Long-term stability of sequential Monte Carlo methods under verifiable conditions

Randal Douc
Eric Moulines
Jimmy Olsson

Preprints in Mathematical Sciences
2012:1

Lund University

Faculty of Science
Centre for Mathematical Sciences
Mathematical Statistics

Centrum Scientiarum Mathematicarum

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

By Randal Douc *,§ Eric Moulines †,§ and Jimmy Olsson ¶

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 non-compact state space. In addition, we derive a similar time uniform bound on the asymptotic L^p error. Importantly, our results hold for misspecified models, i.e. 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 $(\mathsf{X}, \mathcal{X})$, is only partially observed through the sequence $(Y_n)_{n\in\mathbb{N}}$ of *observations* taking values in another general state space $(\mathsf{Y}, \mathcal{Y})$. 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 e.g. [2] and the references therein. We denote by $\mathbf{Q}$ and χ the kernel and initial distribution of

*SAMOVAR, CNRS UMR 5157, Institut Télécom/Télécom SudParis, 9 rue Charles Fourier, 91000 Evry.

†LTCI, CNRS UMR 8151, Institut Télécom/Télécom ParisTech, 46, rue Barrault, 75634 Paris Cedex 13, France.

¶Centre for Mathematical Sciences, Lund University, Box 118, 22100 Lund, Sweden.

§This work is supported by the Agence Nationale de la Recherche through the 2009–2012 project Big MC.

¶This work is supported by the Kungliga Fysiografiska Sällskapet i Lund.

$(X_n)_{n\in\mathbb{N}}$, 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 so-called *filter distributions*, where the filter distribution at time n is defined as the conditional distribution of X_n given the corresponding observation history $Y_0^n = (Y_0, \ldots, Y_n)$ (this will be our generic notation for vectors), and 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*. Alternatively, one may focus on the *predictor distributions*, where the predictor distribution at time n is defined as the conditional distribution of the state X_n given the preceding observation history Y_0^{n-1}, and the predictor distributions are in general obtained as a by-product when computing the filter distributions and vice versa. In this paper we focus on computation of the predictor distributions, which we denote by $\phi_\chi\langle Y_0^{n-1}\rangle$, $n \in \mathbb{N}^*$ (a more precise definition of these measures is given in Section 2). The *filter recursion* defines a measure-valued mapping Φ generating recursively the predictor distribution flow according to $\phi_\chi\langle Y_0^n\rangle = \Phi\langle Y_n\rangle(\phi_\chi\langle Y_0^{n-1}\rangle)$ (we refer again to Section 2 for more 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 in general 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 $\phi_\chi^N\langle Y_0^{n-1}\rangle$ associated with a finite sample $(\xi_n^i)_{i=1}^N$ of *particles* evolving randomly and recursively in time. Particle filters comprise generally two main operations: a *mutation step* and a *selection step*. The mutation step randomly disseminates the particles in the state space while the selection step duplicates or eliminates particles with high or low posterior probability, respectively. The most basic algorithm—proposed in [18] and referred to as the bootstrap particle filter—mutates the particles according to the dynamics of the latent Markov chain and selects the same with probabilities proportional to the local likelihood of the mutated particles. Thus, subjecting a particle sample $(\xi_n^i)_{i=1}^N$ to selection and mutation is in the case of the bootstrap particle filter equivalent to drawing, conditionally independently given $(\xi_n^i)_{i=1}^N$, new particles $(\xi_{n+1}^i)_{i=1}^N$ from the distribution $\Phi\langle Y_n\rangle(\phi_\chi^N\langle Y_0^{n-1}\rangle)$ obtained by plugging the empirical measure $\phi_\chi^N\langle Y_0^{n-1}\rangle$

into the filter recursion, which we denote

$$(\xi_{n+1}^i)_{i=1}^N \sim_{\text{i.i.d.}} \Phi\langle Y_n\rangle(\phi_\chi^N\langle Y_0^{n-1}\rangle)^{\otimes N}. \tag{1}$$

Since the seminal paper [18], particle filters have been successfully applied to nonlinear filtering problems in many different fields; we refer to the collection [15] 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, e.g., L^p error bounds and weak convergence—see the monographs [5, 1] 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*, i.e. when n tends to infinity, convergence is less obvious. Indeed, each recursive update (1) of the particles $(\xi_n^i)_{i=1}^N$ is based on the implicit assumption that the empirical measure $\phi_\chi^N\langle Y_0^{n-1}\rangle$ associated with the ancestor sample approximates perfectly well the predictor $\phi_\chi\langle Y_0^{n-1}\rangle$ 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. [30, 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.* To our knowledge, the first time uniform convergence result for bootstrap-type particle filters was obtained by [7] (see also the book [5] for refinements) using a technique based on the *uniform forgetting* of the predictor distribution. We recall in some detail this technique. By writing

$$\phi_\chi^N\langle Y_0^n\rangle - \phi_\chi\langle Y_0^n\rangle = \underbrace{\phi_\chi^N\langle Y_0^n\rangle - \Phi\langle Y_n\rangle(\phi_\chi^N\langle Y_0^{n-1}\rangle)}_{\text{sampling error}} + \underbrace{\Phi\langle Y_n\rangle(\phi_\chi^N\langle Y_0^{n-1}\rangle) - \Phi\langle Y_n\rangle(\phi_\chi\langle Y_0^{n-1}\rangle)}_{\text{initialization error}}$$

one may decompose the error $\phi_\chi^N\langle Y_0^n\rangle - \phi_\chi\langle Y_0^n\rangle$ into a first error (the sampling error) introduced by replacing $\Phi\langle Y_n\rangle(\phi_\chi^N\langle Y_0^{n-1}\rangle)$ by its empirical estimate $\phi_\chi^N\langle Y_0^n\rangle$ and a second error (the initialization error) originating from the discrepancy between empirical measure $\phi_\chi^N\langle Y_0^{n-1}\rangle$ associated with the ancestor particles and the true predictor $\phi_\chi\langle Y_0^{n-1}\rangle$. 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 $\Phi\langle Y_n\rangle$ is in some sense contracting and thus downscales the discrepancy between $\phi_\chi^N\langle Y_0^{n-1}\rangle$ and $\phi_\chi\langle Y_0^{n-1}\rangle$. 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 $\|\Phi\langle Y_m^n\rangle(\mu) - \Phi\langle Y_m^n\rangle(\nu)\| \le \rho^{n-m+1}\|\mu - \nu\|$ for any integers $0 \le m \le n$ and any probability measures μ and ν, where $\|\cdot\|$ is some suitable norm on the space of probability measures and $\Phi\langle Y_m^n\rangle \triangleq \Phi\langle Y_n\rangle \circ \Phi\langle Y_{n-1}\rangle \circ \cdots \circ \Phi\langle Y_m\rangle$. Since $\Phi\langle Y_m^n\rangle(\mu)$ is the predictor distribution $\phi_\mu\langle Y_m^n\rangle$ obtained when the hidden chain is initialized with the distribution μ at time m, this means that the predictor distribution forgets the initial distribution geometrically fast. In addition, the forgetting rate ρ 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

$$
\begin{aligned}
(2)\quad \phi_\chi^N\langle Y_0^n\rangle - \phi_\chi\langle Y_0^n\rangle &= \phi_\chi^N\langle Y_0^n\rangle - \Phi\langle Y_n\rangle(\phi_\chi^N\langle Y_0^{n-1}\rangle) \\
&+ \sum_{k=1}^{n-1}\left(\Phi\langle Y_{k+1}^n\rangle(\phi_\chi^N\langle Y_0^k\rangle) - \Phi\langle Y_{k+1}^n\rangle \circ \Phi\langle Y_k\rangle(\phi_\chi^N\langle Y_0^{k-1}\rangle)\right) \\
&\qquad + \Phi\langle Y_1^n\rangle(\phi_\chi^N\langle Y_0\rangle) - \Phi\langle Y_1^n\rangle(\phi_\chi\langle Y_0\rangle).
\end{aligned}
$$

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

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

$$\epsilon^- \nu(\mathsf{A}) \leq \mathbf{Q}(x, \mathsf{A}) \leq \epsilon^+ \nu(\mathsf{A}). \tag{3}$$

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, the condition (3) typically requires the state space to be compact). Some refinements have been obtained in e.g. [23, 22, 5, 25, 29, 2, 24, 14, 4, 19].

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 [6]. More specifically, it was shown that the normalized Monte Carlo error $\sqrt{N}(\phi_\chi^N \langle Y_0^{n-1} \rangle h - \phi_\chi \langle Y_0^{n-1} \rangle h)$ tends weakly, for a fixed $n \in \mathbb{N}^*$ and as the particle population size N tends to infinity, to a zero mean normal-distributed variable with variance $\sigma_\chi^2 \langle Y_0^{n-1} \rangle (h)$. Here we have used the notation $\mu h \triangleq \int h(x) \, \mu(\mathrm{d}x)$ to denote expectations. The original proof of the CLT was later simplified and extended to more general particle filtering algorithms in [21, 3, 12, 14, 16]; in Section 2 we recall in detail the version obtained in [12] and provide an explicit expression of the asymptotic variance $\sigma_\chi^2 \langle Y_0^{n-1} \rangle (h)$. 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 $\sigma_\chi^2 \langle Y_0^{n-1} \rangle (h)$ by similar forgetting-based arguments. Here a key ingredient is that the particles $(\xi_n^i)_{i=1}^N$ obtained at the different time steps become, asymptotically as N tends to infinity, statistically independent. Consequently, the total asymptotic variance of $\sqrt{N}(\phi_\chi^N \langle Y_0^{n-1} \rangle h - \phi_\chi \langle Y_0^{n-1} \rangle h)$ is obtained by simply summing up the the asymptotic variances of the error terms $\sqrt{N}(\Phi \langle Y_{k+1}^n \rangle (\phi_\chi^N \langle Y_0^k \rangle) h - \Phi \langle Y_{k+1}^n \rangle \circ \Phi \langle Y_k \rangle (\phi_\chi^N \langle Y_0^{k-1} \rangle) h)$ in the decomposition (2). Finally, applying again the contraction of the composed mapping $\Phi \langle Y_m^n \rangle$ yields a uniform bound on the total asymptotic variance in accordance with the calculation above. In [10], a similar stability result was obtained for a particle-based version of the *forward-filtering backward-simulation algorithm* (proposed in [17]); nevertheless, also the analysis of this work relies completely on the assumption of strong mixing of the latent Markov chain, which, as already pointed out, does not hold for most models used in practice.

A first breakthrough towards stability results for non-compact state spaces was made in [30]. This work establishes, again for bootstrap-type particle filters, a uniform time average convergence result of form

$$\lim_{N\to\infty} \sup_{n\in\mathbb{N}} \mathbb{E}\left(n^{-1}\sum_{k=1}^{n} \|\bar{\phi}^N_\chi\langle Y_0^k\rangle - \bar{\phi}_\chi\langle Y_0^k\rangle\|_{\mathrm{BL}}\right) = 0, \tag{4}$$

where $\|\cdot\|_{\mathrm{BL}}$ denotes the dual bounded-Lipschitz norm and $\bar{\phi}_\chi\langle Y_0^k\rangle$ denotes the filter distribution at time k. 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 non-compact 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 of course never true in practice. In addition, the convergence result (4) does not, on the contrary 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 [16], restricting the analysis to bootstrap-type particle filters does not imply a significant loss of generality, as the CLT for more general *auxiliary particle filters* [26] 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 non-compact 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 $\mathsf{I} = [-a, a] \subset \mathbb{R}$ such that $\mu_n(\mathsf{I}^c) \le \epsilon$ 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 $(\sigma^2_\chi\langle Y_0^{n-1}\rangle(h))_{n\in\mathbb{N}^*}$ of asymptotic variances is tight for *any stationary 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 [13] (and going back to [20] and [9]) with the purpose of establishing the convergence of the *relative entropy* for misspecified HMMs. In our analysis, the strong mixing assumption (3) is replaced by the considerably weaker r-*local Doeblin condition* (14). This assumption is, for instance, trivially satisfied (for $r = 1$) if there exist a measurable set $\mathsf{C} \subseteq \mathsf{X}$, a probability measure λ_{C} on $(\mathsf{X}, \mathcal{X})$ such that $\lambda_{\mathsf{C}}(\mathsf{C}) = 1$, and positive constants $0 < \epsilon_{\mathsf{C}}^- < \epsilon_{\mathsf{C}}^+$ such that for all $x \in \mathsf{X}$ and all $\mathsf{A} \in \mathcal{X}$,

$$\epsilon_{\mathsf{C}}^- \lambda_{\mathsf{C}}(\mathsf{A}) \leq \mathbf{Q}(x, \mathsf{A} \cap \mathsf{C}) \leq \epsilon_{\mathsf{C}}^+ \lambda_{\mathsf{C}}(\mathsf{A}), \tag{5}$$

a condition that is easily verified for many HMMs with non-compact state space (we emphasize however that the assumption (14) is even weaker than (5)).

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

- we present time uniform bounds that also provide the rate of convergence in N of the particle filter for very general HMMs (with possibly non-compact state space).
- we establish long-term stability of the particle filter also in the case of misspecification, i.e. when the stationary law of the observations entering the particle filter differs from that of the HMM governing the dynamics of the particles $(\xi_n^i)_{i=1}^N$.

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 Appendix A provide the full details of our proofs.

