

TWISTED PARTICLE FILTERS

BY NICK WHITELEY¹ AND ANTHONY LEE²

University of Bristol and University of Warwick

We investigate sampling laws for particle algorithms and the influence of these laws on the efficiency of particle approximations of marginal likelihoods in hidden Markov models. Among a broad class of candidates we characterize the essentially unique family of particle system transition kernels which is optimal with respect to an asymptotic-in-time variance growth rate criterion. The sampling structure of the algorithm defined by these optimal transitions turns out to be only subtly different from standard algorithms and yet the fluctuation properties of the estimates it provides can be dramatically different. The structure of the optimal transition suggests a new class of algorithms, which we term “twisted” particle filters and which we validate with asymptotic analysis of a more traditional nature, in the regime where the number of particles tends to infinity.

1. Introduction. A hidden Markov model (HMM) with measurable state space (X, \mathcal{X}) and observation space (Y, \mathcal{Y}) is a process $\{(X_n, Y_n); n \geq 0\}$ where $\{X_n; n \geq 0\}$ is a Markov chain on X , and each observation Y_n , valued in Y , is conditionally independent of the rest of the process given X_n . Let μ_0 and f be respectively a probability distribution and a Markov kernel on (X, \mathcal{X}) , and let g be a Markov kernel acting from (X, \mathcal{X}) to (Y, \mathcal{Y}) , with $g(x, \cdot)$ admitting a strictly positive density, denoted similarly by $g(x, y)$, with respect to some dominating σ -finite measure. The HMM specified by μ_0 , f and g is

$$(1.1) \quad \begin{aligned} X_0 &\sim \mu_0(\cdot), & X_n | \{X_{n-1} = x_{n-1}\} &\sim f(x_{n-1}, \cdot), & n \geq 1, \\ Y_n | \{X_n = x_n\} &\sim g(x_n, \cdot), & n \geq 0. \end{aligned}$$

In practice, one often seeks to fit a HMM to data $\{Y_0, Y_1, \dots\}$. This motivates computation of the marginal likelihood of $\{Y_0, Y_1, \dots\}$ under the model (1.1). We consider methods for approximate performance of this computation.

Let $\Omega := Y^{\mathbb{Z}}$ be the set of doubly infinite sequences valued in Y . For $\omega = \{\omega(n)\}_{n \in \mathbb{Z}} \in \Omega$ we shall write the coordinate projection $Y_n(\omega) = \omega(n)$ and take as a recursive definition of the *prediction filters*, the sequence of distributions

Received October 2012; revised May 2013.

¹Supported by EPSRC Grant EP/K023330/1.

²Supported by CRiSM, an EPSRC-HEFCE UK grant.

MSC2010 subject classifications. Primary 60K35, 62M20; secondary 60G35.

Key words and phrases. Sequential Monte Carlo, filtering.

$\{\pi_n^\omega; n \geq 0\}$ given by

$$(1.2) \quad \begin{aligned} \pi_0^\omega &:= \mu_0, \\ \pi_n^\omega(A) &:= \frac{\int_{\mathcal{X}} \pi_{n-1}^\omega(dx) g(x, Y_{n-1}(\omega)) f(x, A)}{\int_{\mathcal{X}} \pi_{n-1}^\omega(dx) g(x, Y_{n-1}(\omega))}, \quad A \in \mathcal{X}, n \geq 1. \end{aligned}$$

We are centrally concerned with the sequence $\{Z_n^\omega; n \geq 0\}$ defined by

$$(1.3) \quad Z_0^\omega := 1, \quad Z_n^\omega := Z_{n-1}^\omega \int_{\mathcal{X}} \pi_{n-1}^\omega(dx) g(x, Y_{n-1}(\omega)), \quad n \geq 1.$$

Due to the conditional independence structure of the HMM, π_n^ω is the conditional distribution of X_n given $Y_{0:n-1}(\omega)$; and Z_n^ω is the marginal likelihood evaluated at the point $Y_{0:n-1}(\omega)$. The simplest particle filter, known as the ‘‘bootstrap’’ algorithm [19], is given below. It yields an approximation, $Z_n^{\omega, N}$, of each Z_n^ω .

Convergence properties of particle algorithms in the regime $N \rightarrow \infty$ are well understood [7, 9, 13, 22] and their stability properties have been expressed through finite- N error bounds [6, 10, 29], time-uniform convergence [23, 24, 26, 28] and control on $N \rightarrow \infty$ asymptotic variance expressions [11, 16, 18, 29]. Our aim is to rigorously address comparative questions of how and why one algorithm may outperform another, and how it is possible to modify standard algorithms in order to improve performance. Our study is formulated in a generic framework which accommodates standard particle algorithms and novel extensions. As an introduction we discuss some of our intentions and findings in the context of the bootstrap particle filter as per Algorithm 1; more precise statements are given later.

Writing E_N^ω for expectation with respect to the law of the bootstrap particle filter processing a fixed observation sequence $\omega \in \Omega$, the well-known lack-of-bias property ([8], Proposition 7.4.1) reads

$$(1.4) \quad E_N^\omega[Z_n^{\omega, N}] = Z_n^\omega,$$

and holds for any $n \geq 0$ and $N \geq 1$. This property is desirable because it allows particle filters to be used within ‘‘pseudo-marginal’’-type algorithms (see [2] and references therein), and plays a role in explaining the validity of some compound

Algorithm 1 Bootstrap particle filter

For $n = 0$,

Sample $(\zeta_0^i)_{i=1}^N \stackrel{\text{i.i.d.}}{\sim} \mu_0$,

Report $Z_0^{\omega, N} = 1$.

For $n \geq 1$,

Report $Z_n^{\omega, N} = Z_{n-1}^{\omega, N} \cdot \frac{1}{N} \sum_{j=1}^N g(\zeta_{n-1}^j, Y_{n-1}(\omega))$,

Sample $(\zeta_n^i)_{i=1}^N | (\zeta_{n-1}^i)_{i=1}^N \stackrel{\text{i.i.d.}}{\sim} \frac{\sum_{j=1}^N g(\zeta_{n-1}^j, Y_{n-1}(\omega)) f(\zeta_{n-1}^j, \cdot)}{\sum_{j=1}^N g(\zeta_{n-1}^j, Y_{n-1}(\omega))}$.

Monte Carlo techniques such as particle Markov chain Monte Carlo [1]. The accuracy of $Z_n^{\omega, N}$ influences the performance of such schemes [3]. We shall analyze novel particle algorithms arising through changes of measure on the left-hand side of (1.4), similarly enjoying lack-of-bias, and which could therefore be used in lieu of more standard particle filters in compound Monte Carlo algorithms. The resulting approximation of Z_n^ω will be of the form

$$(1.5) \quad \tilde{Z}_n^{\omega, N} := Z_n^{\omega, N} \cdot \prod_{p=1}^n \phi_p^{\omega, N},$$

where $Z_n^{\omega, N}$ is exactly the same functional of the particles as in Algorithm 1 and $\{\phi_n^{\omega, N}; n \geq 1\}$ is a sequence of functionals chosen such that, if we write $\tilde{\mathbb{E}}_N^\omega$ for expectation under the (as yet unspecified) alternative sampling law, then the lack of bias property is preserved:

$$(1.6) \quad \tilde{\mathbb{E}}_N^\omega[\tilde{Z}_n^{\omega, N}] = Z_n^\omega.$$

Our main objective is to identify “good” choices of alternative sampling laws, possibly allowing the transitions of the particles to depend on past and/or future observations. Our criterion for performance arises from a study of the normalized second moment of $\tilde{Z}_n^{\omega, N}$, in the regime where N is fixed and $n \rightarrow \infty$, in an ω -pathwise fashion.

For now, let us still consider $\omega \in \Omega$ as fixed. Then under the probability law corresponding to Algorithm 1, the generations of the particle system, ζ_0, ζ_1, \dots with $\zeta_n := (\zeta_n^1, \dots, \zeta_n^N)$, form an \mathcal{X}^N -valued time-inhomogeneous Markov chain. Let $\{\mathbf{M}^\omega; \omega \in \Omega\}$ be the family of Markov kernels such that for each $\omega \in \Omega$, $\mathbf{M}^\omega: \mathcal{X}^N \times \mathcal{X}^{\otimes N} \rightarrow [0, 1]$ is given by

$$(1.7) \quad \mathbf{M}^\omega(x, dz) = \prod_{i=1}^N \frac{\sum_{j=1}^N g(x^j, Y_0(\omega)) f(x^j, dz^i)}{\sum_{j=1}^N g(x^j, Y_0(\omega))},$$

with $x = (x^1, \dots, x^N) \in \mathcal{X}^N$ and $z = (z^1, \dots, z^N) \in \mathcal{X}^N$. Let $\theta: \Omega \rightarrow \Omega$ be the shift operator, $(\theta\omega)(n) := \omega(n + 1)$, $n \in \mathbb{Z}$, $\omega \in \Omega$, so that, for example, $Y_0(\theta\omega) = Y_1(\omega)$. The n -fold iterate of θ will be written θ^n with $\theta^0 = \text{Id}$. It is then clear that the sampling steps of Algorithm 1 implement

$$(1.8) \quad \zeta_0 \sim \mu_0^{\otimes N}, \quad \zeta_n | \zeta_{n-1} \sim \mathbf{M}^{\theta^{n-1}\omega}(\zeta_{n-1}, \cdot), \quad n \geq 1.$$

Variance growth rates. For a family of Markov kernels $\{\tilde{\mathbf{M}}^\omega; \omega \in \Omega\}$ belonging to a broad class of candidates and which may depend on ω in a rather general fashion, but subject to $\mathbf{M}^\omega(x, \cdot) \ll \tilde{\mathbf{M}}^\omega(x, \cdot)$ and other regularity conditions, we shall consider sampling the particle system according to

$$(1.9) \quad \zeta_0 \sim \mu_0^{\otimes N}, \quad \zeta_n | \zeta_{n-1} \sim \tilde{\mathbf{M}}^{\theta^{n-1}\omega}(\zeta_{n-1}, \cdot), \quad n \geq 1,$$

and simply setting

$$(1.10) \quad \phi_n^{\omega, N} := \frac{d\mathbf{M}^{\theta^{n-1}\omega}(\zeta_{n-1}, \cdot)}{d\tilde{\mathbf{M}}^{\theta^{n-1}\omega}(\zeta_{n-1}, \cdot)}(\zeta_n), \quad n \geq 1.$$

Then letting $\tilde{\mathbb{E}}_N^\omega$ denote expectation under the Markov law (1.9), and with $\tilde{Z}_n^{\omega, N}$ as in (1.5), we of course achieve (1.6).

