \sum_{\ell=0}^{k_n(\omega)-1} U_\ell(\omega)/k_n(\omega) \) remains bounded. However, since \( \omega \in \Omega_2 \) and \( \lim_{n \to \infty} m_n(\omega) = \infty \),

\[
\lim_{n \to \infty} \frac{\sum_{\ell=-m_n(\omega)}^{-1} U_\ell(\omega)}{m_n(\omega)} = \mathbb{E}(U_0),
\]

which implies, together with (72), that

\[
\lim_{n \to \infty} \frac{\sum_{\ell=-m_n(\omega)}^{k_n(\omega)-1} U_\ell(\omega)}{k_n(\omega) + m_n(\omega)} = \mathbb{E}(U_0).
\]

This contradicts (70). Using similar arguments one proves that \( (m_n(\omega))_{n \in \mathbb{N}} \) is unbounded as well. Hence, we have proved that neither \( (k_n(\omega))_{n \in \mathbb{N}} \) nor \( (m_n(\omega))_{n \in \mathbb{N}} \) are bounded.

Then, by extracting a subsequence if necessary, one may assume that \( \lim_{n \to \infty} k_n(\omega) = \lim_{n \to \infty} m_n(\omega) = \infty \). Since \( \omega \in \Omega_2 \), this implies that

\[
\lim_{n \to \infty} \frac{\sum_{\ell=-m_n(\omega)}^{-1} U_\ell(\omega)}{m_n(\omega)} = \lim_{n \to \infty} \frac{\sum_{\ell=0}^{k_n(\omega)-1} U_\ell(\omega)}{k_n(\omega)} = \mathbb{E}(U_0).
\]
Combining this with (71), we obtain that
\[ \lim_{n \to \infty} \sum_{k=n}^{\infty} \frac{U_\ell(\omega)}{k_n(\omega) + m_n(\omega)} = \mathbb{E}(U_0), \]
which again contradicts (70). Finally, \( \Omega_1^c \cap \Omega_2 = \emptyset \), and since \( \mathbb{P}(\Omega_2) = 1 \), we finally obtain that \( \mathbb{P}(\Omega_1) = 1 \). The proof is complete. □

**Lemma 19.** Let \( (U_k)_{k \in \mathbb{Z}} \), \( (V_k)_{k \in \mathbb{Z}} \) and \( (W_k)_{k \in \mathbb{Z}} \) be stationary sequences such that
\[ \mathbb{E}(\ln^+ U_0) < \infty, \quad \mathbb{E}(\ln^+ V_0) < \infty, \quad \mathbb{E}(\ln^+ W_0) < \infty. \]
Then for all \( \eta \) and \( \rho \) in \( [0,1] \) such that \( -\ln \eta > \mathbb{E}(\ln^+ V_0) \) there exist a \( \mathbb{P} \)-a.s. finite random variable \( C \) and a constant \( \beta \in [0,1] \) such that for all \( k \in \mathbb{N}^* \) and \( m \in \mathbb{N} \), \( \mathbb{P} \)-a.s.,
\[ \rho^{k+m} + \eta^{k+m}W_m \left( \prod_{\ell=-m}^{k-1} V_\ell \right) U_k \leq C \beta^{k+m}. \]

**Proof.** See [15], Lemma 6. □

**Acknowledgment.** We thank the anonymous referee for insightful comments that improved the presentation of the paper.

**REFERENCES**

[1] Bain, A. and Crisan, D. (2009). *Fundamentals of Stochastic Filtering*. Stochastic Modelling and Applied Probability 60. Springer, New York. MR2454694

[2] Cappé, O., Moulines, E. and Rydén, T. (2005). *Inference in Hidden Markov Models*. Springer, New York. MR2159833

[3] Chopin, N. (2002). A sequential particle filter method for static models. *Biometrika* 89 539–551. MR1929161

[4] Crisan, D. and Heine, K. (2008). Stability of the discrete time filter in terms of the tails of noise distributions. *J. Lond. Math. Soc. (2)* 78 441–458. MR2439634

[5] Del Moral, P. (2004). *Feynman–Kac Formulae. Genealogical and Interacting Particle Systems with Applications*. Springer, New York. MR2044973

[6] Del Moral, P. and Guionnet, A. (1999). Central limit theorem for nonlinear filtering and interacting particle systems. *Ann. Appl. Probab.* 9 275–297. MR1687359

[7] Del Moral, P. and Guionnet, A. (2001). On the stability of interacting processes with applications to filtering and genetic algorithms. *Ann. Inst. Henri Poincaré Probab. Stat.* 37 155–194. MR1819122

[8] Del Moral, P., Jacod, J. and Protter, P. (2001). The Monte-Carlo method for filtering with discrete-time observations. *Probab. Theory Related Fields* 120 346–368. MR1843179