2. Preliminaries.

2.1. *Notation.* We preface the introduction of HMMs by some notation. Let $(\mathsf{X}, \mathcal{X})$ be a measurable space, where $\mathcal{X}$ is a countably generated σ-field. Denote by $\mathcal{F}(\mathsf{X})$ (resp. $\mathcal{F}_+(\mathsf{X})$) the set of bounded (resp. bounded and positive) $\mathcal{X}/\mathcal{B}(\mathbb{R})$-measurable functions on X and by $\mathcal{P}(\mathsf{X}, \mathcal{X})$ the set of probability measures on $(\mathsf{X}, \mathcal{X})$. Let $\mathbf{K} : \mathsf{X} \times \mathcal{X} \to \mathbb{R}_+$ be a finite kernel on X, i.e. for each $x \in \mathsf{X}$, the mapping $\mathbf{K}(x, \cdot) : \mathsf{A} \mapsto \mathbf{K}(x, \mathsf{A})$ is a finite measure on $\mathcal{X}$ and for each $\mathsf{A} \in \mathcal{X}$, the function $\mathbf{K}(x, \cdot) : x \mapsto \mathbf{K}(x, \mathsf{A})$ is $\mathcal{X}/\mathcal{B}([0, 1])$-measurable. If $\mathbf{K}(x, \cdot)$ is a probability measure on $(\mathsf{X}, \mathcal{X})$ for all $x \in \mathsf{X}$, then the kernel $\mathbf{K}$ is said to be Markov. A kernel induces two integral operators, the first acting on the space $\mathcal{M}(\mathsf{X}, \mathcal{X})$ of σ-finite measures on $(\mathsf{X}, \mathcal{X})$ and the other on $\mathcal{F}(\mathsf{X})$. More specifically, for $\mu \in \mathcal{M}(\mathsf{X}, \mathcal{X})$ and $f \in \mathcal{F}(\mathsf{X})$ we define the measure

$$\mu\mathbf{K} : \mathcal{X} \ni \mathsf{A} \mapsto \int \mathbf{K}(x, \mathsf{A})\, \mu(\mathrm{d}x)$$

and the function

$$\mathbf{K}f : \mathsf{X} \ni x \mapsto \int f(x')\, \mathbf{K}(x, \mathrm{d}x').$$

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

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

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

$$\mathbf{T}((x, y), \mathsf{A}) \triangleq \iint \mathbb{1}_{\mathsf{A}}(x', y')\, \mathbf{Q}(x, \mathrm{d}x')\, \mathbf{G}(x', \mathrm{d}y'), \qquad (x, y) \in \mathsf{X} \times \mathsf{Y}, \quad \mathsf{A} \in \mathcal{X} \otimes \mathcal{Y},$$

on the product space $(\mathsf{X} \times \mathsf{Y}, \mathcal{X} \otimes \mathcal{Y})$. Let $(X_n, Y_n)_{n \in \mathbb{N}}$ be the canonical Markov chain induced by $\mathbf{T}$ and the initial distribution $\mathcal{X} \otimes \mathcal{Y} \ni \mathsf{A} \mapsto \int \mathbb{1}_{\mathsf{A}}(x, y)\, \chi(\mathrm{d}x)\, \mathbf{G}(x, \mathrm{d}y)$. The bivariate process $(X_n, Y_n)_{n \in \mathbb{N}}$ is what we refer to as the HMM. We shall denote by $\mathbb{P}_\chi$ and $\mathbb{E}_\chi$ the probability measure and corresponding expectation associated with the HMM on the canonical space $((\mathsf{X} \times \mathsf{Y})^{\mathbb{N}}, (\mathcal{X} \otimes \mathcal{Y})^{\otimes \mathbb{N}})$. We assume that the observation kernel $\mathbf{G}$ is non-degenerated in the sense that there exists a σ-finite measure ν on $(\mathsf{Y}, \mathcal{Y})$ and a measurable function $g : \mathsf{X} \times \mathsf{Y} \to]0, \infty[$ such that

$$\mathbf{G}(x, \mathsf{A}) = \int \mathbb{1}_{\mathsf{A}}(y)\, g(x, y)\, \nu(\mathrm{d}y), \quad x \in \mathsf{X}, \quad \mathsf{A} \in \mathcal{Y}.$$

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

$$\mathbf{L}\langle y_k^m \rangle(x_k, \mathsf{A}) \triangleq \int \cdots \int \mathbb{1}_{\mathsf{A}}(x_{m+1}) \prod_{\ell=k}^{m} g(x_\ell, y_\ell)\, \mathbf{Q}(x_\ell, \mathrm{d}x_{\ell+1}), \quad x_k \in \mathsf{X}, \quad \mathsf{A} \in \mathcal{X}, \tag{6}$$

with the convention

$$\mathbf{L}\langle y_k^m \rangle(x, \mathsf{A}) \triangleq \delta_x(\mathsf{A}) \text{ for } k > m \tag{7}$$

(where δ_x denotes the Dirac mass at point x). Note that the function $y_0^{n-1} \mapsto \chi \mathbf{L}\langle y_0^{n-1} \rangle \mathbb{1}_{\mathsf{X}}$ 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 $\ell \in \{k, \ldots, m-1\}$,

$$\mathbf{L}\langle y_k^m \rangle = \mathbf{L}\langle y_k^\ell \rangle \mathbf{L}\langle y_{\ell+1}^m \rangle. \tag{8}$$

Let $\phi_\chi \langle y_k^m \rangle$ be the probability measure defined by

$$\phi_\chi \langle y_k^m \rangle(\mathsf{A}) \triangleq \frac{\chi \mathbf{L}\langle y_k^m \rangle \mathbb{1}_{\mathsf{A}}}{\chi \mathbf{L}\langle y_k^m \rangle \mathbb{1}_{\mathsf{X}}}, \quad \mathsf{A} \in \mathcal{X}. \tag{9}$$

Note that this implies that $\phi_\chi \langle y_k^m \rangle = \chi$ when $k > m$. Using the notation, it can be shown (see e.g. [2, Proposition 3.1.4]) that for any $h \in \mathcal{F}(\mathsf{X})$,

$$\mathbb{E}_\chi \left(h(X_n) \,\middle|\, Y_0^{n-1} \right) = \int h(x)\, \phi_\chi \langle Y_0^{n-1} \rangle(\mathrm{d}x),$$

i.e. $\phi_\chi \langle Y_0^{n-1} \rangle$ is the predictor of X_n given the observations Y_0^{n-1}. From the definition (9) one immediately obtains the recursion

$$\phi_\chi \langle y_0^n \rangle(\mathsf{A}) = \frac{\phi_\chi \langle y_0^{n-1} \rangle \mathbf{L}\langle y_n \rangle \mathbb{1}_{\mathsf{A}}}{\phi_\chi \langle y_0^{n-1} \rangle \mathbf{L}\langle y_n \rangle \mathbb{1}_{\mathsf{X}}} = \frac{\int g(x, y_n) \mathbf{Q}(x, \mathsf{A})\, \phi_\chi \langle y_0^{n-1} \rangle(\mathrm{d}x)}{\int g(x, y_n)\, \phi_\chi \langle y_0^{n-1} \rangle(\mathrm{d}x)}, \quad \mathsf{A} \in \mathcal{X},$$

which can be expressed in condensed form as

$$\phi_\chi \langle y_0^n \rangle = \Phi\langle y_n \rangle(\phi_\chi \langle y_0^{n-1} \rangle), \tag{10}$$

where $\Phi\langle y_n \rangle$ transforms a probability measure $\mu \in \mathcal{P}(\mathsf{X}, \mathcal{X})$ into the measure

$$\Phi\langle y_n \rangle(\mu) : \mathcal{X} \ni \mathsf{A} \mapsto \frac{\int g(x, y_n) \mathbf{Q}(x, \mathsf{A})\, \mu(\mathrm{d}x)}{\int g(x, y_n)\, \mu(\mathrm{d}x)}.$$

As mentioned in the introduction, the recursion (10) cannot in general be solved in closed form. In the following section we discuss how approximate solutions to (10) can be obtained using particle filters, with focus set on the bootstrap particle filter proposed in [18].

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

$$\phi_\chi^N\langle y_0^{n-1}\rangle \triangleq \frac{1}{N}\sum_{i=1}^N \delta_{\xi_n^i}$$

associated with these draws *targets* the predictor $\phi_\chi\langle y_0^{n-1}\rangle$ in the sense that $\phi_\chi^N\langle y_0^{n-1}\rangle h = \sum_{i=1}^N h(\xi_n^i)/N$ estimates $\phi_\chi\langle y_0^{n-1}\rangle h$ for any $h \in \mathcal{F}(\mathsf{X})$. In order to form a new particle sample $(\xi_{n+1}^i)_{i=1}^N$ approximating the predictor $\phi_\chi\langle y_0^n\rangle$ at the subsequent time step, we replace, in (10), the true predictor $\phi_\chi\langle y_0^{n-1}\rangle$ by the particle estimate $\phi_\chi^N\langle y_0^{n-1}\rangle$. This yields the approximation

$$\phi_\chi\langle y_0^n\rangle(\mathsf{A}) \approx \sum_{i=1}^N \frac{g(\xi_n^i, y_n)}{\sum_{\ell=1}^N g(\xi_n^\ell, y_n)}\mathbf{Q}(\xi_n^i, \mathsf{A}), \quad \mathsf{A} \in \mathcal{X}. \tag{11}$$

Next, the sample $(\xi_{n+1}^i)_{i=1}^N$ is generated by simulating N conditionally independent draws from the mixture in (11) using the following algorithm.

set $\Omega_n^N \leftarrow 0$
for $i = 1 \to N$ **do**
  set $\omega_n^i \leftarrow g(\xi_n^i, y_n)$
  set $\Omega_n^N \leftarrow \Omega_n^N + \omega_n^i$
end for
for $i = 1 \to N$ **do**
  draw $I_n^i \sim (\omega_n^\ell/\Omega_n^N)_{\ell=1}^N$
  draw $\xi_{n+1}^i \sim \mathbf{Q}(\xi_n^{I_n^i}, \cdot)$
end for

In the scheme above, the operation $\sim$ means implicitly that all draws (for different i's) are conditionally independent. Moreover, the operation $I_n^i \sim (\omega_n^\ell/\Omega_n^N)_{\ell=1}^N$ means that each index I_n^i is simulated according to the discrete probability distribution generated by the normalized importance weights

$(\omega_n^\ell/\Omega_n^N)_{\ell=1}^N$. The algorithm is typically initialized by drawing N i.i.d. particles $(\xi_0^i)_{i=1}^N$ from the initial distribution χ and letting $\sum_{i=1}^N \delta_{\xi_0^i}/N$ be an estimate of χ.

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, [6] established the following CLT. Define for $h \in \mathcal{F}(\mathsf{X})$,

(12)
$$\sigma_\chi^2\langle y_0^{n-1}\rangle(h) \triangleq \sum_{k=0}^n \phi_\chi\langle y_0^{k-1}\rangle \left(\frac{\mathbf{L}\langle y_k^{n-1}\rangle h - \phi_\chi\langle y_0^{n-1}\rangle h \times \mathbf{L}\langle y_k^{n-1}\rangle \mathbb{1}_\mathsf{X}}{\phi_\chi\langle y_0^{k-1}\rangle \mathbf{L}\langle y_k^{n-1}\rangle \mathbb{1}_\mathsf{X}} \right)^2 .$$

THEOREM 1 ([6]). *For all $h \in \mathcal{F}(\mathsf{X})$ and $y_0^{n-1} \in \mathsf{Y}^n$ it holds, as $N \to \infty$,*

$$\sqrt{N}(\phi_\chi^N\langle y_0^{n-1}\rangle h - \phi_\chi\langle y_0^{n-1}\rangle h) \xrightarrow{\mathcal{D}} \sigma_\chi\langle y_0^{n-1}\rangle(h) Z, \tag{13}$$

where $\sigma_\chi\langle y_0^{n-1}\rangle(h)$ is defined in (12) 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 sequence $(\sigma_\chi^2\langle Y_0^{n-1}\rangle(h))_{n\in\mathbb{N}^*}$ of asymptotic variances is an $(\mathcal{F}_n^Y)_{n\in\mathbb{N}}$-adapted stochastic process, where $(\mathcal{F}_n^Y)_{n\in\mathbb{N}}$ is the natural filtration of the observation process. The aim of the next section is to establish that this sequence is tight. *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 2. *Let $r \in \mathbb{N}^*$. A set $\mathsf{C} \in \mathcal{X}$ is r-*local Doeblin* with respect to $\{\mathbf{Q}, g\}$ if there exist positive functions $\epsilon_\mathsf{C}^- : \mathsf{Y}^r \to \mathbb{R}^+$ and $\epsilon_\mathsf{C}^+ : \mathsf{Y}^r \to \mathbb{R}^+$, a family $\{\mu_\mathsf{C}\langle z\rangle; z \in \mathsf{Y}^r\}$ of probability measures, and a family $\{\varphi_\mathsf{C}\langle z\rangle; z \in \mathsf{Y}^r\}$ of positive functions such that for all $z \in \mathsf{Y}^r$, $\mu_\mathsf{C}\langle z\rangle(\mathsf{C}) = 1$ and for all $\mathsf{A} \in \mathcal{X}$ and $x \in \mathsf{C}$,*

$$\epsilon_\mathsf{C}^-\langle z\rangle\ \varphi_\mathsf{C}\langle z\rangle(x)\ \mu_\mathsf{C}\langle z\rangle(\mathsf{A}) \le \mathbf{L}\langle z\rangle(x, \mathsf{A}\cap\mathsf{C}) \le \epsilon_\mathsf{C}^+\langle z\rangle\ \varphi_\mathsf{C}\langle z\rangle(x)\ \mu_\mathsf{C}\langle z\rangle(\mathsf{A}). \tag{14}$$

(A1) The process $(Y_n)_{n\in\mathbb{Z}}$ is strictly stationary. Moreover, there exist an integer $r \in \mathbb{N}^*$ and a set $\mathsf{K} \in \mathcal{Y}^{\otimes r}$ 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 $\mathbb{P}\left(Z_0 \in \mathsf{K}\right) > 2/3$.