Let Ω be endowed with the product σ -algebra $\mathcal{F} = \mathcal{Y}^{\otimes \mathbb{Z}}$ and let \mathbb{P} be a probability measure on (Ω, \mathcal{F}) . We stress that \mathbb{P} is not necessarily a measure on observation sequences derived from the particular HMM (1.1), nor indeed any HMM. Under the assumption that θ is \mathbb{P} -preserving and ergodic, and under certain other regularity conditions, application of our first main result, Proposition 4, establishes, for any fixed $N \geq 1$, existence of a deterministic constant $\Upsilon_N(\tilde{\mathbf{M}})$, depending on $\tilde{\mathbf{M}} = \{\tilde{\mathbf{M}}^\omega; \omega \in \Omega\}$ such that

$$(1.11) \quad \frac{1}{n} \log \frac{\tilde{\mathbb{E}}_N^\omega[(\tilde{Z}_n^{\omega, N})^2]}{(\tilde{Z}_n^\omega)^2} \longrightarrow \Upsilon_N(\tilde{\mathbf{M}}) \quad \text{as } n \longrightarrow \infty, \text{ for } \mathbb{P}\text{-a.a. } \omega.$$

It must be the case that $\Upsilon_N(\tilde{\mathbf{M}}) \geq 0$, because variance is nonnegative and the lack of bias property (1.6) holds. We shall see that typically $\Upsilon_N(\tilde{\mathbf{M}}) > 0$.

Optimal sampling. Our second main result (Theorem 1) identifies, for any fixed $N \geq 1$ and among the class of candidates, the essentially unique choice of the family $\{\tilde{\mathbf{M}}^\omega; \omega \in \Omega\}$ which achieves $\Upsilon_N(\tilde{\mathbf{M}}) = 0$. It turns out that this optimal choice arises from a particular form of re-weighting applied to each transition kernel \mathbf{M}^ω and is defined in terms of a family of functions $\{h^\omega : \mathcal{X} \rightarrow \mathbb{R}_+; \omega \in \Omega\}$ which are, in abstract terms, generalized eigenfunctions associated with algebraic structures underlying the particle algorithm. In the context of the bootstrap particle filter, h^ω has the following interpretation. $\pi_n^{\theta^{-n}\omega}$ is the prediction filter initialized at time $-n$ and run forward to time zero, giving a distribution over X_0 conditional on $Y_{-n}(\omega), \dots, Y_{-1}(\omega)$. Then, letting Π_n^ω be the distribution over X_0 obtained by further conditioning on $Y_0(\omega), \dots, Y_{n-1}(\omega)$, h^ω arises as the pointwise limit:

$$h^\omega(x) = \lim_{n \rightarrow \infty} \frac{d\Pi_n^\omega}{d\pi_n^{\theta^{-n}\omega}}(x).$$

Theorem 1 establishes that for any $N \geq 1$, $\Upsilon_N(\tilde{\mathbf{M}}) = 0$ if and only if, for \mathbb{P} -almost all $\omega \in \Omega$ there exists a set $A_\omega \in \mathcal{X}^{\otimes N}$ such that A_ω^c is null (with respect to an as yet unnamed measure) and for any $x \in A_\omega$,

$$(1.12) \quad \tilde{\mathbf{M}}^\omega(x, B) = \frac{\int_B \mathbf{M}^\omega(x, dx') \mathbf{h}^{\theta\omega}(x')}{\int_{\mathcal{X}^N} \mathbf{M}^\omega(x, dx') \mathbf{h}^{\theta\omega}(x')} \quad \text{for all } B \in \mathcal{X}^{\otimes N},$$

where

$$\mathbf{h}^\omega : x = (x^1, \dots, x^N) \in \mathcal{X}^N \longmapsto N^{-1} \sum_{i=1}^N h^\omega(x^i) \in \mathbb{R}_+.$$

In the rare-event and large deviations literatures, the action of re-weighting Markov kernels using nonnegative eigenfunctions is generically referred to as “twisting.” Since in the present context we are applying re-weighting to the transitions of the entire particle system, we shall adopt this terminology and consider a class of algorithms which we refer to as *twisted particle filters*.

Twisted particle filters. The form of the optimal transition (1.12), where \mathbf{M}^ω is re-weighted by an additive, nonnegative functional, leads us to consider a new class of particle algorithms. Consider a family of functions $\{\psi^\omega : \mathbf{X} \rightarrow \mathbb{R}_+; \omega \in \Omega\}$ and let $\{\tilde{\mathbf{M}}^\omega; \omega \in \Omega\}$ be defined by

$$(1.13) \quad \tilde{\mathbf{M}}^\omega(x, dx') = \frac{\mathbf{M}^\omega(x, dx') \psi^{\theta\omega}(x')}{\int_{\mathbf{X}^N} \mathbf{M}^\omega(x, dz) \psi^{\theta\omega}(z)},$$

$$(1.14) \quad \psi^\omega : x = (x^1, \dots, x^N) \in \mathbf{X}^N \mapsto N^{-1} \sum_{i=1}^N \psi^\omega(x^i) \in \mathbb{R}_+.$$

This setup clearly admits the optimal transition ($\psi^\omega = h^\omega$) and the standard transition (take $\psi^\omega = c$, for some positive constant c) as special cases. Then introducing

$$\tilde{g}^\omega(x) := g(x, Y_0(\omega)) \int_{\mathbf{X}} f(x, dz) \psi^{\theta\omega}(z),$$

we observe that $\phi_n^{\omega, N}$, defined in (1.10), is given by

$$(1.15) \quad \phi_n^{\omega, N} = \left[\frac{1}{N^{-1} \sum_{i=1}^N g(\zeta_{n-1}^i, Y_{n-1}(\omega))} \right] \frac{\sum_{i=1}^N \tilde{g}^{\theta^{n-1}\omega}(\zeta_{n-1}^i)}{\sum_{i=1}^N \psi^{\theta^n\omega}(\zeta_n^i)}.$$

Since ψ^ω is an additive functional, it is clear that $\tilde{\mathbf{M}}^\omega$ as per (1.13) is of mixture form, and introducing the ω -dependent Markov kernel

$$\tilde{f}^\omega(x, dx') := \frac{f(x, dx') \psi^{\theta\omega}(x')}{\int_{\mathbf{X}} f(x, dz) \psi^{\theta\omega}(z)},$$

the procedure of sampling from (1.9) and evaluating $\tilde{Z}_n^{\omega, N}$ can be implemented through Algorithm 2, in which K_n and A_n are auxiliary random variables employed for algorithmic purposes, and the recursion for $\tilde{Z}_n^{\omega, N}$ arises from the definition of $Z_n^{\omega, N}$ combined with (1.5) and (1.15).

The difference between the sampling steps of Algorithm 2 and Algorithm 1 is fairly subtle: loosely speaking, at each time step, $N - 1$ of the particles in Algorithm 2 are propagated by the same mechanism as in Algorithm 1. However, with an appropriate choice of ψ^ω , the fluctuation properties of $\tilde{Z}_n^{\omega, N}$ under (1.13)–(1.14) can be dramatically different to those of $Z_n^{\omega, N}$ under (1.8)–(1.7). Our third main result (Theorem 2) concerns asymptotic fluctuation properties of twisted particle approximations when n and ω are fixed and $N \rightarrow \infty$. Under mild regularity conditions, we prove central limit theorems for generic particle systems under transitions like (1.13)–(1.14). For bounded functions φ centered w.r.t. π_n^ω , we find

Algorithm 2 Twisted bootstrap particle filter

For $n = 0$,

Sample $(\zeta_0^i)_{i=1}^N \stackrel{\text{i.i.d.}}{\sim} \mu_0$,

Report $\tilde{Z}_0^{\omega, N} = 1$.