[9] Del Moral, P. and Ledoux, M. (2000). Convergence of empirical processes for interacting particle systems with applications to nonlinear filtering. *J. Theoret. Probab.* 13 225–257. MR1744985
[10] Del Moral, P. and Miclo, L. (2000). Branching and interacting particle systems approximations of Feynman–Kac formulae with applications to nonlinear filtering. In Séminaire de Probabilités, XXXIV. Lecture Notes in Math. 1729 1–145. Springer, Berlin. MR1768060

[11] Douc, R., Fort, G., Moulines, E. and Priouret, P. (2009). Forgetting the initial distribution for hidden Markov models. Stochastic Process. Appl. 119 1235–1256. MR2508572

[12] Douc, R., Garivier, A., Moulines, E. and Olsson, J. (2011). Sequential Monte Carlo smoothing for general state space hidden Markov models. Ann. Appl. Probab. 21 2109–2145. MR2895411

[13] Douc, R., Guillin, A. and Najim, J. (2005). Moderate deviations for particle filtering. Ann. Appl. Probab. 15 587–614. MR2114983

[14] Douc, R. and Moulines, E. (2008). Limit theorems for weighted samples with applications to sequential Monte Carlo methods. Ann. Statist. 36 2344–2376. MR2458190

[15] Douc, R. and Moulines, E. (2012). Asymptotic properties of the maximum likelihood estimation in misspecified hidden Markov models. Ann. Statist. 40 2697–2732. MR3097617

[16] Douc, R., Moulines, É. and Olsson, J. (2009). Optimality of the auxiliary particle filter. Probab. Math. Statist. 29 1–28. MR2552996

[17] Doucet A., de Freitas N. and Gordon N., eds. (2001). Sequential Monte Carlo Methods in Practice. Springer, New York. MR1847783

[18] Godsill, S. J., Doucet, A. and West, M. (2004). Monte Carlo smoothing for nonlinear time series. J. Amer. Statist. Assoc. 50 438–449.

[19] Gordon, N., Salmond, D. and Smith, A. F. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE Proc., F, Radar Signal Process. 140 107–113.

[20] Heine, K. and Crisan, D. (2008). Uniform approximations of discrete-time filters. Adv. in Appl. Probab. 40 979–1001. MR2488529

[21] Johansen, A. M. and Doucet, A. (2008). A note on auxiliary particle filters. Statist. Probab. Lett. 78 1498–1504. MR2528342

[22] Kingman, J. F. C. (1973). Subadditive ergodic theory. Ann. Probab. 1 883–909. MR0356192

[23] Kleptsyna, M. L. and Veretennikov, A. Yu. (2008). On discrete time ergodic filters with wrong initial data. Probab. Theory Related Fields 141 411–444. MR2391160

[24] Künsch, H. R. (2005). Recursive Monte Carlo filters: Algorithms and theoretical analysis. Ann. Statist. 33 1983–2021. MR2211077

[25] LeGland, F. and Oudian, N. (2003). A robustification approach to stability and to uniform particle approximation of nonlinear filters: The example of pseudo-mixing signals. Stochastic Process. Appl. 106 279–316. MR1986030

[26] Le Gland, F. and Oudian, N. (2004). Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters. Ann. Appl. Probab. 14 144–187. MR2023019

[27] Olsson, J. and Rydén, T. (2008). Asymptotic properties of particle filter-based maximum likelihood estimators for state space models. Stochastic Process. Appl. 118 649–680. MR2394847

[28] Oudian, N. and Rubenthaler, S. (2005). Stability and uniform particle approximation of nonlinear filters in case of nonergodic signals. Stoch. Anal. Appl. 23 421–448. MR2140972
LONG-TERM STABILITY OF SEQUENTIAL MONTE CARLO METHODS

[29] Pitt, M. K. and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. *J. Amer. Statist. Assoc.* **94** 590–599. MR1702328

[30] Pólya, G. and Szegő, G. (1976). *Problems and Theorems in Analysis*. Springer, New York. MR0396134

[31] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York. MR0595165

[32] Shiryayev, A. N. (1996). *Probability*, 2nd ed. Springer, New York. MR1368405

[33] Tadić, V. B. and Doucet, A. (2005). Exponential forgetting and geometric ergodicity for optimal filtering in general state–space models. *Stochastic Process. Appl.* **115** 1408–1436. MR2152381

[34] van Handel, R. (2009). The stability of conditional Markov processes and Markov chains in random environments. *Ann. Probab.* **37** 1876–1925. MR2561436

R. Douc  
SAMOVAR, CNRS UMR 5157  
INSTITUT TÉLÉCOM/TÉLÉCOM SudParis  
9 rue Charles Fourier  
91000 Evry  
France  
E-mail: randal.douc@it-sudparis.eu

E. Moulines  
LTCI, CNRS UMR 8151  
INSTITUT TÉLÉCOM/TÉLÉCOM ParisTech  
46, rue Barrault  
75634 Paris Cedex 13  
France  
E-mail: eric.moulines@telecom-paristech.fr

J. Olsson  
DEPARTMENT OF MATHEMATICS  
KTH Royal Institute of Technology  
SE-100 44, Stockholm  
Sweden  
E-mail: jimmyol@kth.se