(ii) For all $\eta > 0$ there exists an r-local Doeblin set $\mathsf{C} \in \mathcal{X}$ such that for all $y_0^{r-1} \in \mathsf{K}$,

$$\sup_{x\in\mathsf{C}^c} \mathbf{L}\langle y_0^{r-1}\rangle(x,\mathsf{X}) \le \eta \sup_{x\in\mathsf{X}} \mathbf{L}\langle y_0^{r-1}\rangle(x,\mathsf{X}) < \infty \tag{15}$$

and

$$\inf_{y_0^{r-1}\in\mathsf{K}} \frac{\epsilon_{\mathsf{C}}^-\langle y_0^{r-1}\rangle}{\epsilon_{\mathsf{C}}^+\langle y_0^{r-1}\rangle} > 0, \tag{16}$$

where the functions ϵ_{C}^+ and ϵ_{C}^- are given in Definition 2.

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

$$\mathbb{E}\left(\ln^- \inf_{x\in\mathsf{D}} \delta_x \mathbf{L}\langle Y_0^{r-1}\rangle \mathbb{1}_{\mathsf{D}}\right) < \infty. \tag{17}$$

(A2) (i) $g(x,y) > 0$ for all $(x,y) \in \mathsf{X}\times\mathsf{Y}$.

(ii) $\mathbb{E}\left(\ln^+ \sup_{x\in\mathsf{X}} g(x,Y_0)\right) < \infty$.

Remark 3. *In the case $r = 1$ we may replace (**A1**) by the simpler assumption that there exists a set $\mathsf{K} \in \mathcal{Y}$ such that the following holds.*

(i) $\mathbb{P}\left(Y_0 \in \mathsf{K}\right) > 2/3$.

(ii) For all $\eta > 0$ there exists a local Doeblin set $\mathsf{C} \in \mathcal{X}$ such that for all $y \in \mathsf{K}$,

$$\sup_{x\in\mathsf{C}^c} g(x,y) \le \eta \left\|g(\cdot,y)\right\|_\infty < \infty. \tag{18}$$

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

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

For the integer $r \in \mathbb{N}^*$ and the set $\mathsf{D} \in \mathcal{X}$ given in (**A1**), define $\mathcal{M}(\mathsf{D},r) \subseteq \mathcal{P}(\mathsf{X},\mathcal{X})$ by

$$\mathcal{M}(\mathsf{D},r) \triangleq \left\{\chi \in \mathcal{P}(\mathsf{X},\mathcal{X}) : \mathbb{E}\left(\ln^- \chi\mathbf{L}\langle Y_0^{\ell-1}\rangle\mathbb{1}_{\mathsf{D}}\right) < \infty \text{ for all } \ell \in \{0,\ldots,r\}\right\}. \tag{19}$$

A simple sufficient condition can be proposed to ensure that $\chi \in \mathcal{M}(\mathsf{D},r)$.

PROPOSITION 4. *Assume that there exists a sequence of sets* $\mathsf{D}_u \in \mathcal{X}$, $u \in \{0, \ldots, r-1\}$, *such that (setting* $\mathsf{D}_r = \mathsf{D}$ *for notational convenience) for some* $\delta > 0$,

$$\inf_{x \in \mathsf{D}_{u-1}} \mathbf{Q}(x, \mathsf{D}_u) \geq \delta, \quad u \in \{1, \ldots, r\}, \tag{20}$$

and

$$\mathbb{E}\left(\ln^- \inf_{x \in \mathsf{D}_u} g(x, Y_0)\right) < \infty, \quad u \in \{0, \ldots, r\}. \tag{21}$$

Then any initial distribution $\chi \in \mathcal{P}(\mathsf{X}, \mathcal{X})$ *satisfying* $\chi(\mathsf{D}_0) > 0$ *belongs to* $\mathcal{M}(\mathsf{D}, r)$.

REMARK 5. *To check* (21) *we typically assume that for any given* $y \in \mathsf{Y}$, *the function* $x \mapsto g(x, y)$ *is continuous and that* D_i, $i \in \{0, \ldots, r-1\}$, *are compact sets. This condition then translates into an assumption on some generalized moments of the process* $(Y_n)_{n \in \mathbb{Z}}$.

REMARK 6. *Assume that* $\mathsf{X} = \mathbb{R}^d$ *for some* $d \in \mathbb{N}^*$ *(or more generally,* X *is a locally compact separable metric space) and that* $\mathcal{X}$ *is the associated Borel* σ*-field. Assume in addition that for any open subset* $\mathsf{O} \in \mathcal{X}$, *the function* $x \to Q(x, \mathsf{O})$ *is lower semi-continuous on the space* X. *Then for any* $\delta > 0$ *and any compact subset* $\mathsf{D}_0 \in \mathcal{X}$, *there exists a sequence of compact subsets* D_u, $u \in \{0, \ldots, r-1\}$, *satisfying* (20).

We are now ready to state our main result.

THEOREM 7. *Assume* (**A1**–**2**). *Then for all* $\chi \in \mathcal{M}(\mathsf{D}, r)$ *and all* $h \in \mathcal{F}(\mathsf{X})$, *the sequence* $(\sigma^2_\chi \langle Y_0^{n-1} \rangle (h))_{n \in \mathbb{N}^*}$ *(defined in (12)) is tight.*

PROOF OF THEOREM 7. Using the definition (9) of the predictive distribution and the decomposition (8) of the likelihood, we get for all $k \in \{0, \ldots, n-1\}$,

$$\phi_\chi \langle Y_0^{n-1} \rangle h = \frac{\chi \mathbf{L} \langle Y_0^{n-1} \rangle h}{\chi \mathbf{L} \langle Y_0^{n-1} \rangle \mathbb{1}_\mathsf{X}} = \frac{\chi \mathbf{L} \langle Y_0^{k-1} \rangle \mathbf{L} \langle Y_k^{n-1} \rangle h}{\chi \mathbf{L} \langle Y_0^{k-1} \rangle \mathbf{L} \langle Y_k^{n-1} \rangle \mathbb{1}_\mathsf{X}}.$$

Plugging this identity into the expression (12) of the asymptotic variance yields

$$\sigma^2_\chi \langle Y_0^{n-1} \rangle (h) = \sum_{k=0}^{n} \int \phi_\chi \langle Y_0^{k-1} \rangle (\mathrm{d}x) \left[\frac{\Delta_{\delta_x, \phi_\chi \langle Y_0^{k-1} \rangle} \langle Y_k^{n-1} \rangle (h, \mathbb{1}_\mathsf{X})}{(\phi_\chi \langle Y_0^{k-1} \rangle \mathbf{L} \langle Y_k^{n-1} \rangle \mathbb{1}_\mathsf{X})^2} \right]^2,$$

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

$$\Delta_{\chi,\chi'}\langle y_k^{n-1}\rangle(f,h) \triangleq \chi\mathbf{L}\langle y_k^{n-1}\rangle f \times \chi'\mathbf{L}\langle y_k^{n-1}\rangle h - \chi\mathbf{L}\langle y_k^{n-1}\rangle h \times \chi'\mathbf{L}\langle y_k^{n-1}\rangle f. \tag{22}$$

Using (9), we obtain for all sequences $y_0^{n-1} \in \mathsf{Y}^n$,

$$\phi_\chi\langle y_0^{k-1}\rangle\mathbf{L}\langle y_k^{n-1}\rangle\mathbb{1}_\mathsf{X} = \frac{\chi\mathbf{L}\langle y_0^{n-1}\rangle\mathbb{1}_\mathsf{X}}{\chi\mathbf{L}\langle y_0^{k-1}\rangle\mathbb{1}_\mathsf{X}} = \prod_{\ell=k}^{n-1}\frac{\chi\mathbf{L}\langle y_0^{\ell}\rangle\mathbb{1}_\mathsf{X}}{\chi\mathbf{L}\langle y_0^{\ell-1}\rangle\mathbb{1}_\mathsf{X}} = \prod_{\ell=k}^{n-1}\pi_\chi\langle y_0^{\ell-1}\rangle(y_\ell),$$

where $\pi_\chi\langle y_0^{\ell-1}\rangle(y_\ell)$ is the density of the conditional distribution of Y_ℓ given $Y_0^{\ell-1}$ (i.e. the one-step observation predictor at time ℓ) defined by

$$\pi_\chi\langle y_0^{\ell-1}\rangle(y_\ell) \triangleq \int \phi_\chi\langle y_0^{\ell-1}\rangle(\mathrm{d}x)\, g(x, y_\ell). \tag{23}$$

With this notation, the likelihood function $\chi\mathbf{L}\langle y_0^{n-1}\rangle\mathbb{1}_\mathsf{X}$ equals the product $\prod_{k=0}^{n-1}\pi_\chi\langle y_0^{k-1}\rangle(y_k)$ (where we let $\pi_\chi\langle y_0^{-1}\rangle(y_0)$ denote the marginal density of Y_0).

Now, using coupling results obtained in [13] one may prove that the predictor distribution forgets its initial distribution exponentially fast under the r-local Doeblin assumption (14). Moreover, this implies that also the log-density of the one-step observation predictor forgets its initial distribution exponentially fast, i.e. for all initial distributions χ and χ' 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,

$$\left|\ln \pi_\chi\langle Y_{-m}^{k-1}\rangle(Y_k) - \ln \pi_{\chi'}\langle Y_{-m}^{k-1}\rangle(Y_k)\right| \le C_{\chi,\chi'}\beta^{k+m}. \tag{24}$$

Using this, it is shown in [13, Proposition 1] that

(i) there exists a function $\pi : \mathsf{Y}^{\mathbb{Z}^-} \times \mathsf{Y} \to \mathbb{R}$ such that for all probability measures $\chi \in \mathcal{M}(\mathsf{D}, r)$,

$$\lim_{m\to\infty}\pi_\chi\langle Y_{-m}^{-1}\rangle(Y_0) = \pi\langle Y_{-\infty}^{-1}\rangle(Y_0), \quad \mathbb{P}\text{-a.s.}$$

Moreover,

$$\mathbb{E}\left(|\ln \pi\langle Y_{-\infty}^{-1}\rangle(Y_0)|\right) < \infty. \tag{25}$$

(ii) for all probability measures $\chi \in \mathcal{M}(\mathsf{D}, r)$, the normalized log-likelihood function converges according to

$$\lim_{n\to\infty} n^{-1} \ln \chi \mathbf{L}\langle Y_0^{n-1}\rangle \mathbb{1}_{\mathsf{X}} = \ell_\infty, \quad \mathbb{P}\text{-a.s.}, \tag{26}$$

where ℓ_∞ is the negated relative entropy, i.e. the expectation of $\ln \pi\langle Y_{-\infty}^{-1}\rangle(Y_0)$ under the stationary distribution, i.e.

$$\ell_\infty \triangleq \mathbb{E}\left(\ln \pi\langle Y_{-\infty}^{-1}\rangle(Y_0)\right). \tag{27}$$

As a first step, we bound the asymptotic variance $\sigma_\chi^2\langle h\rangle(Y_0^{n-1})$ (defined in (12)) by the product of two quantities, namely $\sigma_\chi^2\langle Y_0^{n-1}\rangle(h) \leq A \times B_n$, where

$$A \triangleq \left(\sup_{(k,m)\in\mathbb{N}^2:\, k\leq m} \prod_{\ell=k}^{m-1} \frac{\pi\langle Y_{-\infty}^{\ell-1}\rangle(Y_\ell)}{\pi_\chi\langle Y_0^{\ell-1}\rangle(Y_\ell)} \right)^4, \tag{28}$$

$$B_n \triangleq \sum_{m=0}^{n} \left(\frac{\sup_{x\in\mathsf{X}} |\Delta_{\delta_x, \phi_\chi\langle Y_0^{m-1}\rangle}\langle Y_m^{n-1}\rangle(h, \mathbb{1}_{\mathsf{X}})|}{[\prod_{\ell=m}^{n-1} \pi\langle Y_{-\infty}^{\ell-1}\rangle(Y_\ell)]^2} \right)^2. \tag{29}$$

The quantity (28) can be bounded using the exponential forgetting (24) of the one-step predictor log-density. More precisely, note that

$$\pi_\chi\langle Y_{-m}^{\ell-1}\rangle(Y_\ell) = \frac{\chi \mathbf{L}\langle Y_{-m}^{\ell}\rangle \mathbb{1}_{\mathsf{X}}}{\chi \mathbf{L}\langle Y_{-m}^{\ell-1}\rangle \mathbb{1}_{\mathsf{X}}};$$

thus, by applying Proposition 11(ii) we conclude that there exist $\beta \in]0,1[$ and a $\mathbb{P}$-a.s. finite random variable C_χ such that for all $n \in \mathbb{N}$,

$$\begin{aligned}\prod_{\ell=k}^{n} \frac{\pi\langle Y_{-\infty}^{\ell-1}\rangle(Y_\ell)}{\pi_\chi\langle Y_0^{\ell-1}\rangle(Y_\ell)} &= \prod_{\ell=k}^{n}\prod_{m=0}^{\infty} \frac{\pi_\chi\langle Y_{-m-1}^{\ell-1}\rangle(Y_\ell)}{\pi_\chi\langle Y_{-m}^{\ell-1}\rangle(Y_\ell)} \\ &\leq \prod_{\ell=k}^{n}\prod_{m=0}^{\infty} \exp(C_\chi \beta^{\ell+m}) \leq \exp(C_\chi/(1-\beta)^2) < \infty, \quad \mathbb{P}\text{-a.s.},\end{aligned} \tag{30}$$

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

Consider now the second quantity (29). Since the process $(Y_n)_{n\in\mathbb{Z}}$ is strictly stationary, Y_0^{n-1} has the same distribution as Y_{-n}^{-1} for all $n \in \mathbb{N}^*$. Therefore, for all $n \in \mathbb{N}^*$, the random variable B_n has the same distribution as

$$\tilde{B}_n \triangleq \sum_{m=0}^{n} \left(\frac{\sup_{x\in\mathsf{X}} |\Delta_{\delta_x, \phi_\chi\langle Y_{-n}^{-m-1}\rangle}\langle Y_{-m}^{-1}\rangle(h, \mathbb{1}_{\mathsf{X}})|}{[\prod_{\ell=1}^{m} \pi\langle Y_{-\infty}^{-\ell-1}\rangle(Y_{-\ell})]^2} \right)^2. \tag{31}$$

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

$$
\begin{aligned}
(32)\quad & \frac{\sup_{x\in\mathsf{X}}|\Delta_{\delta_x,\phi_\chi\langle Y_{-n}^{-m-1}\rangle}\langle Y_{-m}^{-1}\rangle(h,\mathbb{1}_\mathsf{X})|}{[\prod_{\ell=1}^{m}\pi\langle Y_{-\infty}^{-\ell-1}\rangle(Y_{-\ell})]^2} \\
&= \left(\frac{\|\mathbf{L}\langle Y_{-m}^{-1}\rangle\mathbb{1}_\mathsf{X}\|_\infty}{\prod_{\ell=1}^{m}\pi\langle Y_{-\infty}^{-\ell-1}\rangle(Y_{-\ell})}\right)^2\frac{\sup_{x\in\mathsf{X}}|\Delta_{\delta_x,\phi_\chi\langle Y_{-n}^{-m-1}\rangle}\langle Y_{-m}^{-1}\rangle(h,\mathbb{1}_\mathsf{X})|}{\|\mathbf{L}\langle Y_{-m}^{-1}\rangle\mathbb{1}_\mathsf{X}\|_\infty^2},
\end{aligned}
$$

and consider each factor separately.