For $n \geq 1$,

Sample K_n from the uniform distribution on $\{1, \dots, N\}$,

Sample A_n from the distribution on $\{1, \dots, N\}$ with probabilities proportional to

$$\{\tilde{g}^{\theta^{n-1}\omega}(\zeta_{n-1}^1), \dots, \tilde{g}^{\theta^{n-1}\omega}(\zeta_{n-1}^N)\},$$

Sample $\zeta_n^{K_n} | \{A_n, K_n, (\zeta_{n-1}^i)_{i=1}^N\} \sim \tilde{f}^{\theta^n \omega}(\zeta_{n-1}^{A_n}, \cdot)$,

Sample $(\zeta_n^i)_{i \neq K_n} | \{K_n, (\zeta_{n-1}^i)_{i=1}^N\} \stackrel{\text{i.i.d.}}{\sim} \frac{\sum_{j=1}^N g(\zeta_{n-1}^j, Y_{n-1}(\omega)) f(\zeta_{n-1}^j, \cdot)}{\sum_{j=1}^N g(\zeta_{n-1}^j, Y_{n-1}(\omega))}$,

Report $\tilde{Z}_n^{\omega, N} = \tilde{Z}_{n-1}^{\omega, N} \cdot \frac{\sum_{i=1}^N \tilde{g}^{\theta^{n-1}\omega}(\zeta_{n-1}^i)}{\sum_{i=1}^N \psi^{\theta^n \omega}(\zeta_n^i)}$.

that the $N \rightarrow \infty$ asymptotic variance associated with $N^{-1/2} \sum_{i=1}^N \varphi(\zeta_n^i)$ is the same when sampled under Algorithms 1 and 2, but the asymptotic variances of $\sqrt{N}(Z_n^{\omega, N} - Z_n^\omega)$ and $\sqrt{N}(\tilde{Z}_n^{\omega, N} - Z_n^\omega)$ are, in general, different.

The finite- N , finite- n behavior of the relative variance of the standard estimate $Z_n^{\omega, N}$ from Algorithm 1 is well understood. Under certain regularity assumptions, it can be deduced from [6], Theorem 5.1, that in our setting $\Upsilon_N(\mathbf{M})$ must satisfy

$$(1.16) \quad \Upsilon_N(\mathbf{M}) \leq \log \left[1 + \frac{C}{N-1} \right]$$

for some finite constant C which depends on g and f . Our fourth main result (Proposition 5) generalizes (1.16) to the case of twisted particle filters. With $\Upsilon_N(\tilde{\mathbf{M}})$ as in (1.11), $\tilde{\mathbf{M}}^\omega$ as in (1.13), and under some regularity conditions,

$$\Upsilon_N(\tilde{\mathbf{M}}) \leq \log \left[1 + \frac{C'}{N-1} \sup_{\omega, x, x'} \left| \frac{h^\omega(x)}{\psi^\omega(x)} - \frac{h^\omega(x')}{\psi^\omega(x')} \right| \right],$$

where C' is a constant. Thus, whenever $\Upsilon_N(\mathbf{M}) > 0$, by choosing ψ “close” to h , we can in principle achieve $\Upsilon_N(\tilde{\mathbf{M}}) < \Upsilon_N(\mathbf{M})$.

The rest of the paper is structured as follows. Section 2 introduces our general setting, addressing the generalized eigenvalue properties of families of nonnegative kernels and sampling laws of the particle systems we consider. Section 3 narrows attention to twisted particle filters and considers some properties in the regime where N is fixed and $n \rightarrow \infty$, and vice-versa. Section 4 discusses the application of our main results to sequential importance sampling, bootstrap and auxiliary particle filters. The proofs of Lemmas 3–4, Propositions 1–5 and Theorems 1–2 are housed in the supplementary material [30].

2. Nonnegative kernels, sampling particles and variance growth.

2.1. *Notation and conventions.* Let $(X, \mathcal{X}), (Y, \mathcal{Y}), \Omega := Y^{\mathbb{Z}}, \mathcal{F} := \mathcal{Y}^{\otimes \mathbb{Z}}$, and \mathbb{P} and θ be as in Section 1. Expectation w.r.t. \mathbb{P} will be denoted by \mathbb{E} . Let $\mathcal{M}(X), \mathcal{P}(X)$ and $\mathcal{L}(X)$ be respectively the collections of measures, probability measures and real-valued, bounded, \mathcal{X} -measurable functions on X . We write

$$\|\varphi\| := \sup_x |\varphi(x)|$$

and

$$(2.1) \quad \mu(\varphi) := \int_X \varphi(x) \mu(dx) \quad \text{for any } \varphi \in \mathcal{L}(X), \mu \in \mathcal{M}(X).$$

We will be dealing throughout with various real-valued functions on $\Omega \times X$ (and more generally $\Omega \times X^N$, etc.). For any such function φ , we write the ω -section of φ as $\varphi^\omega : X \rightarrow \mathbb{R}, \varphi^\omega(x) := \varphi(\omega, x)$. For a function $\xi : \Omega \rightarrow \mathbb{R}$ it will sometimes be convenient to write ξ_ω instead of the more standard $\xi(\omega)$. We will need to express various integration operations involving functions on $\Omega \times X^N$ and their ω -sections, so for completeness we quote the following facts of measure theory (see, e.g., [12], Chapter VI, which will be used repeatedly without further comment): when $\varphi : \Omega \times X^N \rightarrow \mathbb{R}$ is measurable w.r.t. to $\mathcal{F} \otimes \mathcal{X}^{\otimes N}$, then for every $\omega \in \Omega$, the ω -section φ^ω is measurable w.r.t. $\mathcal{X}^{\otimes N}$; and, furthermore, for any σ -finite measure μ on $(X^N, \mathcal{X}^{\otimes N})$, if φ is integrable w.r.t. to $\mathbb{P} \otimes \mu$, then the function acting $\Omega \rightarrow \mathbb{R}$ which maps $\omega \mapsto \mu(\varphi^\omega)$ is measurable w.r.t. \mathcal{F} and is \mathbb{P} -integrable.

Let $\varphi, \tilde{\varphi}$ be two functions, each acting $\Omega \times X^N \rightarrow \mathbb{R}$ and each measurable w.r.t. $\mathcal{F} \otimes \mathcal{X}^{\otimes N}$. We will need to talk about the sets on which such functions take the same values. For any $\omega \in \Omega$, let $A_\omega := \{x \in X^N : \varphi^\omega(x) = \tilde{\varphi}^\omega(x)\}$ and let μ be a σ -finite measure on $(X^N, \mathcal{X}^{\otimes N})$. In order to avoid having to make the sets $\{A_\omega; \omega \in \Omega\}$ explicit in various statements, we will write by convention

$$\text{for } \mathbb{P}\text{-a.a. } \omega, \quad \varphi^\omega(x) = \tilde{\varphi}^\omega(x) \quad \text{for } \mu\text{-a.a. } x$$

to mean $\mathbb{P}(\{\omega : \mu(A_\omega^c) = 0\}) = 1$.

2.2. *Generalized eigenvalue theory for nonnegative kernels.* Fix arbitrarily $\mu_0 \in \mathcal{P}(X)$ and let $M : \Omega \times X \times \mathcal{X} \rightarrow [0, 1]$ be such that $M(\omega, x, \cdot) \in \mathcal{P}(X)$ for each $(\omega, x) \in \Omega \times X$, and $M(\cdot, \cdot, A)$ is $\mathcal{F} \otimes \mathcal{X}$ -measurable for each $A \in \mathcal{X}$. Then for any $\omega, M(\omega, \cdot, \cdot)$ is a Markov kernel on (X, \mathcal{X}) and when it is important to emphasize this perspective, we shall often write $M^\omega(x, A)$ instead of $M(\omega, x, A)$. We shall adopt similar notation for other kernels.

For any fixed $\omega \in \Omega$, let \mathbb{E}^ω denote expectation with respect to the law of the time-inhomogeneous Markov chain $\{X_n; n \geq 0\}$, with each X_n valued in X , initialized from $X_0 \sim \mu_0$ and $X_n | \{X_{n-1} = x_{n-1}\} \sim M^{\theta^{n-1}\omega}(x_{n-1}, \cdot)$, for $n \geq 1$. Let $G : \Omega \times X \rightarrow \mathbb{R}_+$ be a $\mathcal{F} \otimes \mathcal{X}$ -measurable, strictly positive and bounded function.

REMARK 1. This setup is purposefully generic and accommodates, as one particular instance, the case

$$(2.2) \quad G^\omega(x) = g(x, Y_0(\omega)), \quad M^\omega(x, dx') = f(x, dx') \quad \forall \omega \in \Omega,$$

where g and f are as in Section 1, and then $\mathbb{E}^\omega[\prod_{p=0}^{n-1} g(X_p, Y_p(\omega))] = Z_n^\omega = \mathbb{E}^\omega[\prod_{p=0}^{n-1} G^{\theta^p \omega}(X_p)]$. Other instances will be discussed in Section 4.

We next introduce two hypotheses. Since $\Omega := Y^{\mathbb{Z}}$, (H1) amounts to saying that the observation process is stationary and ergodic. (H2) is a strong mixing condition that rarely holds when X and Y are noncompact, and some results do not rely on both (2.3) and (2.4) simultaneously but their combination allows us to avoid a layer of technical presentation which would further lengthen and complicate our proofs.

(H1) The shift operator θ preserves \mathbb{P} and is ergodic.

(H2) There exist constants $\beta \in [1, \infty)$, $(\varepsilon_-, \varepsilon_+) \in (0, \infty)^2$, and $\nu \in \mathcal{P}(X)$ such that

$$(2.3) \quad \frac{G(\omega, x)}{G(\omega', x')} \leq \beta \quad \forall (\omega, \omega', x, x') \in \Omega^2 \times X^2,$$

$$(2.4) \quad \varepsilon_- \nu(\cdot) \leq M(\omega, x, \cdot) \leq \varepsilon_+ \nu(\cdot) \quad \forall (\omega, x) \in \Omega \times X.$$

We now introduce the nonnegative kernel

$$(2.5) \quad Q : \Omega \times X \times \mathcal{X} \rightarrow \mathbb{R}_+, \quad Q(\omega, x, dx') := G(\omega, x)M(\omega, x, dx').$$

For any fixed $\omega \in \Omega$, we define the operators

$$(2.6) \quad Q^\omega(\varphi)(x) := \int_X Q^\omega(x, dx') \varphi(x'), \quad \varphi \in \mathcal{L}(X),$$

$$(2.7) \quad \mu Q^\omega(\cdot) := \int_X \mu(dx) Q^\omega(x, \cdot), \quad \mu \in \mathcal{M}(X),$$

and let $\{Q_n^\omega; n \in \mathbb{N}\}$ be defined recursively by

$$(2.8) \quad Q_0^\omega := \text{Id}, \quad Q_n^\omega = Q_{n-1}^\omega Q^{\theta^{n-1} \omega}, \quad n \geq 1.$$

This operator notation allows us to express

$$(2.9) \quad \mu_0 Q_n^\omega(\varphi) = \mathbb{E}^\omega \left[\varphi(X_n) \prod_{p=0}^{n-1} G^{\theta^p \omega}(X_p) \right], \quad n \geq 1, \varphi \in \mathcal{L}(X).$$

It is well known that (H1) and (H2) together are sufficient to establish the following result; see [25] for related ideas in the context of HMMs.

PROPOSITION 1. *Assume (H1) and (H2). Then there exists a constant $\Lambda \in (-\infty, \infty)$ independent of the initial distribution $\mu_0 \in \mathcal{P}(X)$ such that*

$$(2.10) \quad \frac{1}{n} \log \mathbb{E}^\omega \left[\prod_{p=0}^{n-1} G^{\theta^p \omega}(X_p) \right] \rightarrow \Lambda \quad \text{as } n \rightarrow \infty, \mathbb{P}\text{-a.s.}$$

It turns out that Proposition 1 is one element of a generalized eigenvalue theory for the nonnegative kernel Q . Another element is Proposition 2, which involves the following objects. Let $\Phi^\omega : \mathcal{P}(X) \rightarrow \mathcal{P}(X)$ be defined by

$$\Phi^\omega(\mu) = \frac{\mu Q^\omega}{\mu Q^\omega(1)}, \quad \mu \in \mathcal{P}(X),$$

and let $\{\Phi_n^\omega; n \in \mathbb{N}\}$ be the family of operators defined recursively by

$$\Phi_0^\omega := \text{Id}, \quad \Phi_n^\omega := \Phi^{\theta^{n-1}\omega} \circ \Phi_{n-1}^\omega,$$

so that each Φ_n^ω acts $\mathcal{P}(X) \rightarrow \mathcal{P}(X)$. Under these definitions, for any $n \in \mathbb{N}$,

$$(2.11) \quad \Phi_n^\omega(\mu) = \frac{\mu Q_n^\omega}{\mu Q_n^\omega(1)},$$

which can be verified by induction, since from the above definitions $\Phi_0^\omega = \text{Id}$, $Q_0^\omega := \text{Id}$ and when (2.11) holds,

$$\Phi_{n+1}^\omega(\mu) = (\Phi^{\theta^n\omega} \circ \Phi_n^\omega)(\mu) = \frac{\Phi_n^\omega(\mu) Q^{\theta^n\omega}}{\Phi_n^\omega(\mu) Q^{\theta^n\omega}(1)} = \frac{\mu Q_n^\omega Q^{\theta^n\omega}}{\mu Q_n^\omega Q^{\theta^n\omega}(1)} = \frac{\mu Q_{n+1}^\omega}{\mu Q_{n+1}^\omega(1)}.$$

REMARK 2. In the setting $M^\omega(x, dx') := f(x, dx')$, $G^\omega(x) := g(x, Y_0(\omega))$, then if μ_0 and π_n^ω are respectively the initial distribution and prediction-filter as in (1.2), we have

$$\pi_{n+1}^\omega = \Phi^{\theta^n\omega}(\pi_n^\omega), \quad n \geq 0.$$

REMARK 3. Under (H2), it is known that Φ_n^ω is exponentially stable with respect to initial conditions (e.g., [8], Chapter 4) that is, there exist constants $C < \infty$ and $\rho < 1$ such that for any $\varphi \in \mathcal{L}(X)$ and any $n \geq 1$,

$$(2.12) \quad \sup_{\omega \in \Omega} \sup_{\mu, \mu' \in \mathcal{P}(X)} |[\Phi_n^\omega(\mu) - \Phi_n^\omega(\mu')](\varphi)| \leq \|\varphi\| C \rho^n.$$

Equation (2.12) is used extensively in the proof of the following proposition, which is a variation on the theme of Kifer’s Perron–Frobenius theorem for positive operators in a random environment [21], Theorem 3.1.

PROPOSITION 2. Assume (H2).

(1) Fix $\mu \in \mathcal{P}(X)$. Then the limits

$$(2.13) \quad \eta^\omega(A) := \lim_{n \rightarrow \infty} \Phi_n^{\theta^{-n}\omega}(\mu)(A), \quad \omega \in \Omega, A \in \mathcal{X},$$

$$(2.14) \quad h(\omega, x) := \lim_{n \rightarrow \infty} \frac{Q_n^\omega(1)(x)}{\Phi_n^{\theta^{-n}\omega}(\mu) Q_n^\omega(1)}, \quad \omega \in \Omega, x \in X,$$

exist and define a family of probability measures $\eta := \{\eta^\omega \in \mathcal{P}(X); \omega \in \Omega\}$ and an $\mathcal{F} \otimes \mathcal{X}$ -measurable function $h : \Omega \times X \rightarrow \mathbb{R}$.

(2) In fact, η and h are independent of the particular μ chosen in part (1) and there exist constants $C < \infty$ and $\rho < 1$ such that for any $\varphi \in \mathcal{L}(X)$,

$$(2.15) \quad \sup_{\omega \in \Omega} \sup_{\mu \in \mathcal{P}(X)} |[\Phi_n^{\theta^{-n}\omega}(\mu) - \eta^\omega](\varphi)| \leq \|\varphi\| C \rho^n, \quad n \geq 1$$

and

$$(2.16) \quad \sup_{\omega \in \Omega} \sup_{x \in X} \sup_{\mu \in \mathcal{P}(X)} \left| \frac{Q_n^\omega(1)(x)}{\Phi_n^{\theta^{-n}\omega}(\mu) Q_n^\omega(1)} - h(\omega, x) \right| \leq C \rho^n, \quad n \geq 1.$$

(3) $\lambda: \omega \in \Omega \mapsto \eta^\omega(G^\omega) \in \mathbb{R}_+$ is measurable w.r.t. \mathcal{F} and we have

$$(2.17) \quad \sup_{(\omega, \omega') \in \Omega^2} \frac{\lambda_\omega}{\lambda_{\omega'}} < \infty, \quad \sup_{(\omega, \omega', x, x') \in \Omega^2 \times X^2} \frac{h(\omega, x)}{h(\omega', x')} < \infty.$$

(4) Among all triples which consist of (i) an Ω -indexed family of probability measures on (X, \mathcal{X}) , (ii) an \mathbb{R}_+ -valued, not identically zero, measurable function on $\Omega \times X$, and (iii) a measurable function on Ω , the triple (η, h, λ) , with η, h as in part (1) and λ as in part (3), uniquely satisfies the system of equations

$$(2.18) \quad \eta^\omega Q^\omega = \lambda_\omega \eta^{\theta\omega}, \quad Q^\omega(h^{\theta\omega}) = \lambda_\omega h^\omega, \quad \eta^\omega(h^\omega) = 1$$

for all $\omega \in \Omega$.

The connection with Proposition 1 is as follows:

PROPOSITION 3. Assume (H1), (H2) and let Λ be as in Proposition 1 and λ be as in Proposition 2. Then

$$(2.19) \quad \Lambda = \mathbb{E}[\log \lambda] = \int_{\Omega} \log \frac{Q^\omega(h^{\theta\omega})(x)}{h^\omega(x)} \mathbb{P}(d\omega) \quad \text{for any } x \in X.$$

In the setting of HMMs as per Remark 1, equalities like the first one in (2.19) appear routinely in the study of likelihood-based estimators [14, 25]. However, it is the second equality in (2.19), and generalizations thereof, which shall be crucial for our purposes in the sequel.

REMARK 4. If one weakens the “1-step” condition (2.4) to an m -step version for some $m \geq 1$, then Propositions 1–3 can easily be generalized, working with the kernel Q_m^ω instead of Q^ω . Part of the utility of the uniform in ω and x bounds in (H2) is that various parts of Proposition 2 hold uniformly over $\omega \in \Omega$. If one allows ω -dependent constants and measures in (2.3) and (2.4), and imposes certain explicit compactness and continuity assumptions and (H1), then [21], Theorem 3.1, provides a partial alternative to our Proposition 2.

We proceed by introducing the laws of the particle systems of interest.

2.3. *Law of the standard particle system.* Unless stated otherwise, in this section we fix arbitrarily $N \geq 1$ and write $\mathcal{P}(\mathbf{X}^N)$ for the collection of probability measures on $(\mathbf{X}^N, \mathcal{X}^{\otimes N})$.