We will show that the first factor in (32) grows at most subgeometrically fast. Indeed, note that

$$
\left(\frac{\|\mathbf{L}\langle Y_{-m}^{-1}\rangle\mathbb{1}_\mathsf{X}\|_\infty}{\prod_{\ell=1}^{m}\pi\langle Y_{-\infty}^{-\ell-1}\rangle(Y_{-\ell})}\right)^2 = \exp(m\varepsilon_m),
$$

where

$$
\varepsilon_m \triangleq \frac{2}{m}\left(\ln\left\|\mathbf{L}\langle Y_{-m}^{-1}\rangle\mathbb{1}_\mathsf{X}\right\|_\infty - \sum_{\ell=1}^{m}\ln\pi\langle Y_{-\infty}^{-\ell-1}\rangle(Y_{-\ell})\right).
$$

According to Lemma 12, $\varepsilon_m \to 2(\ell_\infty - \ell_\infty) = 0$, $\mathbb{P}$-a.s., as $m\to\infty$.

The second factor in (32) is handled using Proposition 11(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 $(m,n)\in(\mathbb{N}^*)^2$,

$$
(33)\qquad \frac{\sup_{x\in\mathsf{X}}|\Delta_{\delta_x,\phi_\chi\langle Y_{-n}^{-m-1}\rangle}\langle Y_{-m}^{-1}\rangle(h,\mathbb{1}_\mathsf{X})|}{\|\mathbf{L}\langle Y_{-m}^{-1}\rangle\mathbb{1}_\mathsf{X}\|_\infty^2} \le C\beta^m\|h\|_\infty .
$$

This concludes the proof. □

Having established tightness of the asymptotic variance, the asymptotic L^p error given in Theorem 8 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 normalized L^p errors. After this, weak convergence implies convergence of moments, implying in turn convergence of the L^p error.

THEOREM 8. *Assume that the sequence $(\sigma_\chi^2\langle Y_0^{n-1}\rangle(h))_{n\in\mathbb{N}^*}$ (defined in (12)) is tight for all functions $h\in\mathcal{F}(\mathsf{X})$. Then, for all functions $h\in\mathcal{F}(\mathsf{X})$,* *constants* $p \in \mathbb{R}_+^*$*, and initial distributions* $\chi \in \mathcal{M}(\mathsf{D}, r)$ *it holds,* $\mathbb{P}$*-a.s.,*

$$\lim_{N \to \infty} \sqrt{N} \mathbb{E}^{1/p} \left(\left| \phi_\chi^N \langle Y_0^{n-1} \rangle h - \phi_\chi \langle Y_0^{n-1} \rangle h \right|^p \middle| Y_0^{n-1} \right) = \sqrt{2} \sigma_\chi \langle Y_0^{n-1} \rangle (h) \left(\frac{\Gamma((p+1)/2)}{\sqrt{2\pi}} \right)^{1/p},$$

where Γ *is the gamma function.*

Proof. Recall that if $(A_N)_{N \in \mathbb{N}}$ is a sequence of random variables such that $A_N \xrightarrow{\mathcal{D}} A$ as $N \to \infty$ 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 e.g. [27, Theorem A, p. 14]. Now set, for $n \in \mathbb{N}^*$,

$$A_{N,\chi} \langle Y_0^{n-1} \rangle (h) \triangleq \sqrt{N} \left(\phi_\chi^N \langle Y_0^{n-1} \rangle h - \phi_\chi \langle Y_0^{n-1} \rangle h \right).$$

For all $q > p$ it holds that

$$\begin{aligned} &\sup_{N \in \mathbb{N}^*} \mathbb{E} \left(\left| A_{N,\chi} \langle Y_0^{n-1} \rangle (h) \right|^q \middle| Y_0^{n-1} \right) \\ &= \sup_{N \in \mathbb{N}^*} \int_0^\infty \mathbb{P} \left(\left| A_{N,\chi} \langle Y_0^{n-1} \rangle (h) \right| \geq \epsilon^{1/q} \middle| Y_0^{n-1} \right) \mathrm{d}\epsilon \\ &= q \sup_{N \in \mathbb{N}^*} \int_0^\infty \epsilon^{q-1} \mathbb{P} \left(\left| A_{N,\chi} \langle Y_0^{n-1} \rangle (h) \right| \geq \epsilon \middle| Y_0^{n-1} \right) \mathrm{d}\epsilon. \end{aligned}$$

Now, note that (**A2**)(ii) implies that $\|g(\cdot, Y_n)\|_\infty$ is $\mathbb{P}$-a.s. finite for all $n \in \mathbb{N}$. Thus, the assumptions of [11, Lemma 2.1] are fulfilled (see also [8, Theorem 3.39]), which implies that there exist, for all $n \in \mathbb{N}$, positive constants B_n and C_n such that for all $N \in \mathbb{N}$, all $h \in \mathcal{F}(\mathsf{X})$, and all $\epsilon > 0$,

$$\mathbb{P} \left(|A_{N,\chi} \langle Y_0^{n-1} \rangle (h)| \geq \epsilon \middle| Y_0^{n-1} \right) \leq B_n \exp(-C_n \epsilon^2). \tag{34}$$

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

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

which establishes, via [28, Lemma II.6.3, p. 190], that $(|A_{N,\chi} \langle Y_0^{n-1} \rangle (h)|^p)_{N \in \mathbb{N}}$ is uniformly integrable conditionally on Y_0^{n-1}, i.e.

$$\lim_{M \to \infty} \sup_{N \in \mathbb{N}^*} \mathbb{E} \left(\left| A_{N,\chi} \langle Y_0^{n-1} \rangle (h) \right|^p \mathbb{1}_{\{|A_{N,\chi} \langle Y_0^{n-1} \rangle (h)| \geq M\}} \middle| Y_0^{n-1} \right) = 0, \quad \mathbb{P}\text{-a.s.}$$

We may now complete the proof by applying Theorem 1, which states that conditionally on Y_0^{n-1}, as $N \to \infty$,

$$A_{N,\chi} \langle Y_0^{n-1} \rangle (h) \xrightarrow{\mathcal{D}} \sigma_\chi \langle Y_0^{n-1} \rangle (h) Z,$$

where Z is a standard normal-distributed random variable. □

4. Applications. 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 (**A1–2**) 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 $\mathsf{X} = \mathbb{R}^{d_x}$ and $\mathsf{Y} = \mathbb{R}^{d_y}$ and define state and observation sequences through the linear dynamic system

$$
\begin{aligned}
X_{k+1} &= AX_k + RU_k, \\
Y_k &= BX_k + SV_k,
\end{aligned}
$$

where $(U_k, V_k)_{k\geq 0}$ is an i.i.d. sequence of 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.

For any $n \in \mathbb{N}$, define the observability and controlability matrices $\mathcal{O}_n$ and $\mathcal{C}_n$ by

$$
\mathcal{O}_n \triangleq \begin{bmatrix} B \\ BA \\ BA^2 \\ \vdots \\ BA^{n-1} \end{bmatrix} \quad \text{and} \quad \mathcal{C}_n \triangleq \left[A^{n-1}R \; A^{n-2}R \ldots R\right], \tag{35}
$$

respectively. We assume the following.

(**LGSS1**) The pair (A, B) is observable and the pair (A, R) is controllable, i.e. there exists $r \in \mathbb{N}$ such that the observability matrix $\mathcal{O}_r$ and the controllability matrix $\mathcal{C}_r$ have full rank.

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

(**LGSS3**) $\mathbb{E}\left(\|Y_0\|^2\right) < \infty$.

We now check Assumptions (**A1–2**). 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 $R^t R$ 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 $\|z\|_A^2 \triangleq {}^t z A^{-1} z$. In addition, define for any $n \in \mathbb{N}$,

$$
\mathcal{F}_n \triangleq \mathcal{D}_n {}^t\mathcal{D}_n + \mathcal{S}_n {}^t\mathcal{S}_n , \tag{36}
$$

where

$$
\mathcal{D}_n \triangleq \begin{bmatrix} 0 & 0 & \cdots & 0 \\ BR & \ddots & & 0 \\ BAR & BR & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ BA^{n-2}R & BA^{n-3}R & \cdots & BR \end{bmatrix} , \quad \mathcal{S}_n \triangleq \begin{bmatrix} S & 0 & \dots & 0 \\ 0 & S & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & S \end{bmatrix} .
$$

Under (**LGSS2**), the matrix $\mathcal{F}_n$ is positive definite for any $n \geq r$. When the state process is initialized at $x_0 \in \mathsf{X}$, the likelihood of the observations $y_0^{n-1} \in \mathsf{Y}^n$ is given by

$$
\delta_{x_0} \mathbf{L}\langle y_0^{n-1} \rangle \mathbb{1}_{\mathsf{X}} = (2\pi)^{-n d_y} \det{}^{-1/2}(\mathcal{F}_n) \exp\left(-\frac{1}{2} \|\mathbf{y}_{n-1} - \mathcal{O}_n x_0\|_{\mathcal{F}_n}^2 \right) ,
$$

where $\mathbf{y}_{n-1} \triangleq {}^t[{}^t y_0, {}^t y_1, \ldots, {}^t y_{n-1}]$ and $\mathcal{O}_n$ is defined in (35).

We first consider (**A1**). Under (**LGSS1**), the observability matrix $\mathcal{O}_r$ is full rank, and we have for any compact subset $\mathsf{K} \subset \mathsf{Y}^r$,

$$
\lim_{\|x_0\| \to \infty} \inf_{y_0^{r-1} \in \mathsf{K}} \|\mathbf{y}_{r-1} - \mathcal{O}_r x_0\|_{\mathcal{F}_r} = \infty ,
$$

showing that for all $\eta > 0$, we may choose a compact set $\mathsf{C} \subset \mathbb{R}^{d_x}$ such that (18) is satisfied. It remains to prove that any compact set C is an r-local Doeblin set satisfying the condition (16). For any $y_0^{r-1} \in \mathsf{Y}^r$ and $x_0 \in \mathsf{X}$, the measure $\delta_{x_0} \mathbf{L}\langle y_0^{r-1} \rangle$ is absolutely continuous with respect to the Lebesgue measure on $(\mathsf{X}, \mathcal{X})$ with Radon-Nikodym derivative $\ell\langle y_0^{r-1} \rangle(x_0, x_r)$ given (up to an irrelevant multiplicative factor) by

$$
\ell\langle y_0^{r-1} \rangle(x_0, x_r) \propto \det{}^{-1/2}(\mathcal{G}_r) \exp\left(-\frac{1}{2} \left\| \begin{bmatrix} \mathbf{y}_{r-1} \\ x_r \end{bmatrix} - \begin{bmatrix} \mathcal{O}_r \\ A^r \end{bmatrix} x_0 \right\|_{\mathcal{G}_r}^2 \right) , \tag{37}
$$

where the covariance matrix $\mathcal{G}_r$ is

$$
\mathcal{G}_r \triangleq \begin{bmatrix} \mathcal{D}_r \\ \mathcal{C}_r \end{bmatrix} \begin{bmatrix} {}^t\mathcal{D}_r {}^t\mathcal{C}_r \end{bmatrix} + \begin{bmatrix} \mathcal{S}_r \\ \mathbf{0} \end{bmatrix} \begin{bmatrix} {}^t\mathcal{S}_r {}^t\mathbf{0} \end{bmatrix} .
$$