Let $\mathbf{M} : \Omega \times \mathbf{X}^N \times \mathcal{X}^{\otimes N} \rightarrow [0, 1]$ be given, in integral form, by

$$(2.20) \quad \mathbf{M}(\omega, x, dz) = \prod_{i=1}^N \left[\frac{\sum_{j=1}^N G(\omega, x^j) M(\omega, x^j, dz^i)}{\sum_{j=1}^N G(\omega, x^j)} \right],$$

where $x = (x^1, \dots, x^N), z = (z^1, \dots, z^N) \in \mathbf{X}^N$. Each member of the family $\{\mathbf{M}^\omega; \omega \in \Omega\}$ is a Markov transition kernel for the entire N -particle system according to a “multinomial” resampling scheme with fitness function $G(\omega, \cdot)$, followed by conditionally independent mutation according to M^ω .

Now for any given $\omega \in \Omega$, we shall denote by \mathbb{E}_N^ω expectation with respect to the law of the Markov chain $\{\zeta_n; n \geq 0\}$, with each $\zeta_n = \{\zeta_n^1, \dots, \zeta_n^N\}$ valued in \mathbf{X}^N and

$$(2.21) \quad \zeta_0 \sim \mu_0^{\otimes N}, \quad \zeta_n | \zeta_{n-1} \sim \mathbf{M}^{\theta^{n-1}\omega}(\zeta_{n-1}, \cdot).$$

We define, with $x = (x^1, \dots, x^N)$,

$$(2.22) \quad \mathbf{G} : (\omega, x) \in \Omega \times \mathbf{X}^N \mapsto \frac{1}{N} \sum_{i=1}^N G(\omega, x^i) \in \mathbb{R}_+.$$

REMARK 5. For any $\varphi \in \mathcal{L}(\mathbf{X})$, if we define the function

$$\varphi : x = (x^1, \dots, x^N) \in \mathbf{X}^N \mapsto \frac{1}{N} \sum_{i=1}^N \varphi(x^i) \in \mathbb{R},$$

then the lack-of-bias property of the particle approximation [8], Proposition 7.4.1, is

$$(2.23) \quad \mathbb{E}_N^\omega \left[\varphi(\zeta_n) \prod_{p=0}^{n-1} \mathbf{G}^{\theta^p \omega}(\zeta_p) \right] = \mathbb{E}^\omega \left[\varphi(X_n) \prod_{p=0}^{n-1} G^{\theta^p \omega}(X_p) \right].$$

REMARK 6. When $M^\omega(x, \cdot) = f(x, \cdot)$ and $G^\omega(x) = g(x, Y_0(\omega))$, the sampling recipe for simulating the process $\{\zeta_n; n \geq 0\}$ according to (2.21) is the bootstrap particle filter: Algorithm 1. Furthermore, the particle approximation of Z_n^ω is then $\prod_{p=0}^{n-1} \mathbf{G}^{\theta^p \omega}(\zeta_p)$. To see it is unbiased, apply (2.23) with $\varphi = 1$.

Part of our investigation will develop some limit theory for

$$(2.24) \quad \frac{\mathbb{E}_N^\omega [\prod_{p=0}^{n-1} \mathbf{G}^{\theta^p \omega}(\zeta_p)^2]}{\mathbb{E}^\omega [\prod_{p=0}^{n-1} G^{\theta^p \omega}(X_p)]^2},$$

when N is fixed and $n \rightarrow \infty$. Our notation \mathbf{G}, \mathbf{M} and (2.23) are intended to hint that the phenomena described in Propositions 1–3 are relevant to the study of (2.24). Indeed, this is the direction in which we are heading. However, we will actually study an object more general than (2.24), arising from a more general form of particle approximation, for the particle system is distributed according to some Markov law, possibly different to (2.21).

2.4. *Alternative sampling of the particle system.* Let us introduce $\tilde{\mathbf{M}}: \Omega \times \mathcal{X}^N \times \mathcal{X}^{\otimes N} \rightarrow [0, 1]$, possibly different from \mathbf{M} . For fixed ω , now denote by $\tilde{\mathbb{E}}_N^\omega$ expectation with respect to law of the Markov chain

$$(2.25) \quad \zeta_0 \sim \mu_0^{\otimes N}, \quad \zeta_n | \zeta_{n-1} \sim \tilde{\mathbf{M}}^{\theta^{n-1}\omega}(\zeta_{n-1}, \cdot).$$

We are going to specify a class of candidates for $\tilde{\mathbf{M}}$, and we first notice that the regularity condition (H2) transfers to \mathbf{G}, \mathbf{M} in the following sense:

LEMMA 1. *Assume (H2). Then for any $N \geq 1$,*

$$\begin{aligned} \frac{\mathbf{G}(\omega, x)}{\mathbf{G}(\omega', x')} &\leq \beta \quad \forall (\omega, \omega', x, x') \in \Omega^2 \times \mathcal{X}^{2N}, \\ \varepsilon_-^N \nu^{\otimes N}(\cdot) &\leq \mathbf{M}(\omega, x, \cdot) \leq \varepsilon_+^N \nu^{\otimes N}(\cdot) \quad \forall (\omega, x) \in \Omega \times \mathcal{X}^N. \end{aligned}$$

The proof is omitted. We shall consider the following family of kernels.

DEFINITION 1 (of \mathbb{M}). Any $\tilde{\mathbf{M}}: \Omega \times \mathcal{X}^N \times \mathcal{X}^{\otimes N} \rightarrow [0, 1]$ is a member of \mathbb{M} if and only if there exist constants $(\tilde{\varepsilon}_-, \tilde{\varepsilon}_+) \in (0, \infty)^2$ and $\tilde{\nu} \in \mathcal{P}(\mathcal{X}^N)$ such that

$$(2.26) \quad \begin{aligned} \tilde{\nu}(\cdot)\tilde{\varepsilon}_- &\leq \tilde{\mathbf{M}}(\omega, x, \cdot) \leq \tilde{\varepsilon}_+\tilde{\nu}(\cdot) \quad \forall (\omega, x) \in \Omega \times \mathcal{X}^N, \\ \nu^{\otimes N} &\ll \tilde{\nu} \quad \text{and} \quad \int_{\mathcal{X}^N} \left(\frac{d\nu^{\otimes N}}{d\tilde{\nu}}(x) \right)^2 \tilde{\nu}(dx) < \infty, \end{aligned}$$

where ν is as in (H2).

When $\tilde{\mathbf{M}}$ is a member of \mathbb{M} we write

$$(2.27) \quad \phi^\omega(x, x') := \frac{d\mathbf{M}^\omega(x, \cdot)}{d\tilde{\mathbf{M}}^\omega(x, \cdot)}(x'), \quad (\omega, x, x') \in \Omega \times \mathcal{X}^{2N},$$

and in the context of sampling the particle system $(\zeta_n; n \geq 0)$ under the law (2.25), we will take

$$(2.28) \quad \prod_{p=0}^{n-1} \mathbf{G}^{\theta^p\omega}(\zeta_p) \phi^{\theta^p\omega}(\zeta_p, \zeta_{p+1})$$

as an approximation of $\mathbb{E}^\omega[\prod_{p=0}^{n-1} G^{\theta^p \omega}(X_p)]$. In light of (2.23) and (2.27), we have

$$\tilde{\mathbb{E}}_N^\omega \left[\prod_{p=0}^{n-1} \mathbf{G}^{\theta^p \omega}(\zeta_p) \phi^{\theta^p \omega}(\zeta_p, \zeta_{p+1}) \right] = \mathbb{E}^\omega \left[\prod_{p=0}^{n-1} G^{\theta^p \omega}(X_p) \right].$$

The following result describes the $n \rightarrow \infty$ behavior of

$$(2.29) \quad \tilde{\mathcal{V}}_{n,N}^\omega := \frac{\tilde{\mathbb{E}}_N^\omega[\prod_{p=0}^{n-1} \mathbf{G}^{\theta^p \omega}(\zeta_p)^2 \phi^{\theta^p \omega}(\zeta_p, \zeta_{p+1})^2]}{\mathbb{E}^\omega[\prod_{p=0}^{n-1} G^{\theta^p \omega}(X_p)]^2}.$$

Its proof starts by considering the family of kernels $\{\tilde{\mathbf{R}}^\omega; \omega \in \Omega\}$, with

$$\tilde{\mathbf{R}}^\omega(x, dx') := \mathbf{G}^\omega(x)^2 \phi^\omega(x, x')^2 \tilde{\mathbf{M}}^\omega(x, dx'),$$

in terms of which the numerator of (2.29) may be written and which exhibit exactly similar properties to Q^ω appearing in the proof of Proposition 1.

PROPOSITION 4. *Assume (H1), (H2) and fix $N \geq 1$ arbitrarily. For every $\tilde{\mathbf{M}} \in \mathbb{M}$ there exists a constant $\Upsilon_N(\tilde{\mathbf{M}}) \in [0, \infty)$, independent of the initial distribution μ_0 such that*

$$\frac{1}{n} \log \tilde{\mathcal{V}}_{n,N}^\omega \longrightarrow \Upsilon_N(\tilde{\mathbf{M}}) \quad \text{as } n \rightarrow \infty, \mathbb{P}\text{-a.s.}$$

We now proceed to address the question of how $\Upsilon_N(\tilde{\mathbf{M}})$ depends on $\tilde{\mathbf{M}}$. To this end, let us introduce two further pieces of notation:

$$\mathbf{Q}(\omega, x, dx') := \mathbf{G}(\omega, x) \mathbf{M}(\omega, x, dx'),$$

and when (H2) holds, so that h as in Proposition 2 is well-defined, consider the function

$$(2.30) \quad \mathbf{h}: (\omega, x) \in \Omega \times \mathbf{X}^N \longmapsto \frac{1}{N} \sum_{i=1}^N h(\omega, x^i) \in \mathbb{R}_+.$$

Our interest in (2.30) stems from the following pivotal lemma, which shows how the generalized eigenfunction h and eigenvalue λ of Q appearing in Proposition 2 define a generalized eigenfunction and eigenvalue for \mathbf{Q} , for any $N \geq 1$. Its proof is quite elementary, but is included here for exposition since the structure it deals with underpins the algorithmic developments in Section 3.

LEMMA 2. For any $\omega \in \Omega$,

$$\mathbf{Q}^\omega(\mathbf{h}^{\theta \omega}) = \lambda_\omega \mathbf{h}^\omega,$$

where λ_ω is as in Proposition 2.

PROOF.

$$\begin{aligned}
\mathbf{Q}^\omega(\mathbf{h}^{\theta\omega})(x) &= \frac{1}{N} \sum_{k=1}^N \int_{\mathcal{X}^N} \mathbf{Q}^\omega(x, dz) h^{\theta\omega}(z^k) \\
&= \frac{1}{N} \sum_{k=1}^N \mathbf{G}^\omega(x) \int_{\mathcal{X}} \frac{\sum_{i=1}^N \mathcal{Q}^\omega(x^i, dz^k)}{\sum_{i=1}^N \mathbf{G}^\omega(x^i)} h^{\theta\omega}(z^k) \\
&= \frac{1}{N} \sum_{k=1}^N \frac{1}{N} \sum_{i=1}^N \int_{\mathcal{X}} \mathcal{Q}^\omega(x^i, dz^k) h^{\theta\omega}(z^k) \\
&= \lambda_\omega \frac{1}{N} \sum_{i=1}^N h^\omega(x^i) = \lambda_\omega \mathbf{h}^\omega(x).
\end{aligned}$$

□

Now consider taking

$$(2.31) \quad \tilde{\mathbf{M}}^\omega(x, dx') = \frac{\mathbf{M}^\omega(x, dx') \mathbf{h}^{\theta\omega}(x')}{\int_{\mathcal{X}^N} \mathbf{M}^\omega(x, dz) \mathbf{h}^{\theta\omega}(z)},$$

which is a member of \mathbb{M} , due to the definition of \mathbf{h} and part (3) of Proposition 2. In this case we have

$$\begin{aligned}
&\prod_{p=0}^{n-1} \mathbf{G}^{\theta^p\omega}(\zeta_p) \phi^{\theta^p\omega}(\zeta_p, \zeta_{p+1}) \\
&= \prod_{p=0}^{n-1} \mathbf{G}^{\theta^p\omega}(\zeta_p) \frac{\int_{\mathcal{X}^N} \mathbf{M}^{\theta^p\omega}(\zeta_p, dz_{p+1}) \mathbf{h}^{\theta^{p+1}\omega}(z_{p+1})}{\mathbf{h}^{\theta^{p+1}\omega}(\zeta_{p+1})} \\
(2.32) \quad &= \prod_{p=0}^{n-1} \frac{\mathbf{Q}^{\theta^p\omega}(\mathbf{h}^{\theta^{p+1}\omega})(\zeta_p)}{\mathbf{h}^{\theta^{p+1}\omega}(\zeta_{p+1})} \\
&= \frac{\mathbf{h}^\omega(\zeta_0)}{\mathbf{h}^{\theta^n\omega}(\zeta_n)} \prod_{p=0}^{n-1} \frac{\mathbf{Q}^{\theta^p\omega}(\mathbf{h}^{\theta^{p+1}\omega})(\zeta_p)}{\mathbf{h}^{\theta^p\omega}(\zeta_p)} \\
&= \frac{\mathbf{h}^\omega(\zeta_0)}{\mathbf{h}^{\theta^n\omega}(\zeta_n)} \prod_{p=0}^{n-1} \lambda_{\theta^p\omega},
\end{aligned}$$

where the final equality is due to Lemma 2. Thus, if we choose $\tilde{\mathbf{M}}$ as per (2.31), then the quantity in (2.32) depends on the particle system trajectory ζ_0, \dots, ζ_n only through the quantities $\mathbf{h}^\omega(\zeta_0)$ and $\mathbf{h}^{\theta^n\omega}(\zeta_n)$, and we then might hope that $\Upsilon_N(\tilde{\mathbf{M}}) = 0$. This turns out to be true, and much more strikingly, up to its definition on certain sets of measure zero, $\tilde{\mathbf{M}}$ as in (2.31) is the unique member of \mathbb{M} which achieves $\Upsilon_N(\tilde{\mathbf{M}}) = 0$, in the sense of the following theorem.

THEOREM 1. *Assume (H1), (H2), let $N \geq 1$ be fixed arbitrarily and assume $\tilde{\mathbf{M}}$ belongs to \mathbb{M} . Then (1)–(3) are equivalent:*

(1) $\Upsilon_N(\tilde{\mathbf{M}}) = 0$.

(2) *For \mathbb{P} -almost all $\omega \in \Omega$, there exists $A_\omega \in \mathcal{X}^{\otimes N}$ such that $v^{\otimes N}(A_\omega^c) = 0$ and for any $x \in A_\omega$,*

$$(2.33) \quad \tilde{\mathbf{M}}^\omega(x, B) = \frac{\int_B \mathbf{M}^\omega(x, dx') \mathbf{h}^{\theta\omega}(x')}{\int_{\mathcal{X}^N} \mathbf{M}^\omega(x, dz) \mathbf{h}^{\theta\omega}(z)} \quad \text{for all } B \in \mathcal{X}^{\otimes N}.$$

(3) *For \mathbb{P} -almost all $\omega \in \Omega$, $\sup_n \tilde{V}_{n,N}^\omega < \infty$.*

The re-weighted particle transitions (2.33) are reminiscent of certain eigenfunction transformations of general type branching processes studied by Athreya [4] and more broadly can be viewed as a randomized version of Doob’s h -process. See [30] for further information. Time-homogeneous counterparts of such transitions arise in the analysis of certain Markov chain rare event problems [5]; in order to prove (1) \Rightarrow (2), we generalize the proof of necessity in [5], Theorem 3, to the case of families of kernels driven by an ergodic, measure-preserving transform.

The following lemma serves to accompany Proposition 4 and Theorem 1, and provides necessary and sufficient conditions for $\Upsilon_N(\tilde{\mathbf{M}}) = 0$ in the case of taking $\tilde{\mathbf{M}} = \mathbf{M}$, that is, the transitions of the standard particle system.

LEMMA 3. *Assume (H1), (H2) and let $N \geq 1$ be fixed arbitrarily. Then (1)–(3) are equivalent:*

(1) $\Upsilon_N(\mathbf{M}) = 0$.

(2) *For \mathbb{P} -a.a. ω , $h^\omega(x) = 1$, for v -a.a. x .*

(3) *There exists a random variable $C : \Omega \rightarrow \mathbb{R}_+$ such that*

$$\text{for } \mathbb{P}\text{-a.a. } \omega, \quad G^\omega(x) = C_\omega \quad \text{for } v\text{-a.a. } x.$$

In situations of practical interest, point (3) of Lemma 3 is usually false, and then it must be the case that $\Upsilon_N(\mathbf{M}) > 0$. It then appears that a choice of $\tilde{\mathbf{M}}$ which approximates the optimal transition, (2.33), may yield a provable performance advantage over \mathbf{M} , in the sense of achieving strict inequality $\Upsilon_N(\tilde{\mathbf{M}}) < \Upsilon_N(\mathbf{M})$. This leads us to consider the class of particle algorithms treated in the next section.

3. Twisted particle algorithms. The form of the optimal transition kernel (2.33) suggests that we consider families of kernels arising from re-weighting of $\mathbf{M}^\omega(x, \cdot)$ by an additive, nonnegative functional. In this section we will analyze particle algorithms arising from kernels of this general form. Let $\psi : \Omega \times \mathcal{X} \rightarrow \mathbb{R}_+$ be a strictly positive, bounded and measurable function and define

$$\psi : (\omega, x) \in \Omega \times \mathcal{X}^N \mapsto \frac{1}{N} \sum_{i=1}^N \psi(\omega, x^i) \in \mathbb{R}_+.$$

For the purposes of this section, let us consider the following mild regularity assumption:

(H3) For each $\omega \in \Omega$, $\sup_x G^\omega(x) < \infty$ and $\sup_x \psi^\omega(x) < \infty$.

When (H3) holds the following Markov kernel is well-defined:

$$(3.1) \quad \tilde{\mathbf{M}}^\omega(x, dx') = \frac{\mathbf{M}^\omega(x, dx') \psi^{\theta\omega}(x')}{\int_{X^N} \mathbf{M}^\omega(x, dz) \psi^{\theta\omega}(z)}.$$

We shall analyze particle approximations which arise from sampling under (3.1). Our motivation here is that we have in mind situations where ψ is chosen to be some approximation of h , assuming the latter exists. The kernel (3.1) accommodates the standard transition (2.20) (e.g., take $\psi^\omega = 1$) and the optimal transition identified in Theorem 1 (take $\psi^\omega = h^\omega$). We note that (3.1) depends on ψ^ω only up to a constant of proportionality.