The proof of (37) relies on the positivity of $\mathcal{G}_r$, which requires further discussion. By construction, the matrix $\mathcal{G}_r$ is non-negative. For all $\mathbf{y}_{r-1} \in \mathsf{Y}^r$ and $x \in \mathsf{X}$, the equation

$$[{}^t\mathbf{y}_{r-1}\,{}^tx]\mathcal{G}_r \begin{bmatrix} \mathbf{y}_{r-1} \\ x \end{bmatrix} = \left\|{}^t\mathcal{D}_r\mathbf{y}_{r-1} + {}^t\mathcal{C}_r x\right\|^2 + \left\|{}^t\mathcal{S}_r\mathbf{y}_{r-1}\right\|^2 = 0$$

implies that $\|{}^t\mathcal{D}_r\mathbf{y}_{r-1} + {}^t\mathcal{C}_r x\|^2 = 0$ and $\|{}^t\mathcal{S}_r\mathbf{y}_{r-1}\|^2 = 0$. Since the matrix $\mathcal{S}_r$ has full rank, this implies that $\mathbf{y}_{r-1} = 0$. Since also $\mathcal{C}_r$ has full rank (the pair (A, R) is commandable), this implies in turn that $x = 0$. Therefore, the matrix $\mathcal{G}_r$ is positive definite and the function

$$(x_0, x_r) \mapsto \left\| \begin{bmatrix} \mathbf{y}_{r-1} \\ x_r \end{bmatrix} - \begin{bmatrix} \mathcal{O}_r \\ A^r \end{bmatrix} x_0 \right\|^2_{\mathcal{G}_r}$$

continuous for all $\mathbf{y}_{r-1}$. It is therefore bounded on any compact subset of X^2. This implies that every non-empty compact set $\mathsf{C} \subset \mathbb{R}^{d_x}$ is an r-local Doeblin set, with $\lambda_\mathsf{C}(\cdot) = \lambda^{\mathrm{Leb}}(\cdot)/\lambda^{\mathrm{Leb}}(\mathsf{C})$ and

$$\begin{aligned} \epsilon^-_\mathsf{C}(y_0^{r-1}) &= \left(\lambda^{\mathrm{Leb}}(\mathsf{C})\right)^{-1} \inf_{(x_0,x_r)\in\mathsf{C}^2} \ell\langle y_0^{r-1}\rangle(x_0, x_r), \\ \epsilon^+_\mathsf{C}(y_0^{r-1}) &= \left(\lambda^{\mathrm{Leb}}(\mathsf{C})\right)^{-1} \sup_{(x_0,x_r)\in\mathsf{C}^2} \ell\langle y_0^{r-1}\rangle(x_0, x_r). \end{aligned}$$

Consequently, condition (16) is satisfied for any compact set $\mathsf{K} \subseteq \mathsf{Y}^{r-1}$. It remains to verify (**A1**) (iii). Under (**LGSS1**), the measure $\delta_{x_0}\mathbf{L}\langle y_0^{r-1}\rangle$ is absolutely continuous with respect to the Lebesgue measure λ^{Leb}; therefore, for any set $\mathsf{D} \subset \mathbb{R}^{d_x}$,

$$\inf_{x_0\in\mathsf{D}} \delta_{x_0}\mathbf{L}\langle y_0^{r-1}\rangle(\mathsf{D}) \geq \inf_{(x_0,x_r)\in\mathsf{D}^2} \ell\langle y_0^{r-1}\rangle(x_0, x_r)\lambda^{\mathrm{Leb}}(\mathsf{D}).$$

Take D to be any compact set with positive Lebesgue measure. Now,

$$\begin{aligned} &\sup_{(x_0,x_r)\in\mathsf{D}^2} \left\| \begin{bmatrix} \mathbf{y}_{r-1} \\ x_r \end{bmatrix} - \begin{bmatrix} \mathcal{O}_r \\ A^r \end{bmatrix} x_0 \right\|^2_{\mathcal{G}_r} \\ &\qquad \leq 2\lambda_{\max}(\mathcal{G}_r) \left\{ \|\mathbf{y}_{r-1}\|^2 + \max_{x\in\mathsf{D}} \|x\|^2 \left[1 + \lambda_{\max}\left({}^t\mathcal{O}_r\mathcal{O}_r + {}^tA^rA^r\right)\right] \right\}, \end{aligned}$$

where $\lambda_{\max}(A)$ is the largest eigenvalue of A. Under (**LGSS3**), $\mathbb{E}\left(\|Y_0\|^2\right) < \infty$, 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 λ^{Leb} as the Lebesgue measure on Y, $g(x,y)$ is, for each $x \in \mathsf{X}$, a Gaussian density with covariance matrix $S^t S$. We therefore have

$$\sup_{x \in \mathsf{X}} g(x,y) = (2\pi)^{-d_y/2} \det{}^{-1/2}(S^t S) < \infty$$

for all $y \in \mathsf{Y}$, which verifies (**A2**)(i–ii).

To conclude this discussion, we need to specify more explicitly the set $\mathcal{M}(\mathsf{D}, r)$ (see (19)) of possible initial distributions. Using Proposition 4, we verify the sufficient conditions (20) and (21). To check (20), we use Remark 6: For any open subset $\mathsf{O} \subset \mathbb{R}^{d_x}$ and $x \in \mathsf{X}$, $\mathbf{Q}(x, \mathsf{O}) = \mathbb{E}\left(\mathbb{1}_{\mathsf{O}}(Ax + RU)\right)$, 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 function $\mathbb{1}_{\mathsf{O}}$ is lower semi-continuous we obtain, via Fatou's Lemma,

$$\liminf_{n \to \infty} \mathbf{Q}(x_n, \mathsf{O}) \geq \mathbb{E}\left(\liminf_{n \to \infty} \mathbb{1}_{\mathsf{O}}(Ax_n + RU)\right) \geq \mathbf{Q}(x, \mathsf{O}),$$

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

Assumption (**LGSS2**) implies that for all $(x,y) \in \mathsf{X} \times \mathsf{Y}$,

$$\ln g(x,y) \geq -\frac{d_y}{2} \ln(2\pi) - \frac{1}{2} \ln \det{}^{-1/2}(S^t S) \\ - \left[\lambda_{\min}\left(S^t S\right)\right]^{-1} \left(\|y\|^2 + \|Bx\|^2\right),$$

where $\lambda_{\min}(S^t S)$ is the minimal eigenvalue of $S^t S$. Therefore (21) is satisfied under (**LGSS3**). Consequently, we may apply Theorem 7 to establish tightness of the asymptotic variance for any initial distribution $\chi \in \mathcal{P}(\mathsf{X}, \mathcal{X})$ as soon as the process $(Y_k)_{k \in \mathbb{Z}}$ is strictly stationary ergodic and $\mathbb{E}\left(\|Y_0\|^2\right) < \infty$.

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

$$X_k = T(X_{k-1}) + \Sigma(X_{k-1}) \zeta_k, \tag{38}$$

where $(\zeta_k)_{k \in \mathbb{N}^*}$ is an i.i.d. sequence of d-dimensional random vectors with density ρ_ζ with respect to the Lebesgue measure λ^{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 $x \in \mathsf{X}$. Such models (38) are sometimes referred to as vector *autoregressive conditional heteroscedasticity* (ARCH) *models* and cover many models 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 $(x, x') \mapsto q(x, x')$ is a positive continuous function on X^2. In addition, $\sup_{(x,x')\in\mathsf{X}^2} q(x, x') < \infty$.

(**NL2**) For any compact subset $\mathsf{K} \subset \mathsf{Y}$,

$$\lim_{\|x\|\to\infty} \sup_{y\in\mathsf{K}} \frac{g(x, y)}{\sup_{x'\in\mathsf{X}} g(x', y)} = 0.$$

(**NL3**) For all $(x, y) \in \mathsf{X} \times \mathsf{Y}$, $g(x, y) > 0$ and

$$\mathbb{E}\left(\ln^+ \sup_{x\in\mathsf{X}} g(x, Y_0)\right) < \infty.$$

(**NL4**) There exists a compact subset $\mathsf{D} \subset \mathsf{Y}$ such that

$$\mathbb{E}\left(\ln^- \inf_{x\in\mathsf{D}} g(x, Y_0)\right) < \infty.$$

Under (**NL1**), every compact set $\mathsf{C} \subset \mathsf{X} = \mathbb{R}^d$ with positive Lebesgue measure is 1-small and therefore local Doeblin with $\lambda_{\mathsf{C}}(\cdot) = \lambda^{\mathrm{Leb}}(\cdot \cap \mathsf{C})/\lambda^{\mathrm{Leb}}(\mathsf{C})$, $\varphi_{\mathsf{C}}\langle y_0\rangle = \lambda^{\mathrm{Leb}}(\mathsf{C})$, and

$$\epsilon_{\mathsf{C}}^- = \inf_{(x,x')\in\mathsf{C}^2} q(x, x'),$$
$$\epsilon_{\mathsf{C}}^+ = \sup_{(x,x')\in\mathsf{C}^2} q(x, x').$$

Under (**NL1**) and (**NL2**), the conditions (18) and (16) are satisfied with $r = 1$. In addition, (17) is implied by (**NL1**) and (**NL4**). Consequently, Assumption (**A1**) holds. Moreover, (**A2**) follows directly from (**NL3**). So, finally, under (**NL1**)–(**NL4**) we conclude, using Theorem 7 and Proposition 4, that the asymptotic variance of the bootstrap particle filter is tight for any initial distribution χ such that $\chi(\mathsf{D}) > 0$.

5. Proofs.

5.1. *Forgetting of the initial distribution.*

LEMMA 9. *Assume (**A1**–**2**). 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 $z_0^{n-1} \in \mathsf{Y}^{nr}$, where $r \in \mathbb{N}^*$ is as in (**A1**) and $z_i = y_{ir}^{(i+1)r-1}$, satisfying*

$$n^{-1} \sum_{i=0}^{n-1} \mathbb{1}_{\mathsf{K}}(z_i) \geq \gamma,$$

all functions f and h in $\mathcal{F}_+(\mathsf{X})$, all finite measures χ and χ' in $\mathcal{M}(\mathsf{X}, \mathcal{X})$, and all $\eta \in]0,1[$,

$$\begin{aligned} (39) \quad & \left|\Delta_{\chi,\chi'}\langle z_0^{n-1}\rangle(f,h)\right| \\ & \leq \rho_\gamma^n(\eta) \left(\chi \mathbf{L}\langle z_0^{n-1}\rangle f \times \chi' \mathbf{L}\langle z_0^{n-1}\rangle h + \chi' \mathbf{L}\langle z_0^{n-1}\rangle f \times \chi \mathbf{L}\langle z_0^{n-1}\rangle h\right) \\ & \qquad + C_\gamma(\eta)\eta^n \|f\|_\infty \|h\|_\infty \prod_{i=0}^{n-1} \|\mathbf{L}\langle z_i\rangle \mathbb{1}_{\mathsf{X}}\|_\infty^2 \chi(\mathsf{X})\chi'(\mathsf{X}), \end{aligned}$$

$$\begin{aligned} (40) \quad & \left|\ln\left(\frac{\chi \mathbf{L}\langle z_0^{n-1}\rangle h}{\chi \mathbf{L}\langle z_0^{n-1}\rangle f}\right) - \ln\left(\frac{\chi' \mathbf{L}\langle z_0^{n-1}\rangle h}{\chi' \mathbf{L}\langle z_0^{n-1}\rangle f}\right)\right| \leq (1-\rho_\gamma(\eta))^{-1} \\ & \times \left(2\rho_\gamma^n(\eta) + \frac{C_\gamma(\eta)\eta^n \|f\|_\infty \|h\|_\infty \prod_{i=0}^{n-1} \|\mathbf{L}\langle z_i\rangle \mathbb{1}_{\mathsf{X}}\|_\infty \chi(\mathsf{X})\chi'(\mathsf{X})}{\chi \mathbf{L}\langle z_0^{n-1}\rangle f \times \chi' \mathbf{L}\langle z_0^{n-1}\rangle h}\right), \end{aligned}$$

$$\begin{aligned} (41) \quad & \left|\frac{\chi \mathbf{L}\langle z_0^{n-1}\rangle h}{\chi \mathbf{L}\langle z_0^{n-1}\rangle f} - \frac{\chi' \mathbf{L}\langle z_0^{n-1}\rangle h}{\chi' \mathbf{L}\langle z_0^{n-1}\rangle f}\right| \leq \rho_\gamma^n(\eta) \left(\frac{\chi \mathbf{L}\langle z_0^{n-1}\rangle h}{\chi \mathbf{L}\langle z_0^{n-1}\rangle f} + \frac{\chi' \mathbf{L}\langle z_0^{n-1}\rangle h}{\chi' \mathbf{L}\langle z_0^{n-1}\rangle f}\right) \\ & \qquad + \frac{C_\gamma(\eta)\eta^n \|h\|_\infty \|f\|_\infty \prod_{i=0}^{n-1} \|\mathbf{L}\langle z_i\rangle \mathbb{1}_{\mathsf{X}}\|_\infty^2 \chi(\mathsf{X})\chi'(\mathsf{X})}{\chi \mathbf{L}\langle z_0^{n-1}\rangle f \times \chi' \mathbf{L}\langle z_0^{n-1}\rangle h}. \end{aligned}$$

PROOF. The proof is straightforwardly adapted from [13, Proposition 5]. □

LEMMA 10. *Assume (**A1**). Then there exists a constant $\kappa > 0$ such that for all $\chi \in \mathcal{M}(\mathsf{D}, r)$ (where $\mathcal{M}(\mathsf{D}, r)$ is defined in (19)),*

$$(42) \qquad \inf_{(k,m) \in \mathbb{N}^* \times \mathbb{N}} \kappa^{(k+m)} \chi \mathbf{L}\langle Y_{-m}^{k-1}\rangle \mathbb{1}_{\mathsf{X}} > 0, \quad \mathbb{P}\text{-}a.s.,$$

and

$$(43) \qquad \inf_{(k,m) \in \mathbb{N}^* \times \mathbb{N}} \kappa^{(k+m)} \|\mathbf{L}\langle Y_{-m}^{k-1}\rangle \mathbb{1}_{\mathsf{X}}\|_\infty > 0, \quad \mathbb{P}\text{-}a.s.$$

PROOF. To derive (42) we first establish that

$$
\liminf_{k+m\to\infty} (k+m)^{-1} \left(\ln \chi \mathbf{L}\langle Y_{-m}^{k-1} \rangle \mathbb{1}_{\mathsf{X}} \right) \geq -r\mathbb{E}\left(\ln^{-} \inf_{x\in\mathsf{D}} \delta_x \mathbf{L}\langle Y_0^{r-1} \rangle \mathbb{1}_{\mathsf{D}} \right) > -\infty, \quad \mathbb{P}\text{-a.s.,} \tag{44}
$$

where the last inequality follows from (**A1**)(iii). We now establish the first inequality in (44). Set $a_{k,m} \triangleq -k + \lfloor (k+m)/r \rfloor r$ and note that $-a_{k,m} \in \{-m, \ldots, -m+r-1\}$. Then, write

$$
\begin{aligned}
&\ln \chi \mathbf{L}\langle Y_{-m}^{k-1} \rangle \mathbb{1}_{\mathsf{X}} \\
&\geq \ln \chi \mathbf{L}\langle Y_{-m}^{-a_{k,m}} \rangle \mathbb{1}_{\mathsf{D}} + \sum_{i=0}^{\lfloor (k+m)/r \rfloor - 1} \ln \inf_{x\in\mathsf{D}} \delta_x \mathbf{L}\langle Y_{-a_{k,m}+ir}^{-a_{k,m}+(i+1)r-1} \rangle \mathbb{1}_{\mathsf{D}} \\
&\geq -\sum_{i=0}^{r-1} \ln^{-} \chi \mathbf{L}\langle Y_{-m}^{-m+i} \rangle \mathbb{1}_{\mathsf{D}} - \sum_{i=0}^{\lfloor (k+m)/r \rfloor - 1} \ln^{-} \inf_{x\in\mathsf{D}} \delta_x \mathbf{L}\langle Y_{-a_{k,m}+ir}^{-a_{k,m}+(i+1)r-1} \rangle \mathbb{1}_{\mathsf{D}}.
\end{aligned} \tag{45}
$$

For $i \in \mathbb{N}$, set $[i]_r \triangleq i - \lfloor i/r \rfloor r$. With this notation, $a_{k,m} = [a_{k,m}]_r + \lfloor a_{k,m}/r \rfloor r$. Then, since $[i]_r \in \{0, \ldots, r-1\}$,

$$
\begin{aligned}
&-\sum_{i=0}^{\lfloor (k+m)/r \rfloor - 1} \ln^{-} \inf_{x\in\mathsf{D}} \delta_x \mathbf{L}\langle Y_{-a_{k,m}+ir}^{-a_{k,m}+(i+1)r-1} \rangle \mathbb{1}_{\mathsf{D}} \\
&= -\sum_{i=0}^{\lfloor (k+m)/r \rfloor - 1} \ln^{-} \inf_{x\in\mathsf{D}} \delta_x \mathbf{L}\langle Y_{-[a_{k,m}]_r + (i - \lfloor a_{k,m}/r \rfloor) r}^{-[a_{k,m}]_r + (i - \lfloor a_{k,m}/r \rfloor + 1) r - 1} \rangle \mathbb{1}_{\mathsf{D}} \\
&\geq -\sum_{j=0}^{r-1} \sum_{i=0}^{\lfloor (k+m)/r \rfloor - 1} \ln^{-} \inf_{x\in\mathsf{D}} \delta_x \mathbf{L}\langle Y_{-j + (i - \lfloor a_{k,m}/r \rfloor) r}^{-j + (i - \lfloor a_{k,m}/r \rfloor + 1) r - 1} \rangle \mathbb{1}_{\mathsf{D}} \\
&= -\sum_{j=0}^{r-1} \sum_{\ell = -\lfloor a_{k,m}/r \rfloor}^{\lfloor (k+m)/r \rfloor - \lfloor a_{k,m}/r \rfloor - 1} \ln^{-} \inf_{x\in\mathsf{D}} \delta_x \mathbf{L}\langle Y_{-j+\ell r}^{-j+(\ell+1)r-1} \rangle \mathbb{1}_{\mathsf{D}},
\end{aligned} \tag{46}
$$

where the last identity follows by reindexing the summation. We now plug (46) into (45); the ergodicity of the process $(Z_n)_{n\in\mathbb{Z}}$ (Assumption (**A1**)(i)) then implies, via Lemma 13, $\mathbb{P}$-a.s.,

$$\liminf_{k+m\to\infty} (k+m)^{-1} \left(\ln \chi \mathbf{L}\langle Y_{-m}^{k-1} \rangle \mathbb{1}_{\mathsf{X}} \right) \\ \geq \sum_{j=0}^{r-1} \mathbb{E} \left(\ln^- \inf_{x\in\mathsf{D}} \delta_x \mathbf{L}\langle Y_{-j}^{-j+r-1} \rangle \mathbb{1}_{\mathsf{D}} \right) = -r\mathbb{E} \left(\ln^- \inf_{x\in\mathsf{D}} \delta_x \mathbf{L}\langle Y_0^{r-1} \rangle \mathbb{1}_{\mathsf{D}} \right) ,$$

which shows (44). Now, choose a constant κ such that

$$-r\mathbb{E} \left(\ln^- \inf_{x\in\mathsf{D}} \delta_x \mathbf{L}\langle Y_0^{r-1} \rangle \mathbb{1}_{\mathsf{D}} \right) > -\ln \kappa > -\infty .$$

According to (44), there exists a $\mathbb{P}$-a.s. finite $\mathbb{N}^*$-valued random variable N such that if $k+m \geq N$,

$$\ln \chi \mathbf{L}\langle Y_0^{r-1} \rangle \mathbb{1}_{\mathsf{X}} \geq (-\ln \kappa)(k+m),$$

which implies that

$$\inf_{k+m\geq N} \kappa^{k+m} \chi \mathbf{L}\langle Y_0^{r-1} \rangle \mathbb{1}_{\mathsf{X}} \geq 1 .$$

On the other hand, Assumption (**A2**) implies that for all $(k,m) \in \mathbb{N}^* \times \mathbb{N}$, $\chi \mathbf{L}\langle Y_0^{r-1} \rangle \mathbb{1}_{\mathsf{X}} > 0$, $\mathbb{P}$-a.s. This completes the proof of (42). Finally, the proof of (43) follows by combining

$$\|\mathbf{L}\langle Y_{-m}^{k-1} \rangle \mathbb{1}_{\mathsf{X}}\|_\infty \geq \chi \mathbf{L}\langle Y_{-m}^{k-1} \rangle \mathbb{1}_{\mathsf{X}}$$

and (42). □

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

$$\mathcal{M}\langle y_{-m}^k \rangle (\chi) \\ \triangleq \left\{ \tilde{\chi} \in \mathcal{P}(\mathsf{X}, \mathcal{X}) : \|g(\cdot, y_k)\|_\infty \times \tilde{\chi} \mathbf{L}\langle y_{-m}^{k-1} \rangle \mathbb{1}_{\mathsf{X}} \geq (1/2) \chi \mathbf{L}\langle y_{-m}^k \rangle \mathbb{1}_{\mathsf{X}} \right\} \tag{47}$$

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

PROPOSITION 11. *Assume (**A1–2**). Then there exists a constant $\beta \in]0,1[$ such that the following holds.*

(i) *For all probability measures χ and χ' in $\mathcal{M}(\mathsf{D}, r)$ 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 $\tilde{\chi} \in \mathcal{M}\langle Y_{-m}^k \rangle(\chi)$,*

$$\ln\left(\frac{\tilde{\chi}\mathbf{L}\langle Y_{-m}^k\rangle \mathbb{1}_\mathsf{X}}{\tilde{\chi}\mathbf{L}\langle Y_{-m}^{k-1}\rangle \mathbb{1}_\mathsf{X}}\right) - \ln\left(\frac{\chi'\mathbf{L}\langle Y_{-m}^k\rangle \mathbb{1}_\mathsf{X}}{\chi'\mathbf{L}\langle Y_{-m}^{k-1}\rangle \mathbb{1}_\mathsf{X}}\right) \leq C_{\chi,\chi'}\beta^{k+m}, \quad \mathbb{P}\text{-a.s.}$$

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

$$\left|\ln\left(\frac{\chi\mathbf{L}\langle Y_{-m}^k\rangle \mathbb{1}_\mathsf{X}}{\chi\mathbf{L}\langle Y_{-m}^{k-1}\rangle \mathbb{1}_\mathsf{X}}\right) - \ln\left(\frac{\chi\mathbf{L}\langle Y_{-m-1}^k\rangle \mathbb{1}_\mathsf{X}}{\chi\mathbf{L}\langle Y_{-m-1}^{k-1}\rangle \mathbb{1}_\mathsf{X}}\right)\right| \leq C_\chi\beta^{k+m}, \quad \mathbb{P}\text{-a.s.}$$

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

$$\frac{|\Delta_{\chi,\chi'}\langle Y_{-m}^{-1}\rangle(h, \mathbb{1}_\mathsf{X})|}{\|\mathbf{L}\langle Y_{-m}^{-1}\rangle \mathbb{1}_\mathsf{X}\|_\infty^2} \leq C\beta^m \|h\|_\infty, \quad \mathbb{P}\text{-a.s.}$$

PROOF. ***Proof of* (i) *and* (ii)**. Let $\tilde{\chi} \in \mathcal{M}\langle Y_{-m}^k\rangle(\chi)$. Recall the notation $Z_i = Y_{ir}^{(i+1)r-1}$ and consider the decompositions

$$\chi\mathbf{L}\langle Y_{-m}^k\rangle \mathbb{1}_\mathsf{X} = \chi\mathbf{L}\langle Y_{-m}^{-\lfloor m/r\rfloor r-1}\rangle \mathbf{L}\langle Z_{-\lfloor m/r\rfloor}^{\lfloor k/r\rfloor-1}\rangle \mathbf{L}\langle Y_{\lfloor k/r\rfloor r}^k\rangle \mathbb{1}_\mathsf{X}$$
$$\chi\mathbf{L}\langle Y_{-m}^{k-1}\rangle \mathbb{1}_\mathsf{X} = \chi\mathbf{L}\langle Y_{-m}^{-\lfloor m/r\rfloor r-1}\rangle \mathbf{L}\langle Z_{-\lfloor m/r\rfloor}^{\lfloor k/r\rfloor-1}\rangle \mathbf{L}\langle Y_{\lfloor k/r\rfloor r}^{k-1}\rangle \mathbb{1}_\mathsf{X},$$

where we make use of the convention (7) if necessary.

Choose γ such that $2/3 < \gamma < \mathbb{P}(Z_0 \in \mathsf{K})$, 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

$$\Omega_{k,m} \triangleq \left\{\left(\left\lfloor\frac{k}{r}\right\rfloor + \left\lfloor\frac{m}{r}\right\rfloor\right)^{-1} \sum_{\ell=-\lfloor m/r\rfloor}^{\lfloor k/r\rfloor-1} \mathbb{1}_\mathsf{K}(Z_\ell) \geq \gamma\right\}.$$

By Lemma 9 (Eq. (40)) it holds for all $\eta \in]0,1[$, on the event $\Omega_{k,m}$,

$$(48)\qquad (1-\rho_\gamma(\eta))\left(\ln\left(\frac{\tilde{\chi}\mathbf{L}\langle Y_{-m}^k\rangle \mathbb{1}_\mathsf{X}}{\tilde{\chi}\mathbf{L}\langle Y_{-m}^{k-1}\rangle \mathbb{1}_\mathsf{X}}\right) - \ln\left(\frac{\chi'\mathbf{L}\langle Y_{-m}^k\rangle \mathbb{1}_\mathsf{X}}{\chi'\mathbf{L}\langle Y_{-m}^{k-1}\rangle \mathbb{1}_\mathsf{X}}\right)\right)$$
$$\overset{(a)}{\leq} 2\rho_\gamma^{b_{k,m}}(\eta) + \frac{C_\gamma(\eta)\eta^{b_{k,m}}\|g(\cdot,Y_k)\|_\infty \prod_{i=-m}^{k-1}\|g(\cdot,Y_i)\|_\infty^2}{\tilde{\chi}\mathbf{L}\langle Y_{-m}^{k-1}\rangle \mathbb{1}_\mathsf{X} \times \chi'\mathbf{L}\langle Y_{-m}^k\rangle \mathbb{1}_\mathsf{X}}$$
$$\overset{(b)}{\leq} 2\rho_\gamma^{b_{k,m}}(\eta) + \frac{2C_\gamma(\eta)\eta^{b_{k,m}}\prod_{i=-m}^{k}\|g(\cdot,Y_i)\|_\infty^2}{\chi\mathbf{L}\langle Y_{-m}^k\rangle \mathbb{1}_\mathsf{X} \times \chi'\mathbf{L}\langle Y_{-m}^k\rangle \mathbb{1}_\mathsf{X}},$$

where

(a) follows from (40) and the bound $\delta_x \mathbf{L}\langle Y_u^v \rangle \mathbb{1}_{\mathsf{X}} \leq \prod_{\ell=u}^{v} \|g(\cdot, Y_\ell)\|_\infty$, valid for $u \leq v$, and

(b) follows from the fact that $\tilde{\chi} \in \mathcal{M}\langle Y_{-m}^k \rangle(\chi)$.

Since, under (**A1**)(i), the sequence $(Z_n)_{n\in\mathbb{Z}}$ is ergodic and $\mathbb{P}(Z_0 \in \mathsf{K}) > \gamma$, Lemma 13 implies that

$$\mathbb{P}\left(\bigcup_{j\geq 0} \bigcap_{\substack{(k,m)\in\mathbb{N}^*\times\mathbb{N} \\ k+m\geq j}} \Omega_{k,m} \right) = 1.$$

Hence, there exists a $\mathbb{P}$-a.s. finite integer-valued random variable U such that (48) is satisfied for all $(k,m) \in \mathbb{N}^* \times \mathbb{N}$ such that $k + m \geq U$.