This section addresses two main objectives: First, to validate the particle approximations delivered when sampling under (3.1), by analyzing some of their convergence and fluctuation properties in the regime where $N \rightarrow \infty$. Second, to provide an estimate of $\Upsilon_N(\tilde{\mathbf{M}})$ which exhibits dependence on N and on the discrepancy between ψ^ω and h^ω .

Let us introduce a little more notation. Define, for each $\omega \in \Omega$, the sequence of probability measures:

$$\eta_0^\omega := \mu_0, \quad \eta_n^\omega := \Phi^{\theta^{n-1}\omega}(\eta_{n-1}^\omega), \quad n \geq 1.$$

With $\{\zeta_n; n \geq 0\}$ the sequence of generations of the particles, we write

$$\eta_n^N := \frac{1}{N} \sum_{i=1}^N \delta_{\zeta_n^i}, \quad n \geq 0,$$

$$\phi_n^{\omega, N} := \frac{\eta_{n-1}^N \mathcal{Q}^{\theta^{n-1}\omega}(\psi^{\theta^n\omega})}{\eta_{n-1}^N(G^{\theta^{n-1}\omega})} \frac{1}{\eta_n^N(\psi^{\theta^n\omega})} = \frac{\Phi^{\theta^{n-1}\omega}(\eta_{n-1}^N)(\psi^{\theta^n\omega})}{\eta_n^N(\psi^{\theta^n\omega})}, \quad n \geq 1,$$

$$\gamma_0^\omega := \mu_0, \quad \gamma_n^\omega(\varphi) := \eta_n^\omega(\varphi) \prod_{p=0}^{n-1} \eta_p^\omega(G^{\theta^p\omega}), \quad n \geq 1,$$

$$\gamma_0^{\omega, N} := \eta_0^N, \quad \gamma_n^{\omega, N}(\varphi) := \eta_n^N(\varphi) \prod_{p=0}^{n-1} \eta_p^N(G^{\theta^p\omega}) \phi_{p+1}^{\omega, N}, \quad n \geq 1.$$

To connect with (2.27), we note that for $\tilde{\mathbf{M}}$ as in (3.1), we have

$$\begin{aligned} \phi_n^{\omega, N} &= \left(\frac{1}{N} \sum_{i=1}^N G^{\theta^{n-1}\omega}(\zeta_{n-1}^i) \right)^{-1} \frac{\sum_{i=1}^N \mathcal{Q}^{\theta^{n-1}\omega}(\psi^{\theta^n\omega})(\zeta_{n-1}^i)}{\sum_{i=1}^N \psi^{\theta^n\omega}(\zeta_{n-1}^i)} \\ &= \phi^{\theta^n\omega}(\zeta_{n-1}, \zeta_n) \end{aligned}$$

Algorithm 3 Twisted particle algorithm

For $n = 0$,

Sample $(\zeta_0^i)_{i=1}^N \stackrel{\text{i.i.d.}}{\sim} \mu_0$ and report $\tilde{Z}_0^{\omega, N} = 1$.

For $n \geq 1$,

Sample K_n from the uniform distribution on $\{1, \dots, N\}$,

Sample A_n from the distribution on $\{1, \dots, N\}$ with probabilities proportional to

$$\{Q^{\theta^{n-1}\omega}(\psi^{\theta^n\omega})(\zeta_{n-1}^1), \dots, Q^{\theta^{n-1}\omega}(\psi^{\theta^n\omega})(\zeta_{n-1}^N)\},$$

Sample $\zeta_n^{K_n} | \{A_n, K_n, (\zeta_{n-1}^i)_{i=1}^N\} \sim \tilde{M}^{\theta^n\omega}(\zeta_{n-1}^{A_n}, \cdot)$,

Sample $(\zeta_n^i)_{i \neq K_n} | \{K_n, (\zeta_{n-1}^i)_{i=1}^N\} \stackrel{\text{i.i.d.}}{\sim} \frac{\sum_{j=1}^N G^{\theta^{n-1}\omega}(\zeta_{n-1}^j) M^{\theta^{n-1}\omega}(\zeta_{n-1}^j, \cdot)}{\sum_{j=1}^N G^{\theta^{n-1}\omega}(\zeta_{n-1}^j)}$,

Report $\tilde{Z}_n^{\omega, N} = \tilde{Z}_{n-1}^{\omega, N} \cdot \frac{\sum_{i=1}^N Q^{\theta^{n-1}\omega}(\psi^{\theta^n\omega})(\zeta_{n-1}^i)}{\sum_{i=1}^N \psi^{\theta^n\omega}(\zeta_n^i)}$.

and

$$\prod_{p=0}^{n-1} \mathbf{G}^{\theta^p\omega}(\zeta_p) \phi^{\theta^p\omega}(\zeta_p, \zeta_{p+1}) = \prod_{p=0}^{n-1} \frac{\sum_{i=1}^N Q^{\theta^p\omega}(\psi^{\theta^{p+1}\omega})(\zeta_p^i)}{\sum_{i=1}^N \psi^{\theta^{p+1}\omega}(\zeta_{p+1}^i)} = \gamma_n^{\omega, N}(1).$$

Introducing the Markov kernel $\tilde{M}^\omega(x, dx') \propto M^\omega(x, dx') \psi^{\theta\omega}(x')$, Algorithm 3 gives a recipe for sampling the particle system according to (2.25) with $\tilde{\mathbf{M}}$ as in (3.1) (details of the derivation of this algorithm are given in [30]). Here K_n and A_n are just some auxiliary random variables introduced for algorithmic convenience.

3.1. Analysis for $N \rightarrow \infty$.

LEMMA 4. For each, $n \geq 0$, fixed $\omega \in \Omega$ and $\varphi \in \mathcal{L}(\mathbf{X})$,

$$(3.2) \quad \eta_n^N(\varphi) - \eta_n^\omega(\varphi) \longrightarrow 0,$$

$$(3.3) \quad \gamma_n^{\omega, N}(\varphi) - \gamma_n^\omega(\varphi) \longrightarrow 0$$

almost surely, as $N \rightarrow \infty$.

Now define

$$\overline{Q}_{n,n}^\omega := \text{Id}, \quad n \geq 0, \quad \overline{Q}_{p,n}^\omega := \frac{Q^{\theta^p\omega} \dots Q^{\theta^{n-1}\omega}}{\prod_{q=p}^{n-1} \eta_q^\omega(G^{\theta^q\omega})}, \quad n \geq 1, 0 \leq p < n,$$

and notice that $\mu_0 \overline{Q}_{0,n}^\omega = \eta_n^\omega$.

THEOREM 2. Assume (H3). Then for any $n \geq 0$, fixed $\omega \in \Omega$ and $\varphi \in \mathcal{L}(X)$,

$$(3.4) \quad \sqrt{N}[\gamma_n^{\omega, N}(\varphi) - \gamma_n^\omega(\varphi)] \Rightarrow \mathcal{N}(0, \varsigma_{n, \omega}^2(\varphi)),$$

$$(3.5) \quad \sqrt{N}[\eta_n^N(\varphi) - \eta_n^\omega(\varphi)] \Rightarrow \mathcal{N}(0, \sigma_{n, \omega}^2(\varphi))$$

as $N \rightarrow \infty$ where

$$(3.6) \quad \varsigma_{n, \omega}^2(\varphi) = \sum_{p=0}^n \gamma_p^\omega(1)^2 \eta_p^\omega \left[\left(Q_{n-p}^{\theta^p \omega}(\varphi) - \frac{\psi^{\theta^p \omega}}{\eta_p^\omega(\psi^{\theta^p \omega})} \eta_p^\omega Q_{n-p}^{\theta^p \omega}(\varphi) \right)^2 \right],$$

with the convention $\psi^\omega / \eta_0^\omega(\psi^\omega) = 1$, and

$$(3.7) \quad \sigma_{n, \omega}^2(\varphi) = \sum_{p=0}^n \eta_p^\omega [(\bar{Q}_{p, n}^\omega(\varphi) - \eta_n^\omega(\varphi))^2].$$

REMARK 7. The asymptotic variance expression (3.7) is independent of the particular choice of ψ [the CLT holding subject to (H3), of course] and is exactly the same expression obtained in the CLT for the standard particle system (i.e., ψ constant); see, for example, [8], Proposition 9.4.2. However, the asymptotic variance in (3.6) clearly does depend on ψ in general.

3.2. *Analysis for $n \rightarrow \infty$.* For $\tilde{\mathbf{M}}$ as in (3.1), we obtain an estimate of $\Upsilon_N(\tilde{\mathbf{M}})$ which exhibits its dependence on N and the discrepancy between ψ and h .

PROPOSITION 5. Assume (H1), (H2) and $\sup_{\omega, \omega', x, x'} \psi^\omega(x) / \psi^{\omega'}(x') < \infty$. Then for any $N \geq 2$,

$$\Upsilon_N(\tilde{\mathbf{M}}) \leq \log \left[1 + \frac{1}{N-1} \mathcal{D}_{\mathbb{P}}(\psi, h) \right],$$

where

$$\mathcal{D}_{\mathbb{P}}(\psi, h) := \mathbb{P} - \text{ess sup}_\omega \left\{ C_\omega \sup_{(z, z') \in X^2} \left| \frac{h^\omega(z)}{\psi^\omega(z)} - \frac{h^\omega(z')}{\psi^\omega(z')} \right| \right\},$$

$$C_\omega := \left(2 \sup_{z, z' \in X} \frac{\psi^\omega(z)}{\psi^\omega(z')} - 1 \right) \sup_{z \in X} \left(\frac{\psi^\omega(z)}{h^\omega(z)} \right).$$

4. Discussion.

4.1. *Sequential importance sampling.* In the case $N = 1$, we have by inspection of (1.7) and (2.22) the identity $\mathbf{M}^\omega \equiv M^\omega$, so the particle process $\{\zeta_n; n \geq 0\}$ reduces to a Markov chain with state-space X and also $\mathbf{G}^\omega(x) \equiv G^\omega(x)$. With these observations in hand, we may apply our results to analyze sequential importance sampling (SIS) estimators: arithmetic averages involving independent copies of this (and other) Markov chains on X .