The lower bound obtained in Lemma 10 implies that there exists a constant $\kappa > 0$ such that for all probability measures χ and χ' in $\mathcal{M}(\mathsf{D}, r)$ and all $(k,m) \in \mathbb{N}^* \times \mathbb{N}$, $\mathbb{P}$-a.s.,

$$\chi \mathbf{L}\langle Y_{-m}^k \rangle \mathbb{1}_{\mathsf{X}} \geq \bar{C}_{\chi,\chi'} \kappa^{-(k+m+1)},$$
$$\chi' \mathbf{L}\langle Y_{-m}^k \rangle \mathbb{1}_{\mathsf{X}} \geq \bar{C}_{\chi,\chi'} \kappa^{-(k+m+1)},$$

where $\bar{C}_{\chi,\chi'}$ is a $\mathbb{P}$-a.s. finite constant.

By plugging these bounds into (48) and using Lemma 14 with η sufficiently small (note that (48) 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.,

$$\ln\left(\frac{\tilde{\chi}\mathbf{L}\langle Y_{-m}^k \rangle \mathbb{1}_{\mathsf{X}}}{\tilde{\chi}\mathbf{L}\langle Y_{-m}^{k-1} \rangle \mathbb{1}_{\mathsf{X}}} \right) - \ln\left(\frac{\chi'\mathbf{L}\langle Y_{-m}^k \rangle \mathbb{1}_{\mathsf{X}}}{\chi'\mathbf{L}\langle Y_{-m}^{k-1} \rangle \mathbb{1}_{\mathsf{X}}} \right) \leq C_{\chi,\chi'} \beta^{k+m},$$

which completes the proof of (i). Note that $\chi \in \mathcal{M}\langle Y_{-m}^k \rangle(\chi)$ implies that the previous relation is satisfied with $\tilde{\chi} = \chi$.

The proof of (ii) follows the same lines as the proof of (i) and is omitted for brevity. ◀

Proof of (iii). As in the proof of (i), write

$$\chi \mathbf{L}\langle Y_{-m}^{-1} \rangle h = \chi \mathbf{L}\langle Y_{-m}^{-\lfloor m/r \rfloor r - 1} \rangle \mathbf{L}\langle Z_{-\lfloor m/r \rfloor}^{-1} \rangle h$$

and define the event

$$\Omega_m \triangleq \left\{ \left\lfloor \frac{m}{r} \right\rfloor^{-1} \sum_{\ell=-\lfloor m/r \rfloor}^{-1} \mathbb{1}_{\mathsf{K}}(Z_\ell) \geq \gamma \right\}.$$

By Lemma 9 (Eq. 41) it holds, on the event Ω_m,

$$\left| \frac{\chi \mathbf{L}\langle Y_{-m}^{-1}\rangle h}{\chi \mathbf{L}\langle Y_{-m}^{-1}\rangle \mathbb{1}_{\mathsf{X}}} - \frac{\chi' \mathbf{L}\langle Y_{-m}^{-1}\rangle h}{\chi' \mathbf{L}\langle Y_{-m}^{-1}\rangle \mathbb{1}_{\mathsf{X}}} \right| \leq 2 \|h\|_\infty \rho_\gamma^{\lfloor m/r \rfloor}(\eta) + \frac{C_\gamma(\eta)\eta^{\lfloor m/r \rfloor} \|h\|_\infty \prod_{i=-m}^{-1} \|g(\cdot, Y_i)\|_\infty^2}{\chi \mathbf{L}\langle Y_{-m}^{-1}\rangle \mathbb{1}_{\mathsf{X}} \times \chi' \mathbf{L}\langle Y_{-m}^{-1}\rangle \mathbb{1}_{\mathsf{X}}}, \tag{49}$$

where we used that for $u \leq v$, $\delta_x \mathbf{L}\langle Y_u^v\rangle \mathbb{1}_{\mathsf{X}} \leq \prod_{\ell=u}^{v} \|g(\cdot, Y_\ell)\|_\infty$. Under (**A1**) (i), Birkhoff's ergodic theorem 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 (49) is satisfied for $m \geq U$. Then, for $m \geq U$,

$$\begin{aligned} \frac{|\Delta_{\chi,\chi'}\langle Y_{-m}^{-1}\rangle (h, \mathbb{1}_{\mathsf{X}})|}{\|\mathbf{L}\langle Y_{-m}^{-1}\rangle \mathbb{1}_{\mathsf{X}}\|^2} &= \frac{\chi \mathbf{L}\langle Y_{-m}^{-1}\rangle \mathbb{1}_{\mathsf{X}} \times \chi' \mathbf{L}\langle Y_{-m}^{-1}\rangle \mathbb{1}_{\mathsf{X}}}{\|\mathbf{L}\langle Y_{-m}^{-1}\rangle \mathbb{1}_{\mathsf{X}}\|^2} \left| \frac{\chi \mathbf{L}\langle Y_{-m}^{-1}\rangle h}{\chi \mathbf{L}\langle Y_{-m}^{-1}\rangle \mathbb{1}_{\mathsf{X}}} - \frac{\chi' \mathbf{L}\langle Y_{-m}^{-1}\rangle h}{\chi' \mathbf{L}\langle Y_{-m}^{-1}\rangle \mathbb{1}_{\mathsf{X}}} \right| \\ &\leq 2 \|h\|_\infty \rho_\gamma^{\lfloor m/r \rfloor}(\eta) + \frac{C_\gamma(\eta)\eta^{\lfloor m/r \rfloor} \|h\|_\infty \prod_{i=-m}^{-1} \|g(\cdot, Y_i)\|_\infty^2}{\|\mathbf{L}\langle Y_{-m}^{-1}\rangle \mathbb{1}_{\mathsf{X}}\|^2}, \end{aligned} \tag{50}$$

we have used that $\chi \mathbf{L}\langle Y_{-m}^{-1}\rangle \mathbb{1}_{\mathsf{X}} \leq \|\mathbf{L}\langle Y_{-m}^{-1}\rangle \mathbb{1}_{\mathsf{X}}\|_\infty$. By Lemma 10, Eq. (43), there exist a constant $\kappa > 0$ and a $\mathbb{P}$-a.s. finite random variable $\bar{C}$ such that

$$\|\mathbf{L}\langle Y_{-m}^{-1}\rangle \mathbb{1}_{\mathsf{X}}\|_\infty \geq \bar{C}\kappa^{-m}, \quad \mathbb{P}\text{-a.s.}$$

Finally, we complete the proof by inserting this bound into (50) and applying Lemma 14 to the right hand side of the resulting inequality.

◀
□

5.2. *Convergence of the log-likelihood.*

LEMMA 12. *Assume (**A1–2**). Then, $\mathbb{P}$-a.s.,*

$$\lim_{n\to\infty} n^{-1} \ln \|\mathbf{L}\langle Y_0^n\rangle \mathbb{1}_{\mathsf{X}}\|_\infty = \ell_\infty, \tag{51}$$

$$\lim_{n\to\infty} n^{-1} \ln \|\mathbf{L}\langle Y_{-n}^0\rangle \mathbb{1}_{\mathsf{X}}\|_\infty = \ell_\infty, \tag{52}$$

$$\lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} \ln \pi\langle Y_{-\infty}^{-k-1}\rangle (Y_{-k}) = \ell_\infty, \tag{53}$$

where ℓ_∞ is defined in (27).

PROOF. ***Proof of*** (51). Let $(\alpha_n)_{n\in\mathbb{N}^*}$ be a non-decreasing 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 $\tilde{x}_n \in \mathsf{X}$ such that

$$\alpha_n \left\| \mathbf{L}\langle Y_0^n\rangle \mathbb{1}_\mathsf{X}\right\|_\infty \leq \delta_{\tilde{x}_n}\mathbf{L}\langle Y_0^n\rangle \mathbb{1}_\mathsf{X} \leq \left\| \mathbf{L}\langle Y_0^n\rangle \mathbb{1}_\mathsf{X}\right\|_\infty . \tag{54}$$

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

$$\delta_{\tilde{x}_{k-1}}\mathbf{L}\langle Y_0^{k-1}\rangle \mathbb{1}_\mathsf{X} \geq \alpha_{k-1}\|\mathbf{L}\langle Y_0^{k-1}\rangle \mathbb{1}_\mathsf{X}\|_\infty \geq \alpha_{k-1}\delta_{\tilde{x}_k}\mathbf{L}\langle Y_0^{k-1}\rangle \mathbb{1}_\mathsf{X}. \tag{55}$$

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

$$\delta_{\tilde{x}_k}\mathbf{L}\langle Y_0^{k-1}\rangle \mathbb{1}_\mathsf{X} \overset{(a)}{\geq} \frac{\delta_{\tilde{x}_k}\mathbf{L}\langle Y_0^{k}\rangle \mathbb{1}_\mathsf{X}}{\|g(\cdot,Y_k)\|_\infty} \overset{(b)}{\geq} \alpha_k\frac{\|\mathbf{L}\langle Y_0^{k}\rangle \mathbb{1}_\mathsf{X}\|_\infty}{\|g(\cdot,Y_k)\|_\infty} \geq \alpha_k\frac{\chi\mathbf{L}\langle Y_0^{k}\rangle \mathbb{1}_\mathsf{X}}{\|g(\cdot,Y_k)\|_\infty}, \tag{56}$$

where (a) follows from the bound $\delta_{\tilde{x}_k}\mathbf{L}\langle Y_0^{k}\rangle \mathbb{1}_\mathsf{X} \leq \|g(\cdot,Y_k)\|_\infty\, \delta_{\tilde{x}_k}\mathbf{L}\langle Y_0^{k-1}\rangle \mathbb{1}_\mathsf{X}$ and (b) stems from the definition (54) of α_n. Then,

$$\begin{aligned}
0 &\leq n^{-1}\left(\ln\|\mathbf{L}\langle Y_0^n\rangle \mathbb{1}_\mathsf{X}\|_\infty - \ln \chi\mathbf{L}\langle Y_0^n\rangle \mathbb{1}_\mathsf{X}\right)\\
&\leq -n^{-1}\ln\alpha_n + n^{-1}\left(\ln\left(\alpha_n\|\mathbf{L}\langle Y_0^n\rangle \mathbb{1}_\mathsf{X}\|_\infty\right) - \ln \chi\mathbf{L}\langle Y_0^n\rangle \mathbb{1}_\mathsf{X}\right)\\
&\leq -n^{-1}\ln\alpha_n + n^{-1}\left(\ln\delta_{\tilde{x}_n}\mathbf{L}\langle Y_0^n\rangle \mathbb{1}_\mathsf{X} - \ln \chi\mathbf{L}\langle Y_0^n\rangle \mathbb{1}_\mathsf{X}\right)\\
&= -n^{-1}\ln\alpha_n + n^{-1}\left(\ln\delta_{\tilde{x}_0}\mathbf{L}\langle Y_0\rangle \mathbb{1}_\mathsf{X} - \ln \chi\mathbf{L}\langle Y_0\rangle \mathbb{1}_\mathsf{X}\right)\\
&\quad + n^{-1}\sum_{k=1}^{n}\left[\ln\left(\frac{\delta_{\tilde{x}_k}\mathbf{L}\langle Y_0^k\rangle \mathbb{1}_\mathsf{X}}{\delta_{\tilde{x}_{k-1}}\mathbf{L}\langle Y_0^{k-1}\rangle \mathbb{1}_\mathsf{X}}\right) - \ln\left(\frac{\chi\mathbf{L}\langle Y_0^k\rangle \mathbb{1}_\mathsf{X}}{\chi\mathbf{L}\langle Y_0^{k-1}\rangle \mathbb{1}_\mathsf{X}}\right)\right].
\end{aligned} \tag{57}$$

For each term in the sum it holds, by (55),

$$\begin{aligned}
&\ln\left(\frac{\delta_{\tilde{x}_k}\mathbf{L}\langle Y_0^k\rangle \mathbb{1}_\mathsf{X}}{\delta_{\tilde{x}_{k-1}}\mathbf{L}\langle Y_0^{k-1}\rangle \mathbb{1}_\mathsf{X}}\right) - \ln\left(\frac{\chi\mathbf{L}\langle Y_0^k\rangle \mathbb{1}_\mathsf{X}}{\chi\mathbf{L}\langle Y_0^{k-1}\rangle \mathbb{1}_\mathsf{X}}\right)\\
&\quad\leq -\ln\alpha_{k-1} + \ln\left(\frac{\delta_{\tilde{x}_k}\mathbf{L}\langle Y_0^k\rangle \mathbb{1}_\mathsf{X}}{\delta_{\tilde{x}_k}\mathbf{L}\langle Y_0^{k-1}\rangle \mathbb{1}_\mathsf{X}}\right) - \ln\left(\frac{\chi\mathbf{L}\langle Y_0^k\rangle \mathbb{1}_\mathsf{X}}{\chi\mathbf{L}\langle Y_0^{k-1}\rangle \mathbb{1}_\mathsf{X}}\right).
\end{aligned}$$

For all $k\in\mathbb{N}^*$, (56) implies that

$$\delta_{\tilde{x}_k}\mathbf{L}\langle Y_0^{k-1}\rangle \mathbb{1}_\mathsf{X} \geq \frac{1}{2}\frac{\chi\mathbf{L}\langle Y_0^k\rangle \mathbb{1}_\mathsf{X}}{\|g(\cdot,Y_k)\|_\infty},$$

so that $\delta_{\tilde{x}_k}$ belongs to the set $\mathcal{M}\langle Y_0^{k-1}\rangle(\chi)$ (defined in (47)). Proposition 11(i) then provides a constant $\beta\in\,]0,1[$ and a $\mathbb{P}$-a.s. finite random variable C_χ such that

$$\ln\left(\frac{\delta_{\tilde{x}_k}\mathbf{L}\langle Y_0^k\rangle \mathbb{1}_\mathsf{X}}{\delta_{\tilde{x}_k}\mathbf{L}\langle Y_0^{k-1}\rangle \mathbb{1}_\mathsf{X}}\right) - \ln\left(\frac{\chi\mathbf{L}\langle Y_0^k\rangle \mathbb{1}_\mathsf{X}}{\chi\mathbf{L}\langle Y_0^{k-1}\rangle \mathbb{1}_\mathsf{X}}\right) \leq C_\chi\beta^k. \tag{58}$$

Finally, the statement (51) follows by plugging the bound (58) into (57), letting n tend to infinity, and using (26). ◀

Proof of (52). For all $(p,n) \in \mathbb{N}^2$ such that $p \leq n$, define $W_{p,n} \triangleq \ln \|\mathbf{L}\langle Y_p^{n-1}\rangle \mathbb{1}_{\mathsf{X}}\|_\infty$ and $\tilde{W}_{p,n} \triangleq \ln \|\mathbf{L}\langle Y_{-n+1}^{-p}\rangle \mathbb{1}_{\mathsf{X}}\|_\infty$. 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 $x \in \mathsf{D}$, $m \in \mathbb{N}$, and $y_0^{mr-1} \in \mathsf{Y}^{mr}$, it holds that

$$\|\mathbf{L}\langle y_0^{mr-1}\rangle \mathbb{1}_{\mathsf{X}}\|_\infty \geq \delta_x \mathbf{L}\langle y_0^{mr-1}\rangle \mathbb{1}_{\mathsf{X}} \geq \prod_{\ell=0}^{m-1} \inf_{x\in\mathsf{D}} \delta_x \mathbf{L}\langle y_{kr}^{(k+1)r-1}\rangle \mathbb{1}_{\mathsf{D}}. \tag{59}$$

Using the stationarity of the observation process $(Y_k)_{k\in\mathbb{Z}}$, we get, via Assumption (**A1**)(iii), for all $m \in \mathbb{N}^*$,

(60)
$$(mr)^{-1}\mathbb{E}\left(W_{0,mr}\right) = (mr)^{-1}\mathbb{E}\left(\tilde{W}_{0,mr}\right) \geq (mr)^{-1}\mathbb{E}\left(\ln \|\mathbf{L}\langle y_0^{mr-1}\rangle \mathbb{1}_{\mathsf{X}}\|_\infty\right)$$
$$\geq r^{-1}\mathbb{E}\left(\ln \inf_{x\in\mathsf{D}} \delta_x \mathbf{L}\langle y_{kr}^{(k+1)r-1}\rangle \mathbb{1}_{\mathsf{D}}\right) > -\infty.$$

The sequences $(\mathbb{E}\left(W_{0,n}\right))_{n\in\mathbb{N}^*}$ and $(\mathbb{E}(\tilde{W}_{0,n}))_{n\in\mathbb{N}^*}$ are subadditive; Fekete's lemma thus implies that the sequences $(n^{-1}\mathbb{E}\left(W_{0,n}\right))_{n\in\mathbb{N}^*}$ and $(n^{-1}\mathbb{E}(\tilde{W}_{0,n}))_{n\in\mathbb{N}^*}$ have limits in $[-\infty, \infty[$ and that

$$\lim_{n\to\infty} n^{-1}\mathbb{E}\left(W_{0,n}\right) = \lim_{n\to\infty} n^{-1}\mathbb{E}\left(\tilde{W}_{0,n}\right)$$
$$= \inf_{n\in\mathbb{N}^*} n^{-1}\mathbb{E}\left(W_{0,n}\right) = \inf_{n\in\mathbb{N}^*} n^{-1}\mathbb{E}\left(\tilde{W}_{0,n}\right).$$

However, by (60) there exists a subsequence that is bounded away from $-\infty$, showing that

$$\inf_{n\in\mathbb{N}^*} n^{-1}\mathbb{E}\left(W_{0,n}\right) = \lim_{n\to\infty} n^{-1}\mathbb{E}\left(W_{0,n}\right) > -\infty,$$
$$\inf_{n\in\mathbb{N}^*} n^{-1}\mathbb{E}\left(\tilde{W}_{0,n}\right) = \lim_{n\to\infty} n^{-1}\mathbb{E}\left(\tilde{W}_{0,n}\right) > -\infty.$$

Now, by applying Kingman's subadditive ergodic theorem and using again that $\mathbb{E}(\tilde{W}_{0,k}) = \mathbb{E}(W_{0,k})$ under stationarity, we obtain

$$\lim_{n\to\infty} n^{-1}\tilde{W}_{0,n} = \inf_{n\in\mathbb{N}^*} n^{-1}\mathbb{E}\left(\tilde{W}_{0,n}\right) = \inf_{n\in\mathbb{N}^*} n^{-1}\mathbb{E}\left(W_{0,n}\right)$$
$$= \lim_{n\to\infty} n^{-1}W_{0,n} = \ell_\infty, \quad \mathbb{P}\text{-a.s.},$$

where the last limit follows from (51). This completes the proof of statement (52). ◀

Proof of (53). Since $\mathbb{E}(|\ln \pi\langle Y_{-\infty}^{-1}\rangle(Y_0)|) < \infty$ and the process $(Y_k)_{k\in\mathbb{Z}}$ is stationary and ergodic, (53) follows from Birkhoff's ergodic theorem. ◀

□

APPENDIX A: TECHNICAL LEMMAS

LEMMA 13. *If $(U_n)_{n\in\mathbb{Z}}$ is a stationary and ergodic sequence of random variables such that $\mathbb{E}(|U_0|) < \infty$, then*

$$\lim_{k+m\to\infty} (k+m)^{-1}\left(\sum_{\ell=-m}^{k-1} U_\ell\right) = \mathbb{E}(U_0)\ , \quad \mathbb{P}\text{-}a.s. \tag{61}$$

PROOF. Denote

$$\Omega_1 \triangleq \left\{\omega\in\Omega;\ \lim_{k+m\to\infty}(k+m)^{-1}\left(\sum_{\ell=-m}^{k-1} U_\ell(\omega)\right) = \mathbb{E}(U_0)\right\},$$

$$\Omega_2 \triangleq \left\{\omega\in\Omega;\ \lim_{m\to\infty}\frac{\sum_{\ell=-m}^{-1} U_\ell(\omega)}{m} = \lim_{k\to\infty}\frac{\sum_{\ell=0}^{k-1} U_\ell(\omega)}{k} = \mathbb{E}(U_0)\right\}.$$

By Birkhoff's ergodic theorem, $\mathbb{P}(\Omega_2) = 1$. To obtain (61), it is thus sufficient to show that $\Omega_1^c \cap \Omega_2 = \emptyset$. The proof is by contradiction. Assume $\Omega_1^c\cap\Omega_2 \neq \emptyset$, so that there exists $\omega \in \Omega_1^c \cap \Omega_2$. For such ω, 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 $n \geq 0$,

$$\left|\frac{\sum_{\ell=-m_n(\omega)}^{k_n(\omega)-1} U_\ell(\omega)}{k_n(\omega)+m_n(\omega)} - \mathbb{E}(U_0)\right| \geq \epsilon(\omega). \tag{62}$$

Consider the following decomposition:

$$\frac{\sum_{\ell=-m_n(\omega)}^{k_n(\omega)-1} U_\ell(\omega)}{k_n(\omega)+m_n(\omega)} = \frac{m_n(\omega)}{k_n(\omega)+m_n(\omega)}\frac{\sum_{\ell=-m_n(\omega)}^{-1} U_\ell(\omega)}{m_n(\omega)} + \frac{k_n(\omega)}{k_n(\omega)+m_n(\omega)}\frac{\sum_{\ell=0}^{k_n(\omega)-1} U_\ell(\omega)}{k_n(\omega)}. \tag{63}$$

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

$$\lim_{n\to\infty}\frac{m_n(\omega)}{k_n(\omega)+m_n(\omega)} = 1\,,\quad \lim_{n\to\infty}\frac{k_n(\omega)}{k_n(\omega)+m_n(\omega)} = 0, \tag{64}$$

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

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

which implies, together with (64), that

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

This contradicts (62). 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

$$\lim_{n\to\infty} \frac{\sum_{\ell=-m_n(\omega)}^{-1} U_\ell(\omega)}{m_n(\omega)} = \lim_{n\to\infty} \frac{\sum_{\ell=0}^{k_n(\omega)-1} U_\ell(\omega)}{k_n(\omega)} = \mathbb{E}\left(U_0\right).$$

Combining this with (63), we obtain that

$$\lim_{n\to\infty} \frac{\sum_{\ell=-m_n(\omega)}^{k_n(\omega)-1} U_\ell(\omega)}{k_n(\omega)+m_n(\omega)} = \mathbb{E}\left(U_0\right),$$

which again contradicts (62). Finally, $\Omega_1^c \cap \Omega_2 = \emptyset$, and since $\mathbb{P}\left(\Omega_2\right) = 1$, we finally obtain that $\mathbb{P}\left(\Omega_1\right) = 1$. The proof is completed. □

LEMMA 14. *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*

$$\mathbb{E}\left(\ln^+ U_0\right) < \infty, \quad \mathbb{E}\left(\ln^+ V_0\right) < \infty, \quad \mathbb{E}\left(\ln^+ W_0\right) < \infty.$$

Then for all η and ρ 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.,

$$\rho^{k+m} + \eta^{k+m} W_{-m} \left(\prod_{\ell=-m}^{k-1} V_\ell\right) U_k \le C\beta^{k+m}.$$

PROOF. See [13, Lemma 6]. □

REFERENCES

[1] A. Bain and D. Crisan. *Fundamentals of stochastic filtering*, volume 60 of *Stochastic Modelling and Applied Probability*. Springer, New York, 2009.

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

[3] N. Chopin. A sequential particle filter method for static models. *Biometrika*, 89:539–552, 2002.

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

[5] P. Del Moral. *Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications*. Springer, 2004.

[6] P. Del Moral and A. Guionnet. Central limit theorem for nonlinear filtering and interacting particle systems. *Ann. Appl. Probab.*, 9(2):275–297, 1999.

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

[8] P. Del Moral and L. Miclo. Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering. In *Séminaire de Probabilités, XXXIV*, volume 1729 of *Lecture Notes in Math.*, pages 1–145. Springer, Berlin, 2000.

[9] R. Douc, G. Fort, E. Moulines, and P. Priouret. Forgetting the initial distribution for hidden markov models. *Stochastic Processes and their Applications*, (4):1235–1256, 2009.

[10] R. Douc, A. Garivier, E. Moulines, and J. Olsson. On the forward filtering backward smoothing particle approximations of the smoothing distribution in general state spaces models. *Ann. Appl. Probab.*, 21(6):1201–2145, 2011.

[11] R. Douc, A. Guillin, and J. Najim. Moderate deviations for particle filtering. *Ann. Appl. Probab.*, 15(1B):587–614, 2005.

[12] R. Douc and E. Moulines. Limit theorems for weighted samples with applications to sequential Monte Carlo methods. *Ann. Statist.*, 36(5):2344–2376, 2008.

[13] R. Douc and E. Moulines. Asymptotic properties of the maximum likelihood estimation in misspecified hidden Markov models. *Preprint*, (Arxiv: 1110.0356), 2011.

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

[15] A. Doucet, N. De Freitas, and N. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer, New York, 2001.

[16] A. Doucet and A. M. Johansen. A note on auxiliary particle filters. *Statist. Probab. Lett.*, 78:1498–1504, 2008.

[17] S. J. Godsill, A. Doucet, and M. West. Monte carlo smoothing for non-linear time series. *J. Am. Statist. Assoc.*, 50:438–449, 2004.

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

[19] K. Heine and D. Crisan. Uniform approximations of discrete-time filters. *Adv. in Appl. Probab.*, 40(4):979–1001, 2008.

[20] M. L. Kleptsyna and A. Yu. Veretennikov. On discrete time ergodic filters with wrong initial data. *Probab. Theory Related Fields*, 141(3-4):411–444, 2008.

[21] H. R. Künsch. Recursive Monte-Carlo filters: algorithms and theoretical analysis. *Ann. Statist.*, 33(5):1983–2021, 2005.

[22] F. Le Gland and N. Oudjane. Stability and uniform approximation of nonlinear

filters using the hilbert metric and application to particle filters. *Ann. Appl. Probab.*, 14:144–187, 2004.

[23] François LeGland and Nadia Oudjane. A robustification approach to stability and to uniform particle approximation of nonlinear filters: the example of pseudo-mixing signals. *Stochastic Process. Appl.*, 106(2):279–316, 2003.

[24] J. Olsson and T. Rydén. Asymptotic properties of the bootstrap particle filter maximum likelihood estimator for state space models. *Stoch. Proc. Appl.*, 118:649–680, 2008.

[25] Nadia Oudjane and Sylvain Rubenthaler. Stability and uniform particle approximation of nonlinear filters in case of non ergodic signals. *Stoch. Anal. Appl.*, 23(3):421–448, 2005.

[26] M. K. Pitt and N. Shephard. Filtering via simulation: Auxiliary particle filters. *J. Am. Statist. Assoc.*, 94(446):590–599, 1999.

[27] R. J. Serfling. *Approximation Theorems of Mathematical Statistics*. Wiley, New York, 1980.

[28] A. N. Shiryaev. *Probability*. Springer, 2nd edition, 1996.

[29] V. B. Tadić and A. Doucet. Exponential forgetting and geometric ergodicity for optimal filtering in general state-space models. *Stochastic Process. Appl.*, 115(8):1408–1436, 2005.

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

Preprints in Mathematical Sciences 2012:1
ISSN 1403-9338

LUNFMS-5050-2012

Mathematical Statistics
Centre for Mathematical Sciences
Lund University
Box 118, SE-221 00 Lund, Sweden

http://www.maths.lth.se/