Let $\tilde{M}: \Omega \times \mathbb{X} \times \mathcal{X} \rightarrow [0, 1]$ be a Markov kernel, and for some $L \geq 1$ and any fixed $\omega \in \Omega$, let $\{X_n^i; n \geq 0\}_{i=1}^L$ be L i.i.d. time-inhomogeneous Markov chains, each with law

$$(4.1) \quad X_0^i \sim \mu_0, \quad X_n^i | \{X_{n-1}^i = x_{n-1}^i\} \sim \tilde{M}^{\theta^{n-1}\omega}(x_{n-1}^i, \cdot), \quad n \geq 1.$$

To connect with the setting of Sections 2.3 and 2.4, let $N = 1$, and set $\tilde{\mathbf{M}} := \tilde{M}$. We shall assume that (H1) and (H2) hold and that \tilde{M} is a member of \mathbb{M} . With each $\{X_n^i; n \geq 0\}$ distributed according to (4.1), the quantity

$$(4.2) \quad \frac{1}{L} \sum_{i=1}^L \left[\prod_{p=0}^{n-1} G^{\theta^p \omega}(X_p^i) \phi^{\theta^p \omega}(X_p^i, X_{p+1}^i) \right]$$

is clearly an unbiased estimator of $\mathbb{E}^\omega[\prod_{p=0}^{n-1} G^{\theta^p \omega}(X_p^i)]$. Furthermore, since the L Markov chains are independent and $\mathbf{G}^\omega(x) \equiv G^\omega(x)$, for any fixed ω the relative variance of (4.2) is $L^{-1}(\tilde{\mathcal{V}}_{n,1}^\omega - 1)$, where $\tilde{\mathcal{V}}_{n,1}^\omega$ is as in (2.29). By application of Proposition (4) (again with $N = 1$), we have the \mathbb{P} -almost-sure convergence

$$(4.3) \quad \frac{1}{n} \log \tilde{\mathcal{V}}_{n,1}^\omega \longrightarrow \Upsilon_1(\tilde{M}),$$

and so if $L = L(n)$,

$$(4.4) \quad \liminf_{n \rightarrow \infty} \frac{1}{n} \log \left(\frac{1}{L(n)} \tilde{\mathcal{V}}_{n,1}^\omega \right) = \Upsilon_1(\tilde{M}) - \limsup_{n \rightarrow \infty} \frac{1}{n} \log L(n),$$

\mathbb{P} -almost-surely. By Theorem 1, except in the case (up to the sets of measure zero mentioned therein) that $\tilde{M}^\omega(x, dx') \propto M^\omega(x, dx') h^{\theta\omega}(x')$, $\Upsilon_1(\tilde{M}) > 0$ and so the number of chains $L(n)$ must be scaled up exponentially in n in order to prevent exponential growth of the relative variance of (4.2). In this sense the SIS approach is typically an inefficient method for approximating $\mathbb{E}^\omega[\prod_{p=0}^{n-1} G^{\theta^p \omega}(X_p^i)]$, at least relative to particle methods, which we shall now discuss.

4.2. *The bootstrap particle filter.* In the case

$$(4.5) \quad G(\omega, x) := g(x, Y_0(\omega)), \quad M(\omega, x, dx') := f(x, dx'),$$

we have that $\{\mathbf{M}^\omega; \omega \in \Omega\}$ is the collection of the transitions of the bootstrap particle filter, as described in the Introduction. When (H1) and (H2) hold, by Lemma 3 we find that in this scenario, for any $N \geq 1$, $\Upsilon_N(\mathbf{M}) = 0$ if and only if for \mathbb{P} -a.a. ω , $\exists A_\omega \in \mathcal{X}$ s.t. $\nu(A_\omega^c) = 0$ and $g(x, Y_0(\omega))$ is constant on A_ω . The condition of $g(x, Y_0(\omega))$ being constant in x represents an entirely degenerate HMM in which the observations do not provide any information about the hidden state. Thus, we concentrate on the situation $\Upsilon_N(\mathbf{M}) > 0$. By an application of Proposition 5 in the case that $\psi^\omega(x) = 1$ for all ω and x , and using the bound (2.17) of Proposition 2, we find that there exists a constant $c < \infty$ such that

$$(4.6) \quad \frac{1}{n} \log \tilde{\mathcal{V}}_{n,N}^\omega \rightarrow \Upsilon_N(\mathbf{M}) \leq \log \left[1 + \frac{c}{N-1} \right],$$

where the convergence holds \mathbb{P} -almost surely. The practical importance of this result is that it shows why even the rather basic bootstrap filter is to be preferred over the SIS method in terms of variance growth behavior, as seen by comparing (4.6) with (4.4).

It should be noted that under our assumptions (H1) and (H2), the bound (4.6) is implied by the bound of [6], Theorem 5.1. The latter also provides important information about the nonasymptotic-in- n behavior of the relative variance, which our Proposition 5 does not. On the other hand, Proposition 5 applies not just to the standard particle transition \mathbf{M} , but also to twisted transitions, to which the analysis of [6] does not extend.

Continuing with the setting (4.5), and assuming that (H1) and (H2) hold, we shall now discuss h . The objects appearing in part (1) of Proposition 2 have the following interpretations: $\Phi_n^{\theta^{-n}\omega}(\mu) \equiv \pi_n^{\theta^{-n}\omega}$ is the prediction filter initialized at time $-n$ using μ , and run forward to time zero, thus conditioning on the observations $Y_{-n}(\omega), \dots, Y_{-1}(\omega)$. The quantity $Q_n^\omega(1)(x)$ is the conditional likelihood of observations $Y_0(\omega), \dots, Y_{n-1}(\omega)$ given that the hidden state in the HMM at time zero is x . Thus, if we denote by Π_n^ω the probability measure

$$\Pi_n^\omega(A) := \frac{\int_A \pi_n^{\theta^{-n}\omega}(dx) Q_n^\omega(1)(x)}{\int_{\mathcal{X}} \pi_n^{\theta^{-n}\omega}(dz) Q_n^\omega(1)(z)}, \quad A \in \mathcal{X},$$

we find by inspection of part (1) of Proposition 2 that h can be interpreted as the pointwise limit

$$(4.7) \quad h(\omega, x) \equiv \lim_{n \rightarrow \infty} \frac{d\Pi_n^\omega}{d\pi_n^{\theta^{-n}\omega}}(x).$$

Moreover, by part (2) of Proposition 2, we find that

$$(4.8) \quad \sup_{\omega, x} \left| \frac{d\Pi_n^\omega}{d\pi_n^{\theta^{-n}\omega}}(x) - h(\omega, x) \right| \leq C\rho^n$$

for some constants $C < \infty$ and $\rho \in (0, 1)$.

Let us now consider a twisted bootstrap particle filter (as per Section 1), in the case that for some fixed $\ell \geq 1$, we take $\psi^\omega := d\Pi_\ell^\omega/d\pi_\ell^{\theta^{-\ell}\omega}$, and as an instance of the setup in Section 3, we let $\tilde{\mathbf{M}}_\ell = \tilde{\mathbf{M}}$ be as per (3.1) with this choice of ψ^ω . We note that $\psi^\omega(x)$ is proportional to the conditional likelihood, under the HMM, of observations $Y_0(\omega), \dots, Y_{\ell-1}(\omega)$ given $X_0 = x$, and that Algorithm 2 can be implemented with ψ^ω only specified up to a constant of proportionality. Although typically unavailable in practice, this ψ^ω allows an illustrative application of Proposition 5. Indeed, using (4.8), and the fact that under the bounds of part (3) of Proposition 2 $h(\omega, x)$ is uniformly bounded above and below away from zero, elementary manipulations show that there exists some finite constant $C' < \infty$ such that

$$(4.9) \quad \Upsilon_N(\tilde{\mathbf{M}}_\ell) \leq \log \left[1 + \frac{C'\rho^\ell}{N-1} \right].$$

We see that, in principle, increasing the lag length ℓ is useful in helping to control $\Upsilon_N(\tilde{\mathbf{M}}_\ell)$.

Now under the mild regularity condition (H3), for fixed ω and n , and φ a bounded measurable function on \mathbf{X} , Lemma 4 shows that for the twisted particle filter,

$$(4.10) \quad N^{-1} \sum_{i=1}^N \varphi(\zeta_n^i) - \pi_n^\omega(\varphi) \rightarrow 0$$

as $N \rightarrow \infty$, with probability one, independently of ψ . Furthermore, by Theorem 2, $N^{-1/2} \sum_{i=1}^N [\varphi(\zeta_n^i) - \pi_n^\omega(\varphi)]$ converges in distribution to a centered Gaussian random variable with variance independent of ψ , that is, the same asymptotic variance obtained under the standard bootstrap particle filter.

4.3. *Auxiliary particle filters.* There exist many popular alternatives to the bootstrap particle filter. One such algorithm is the auxiliary particle filter (APF) [27], in which current and/or future observations can influence both the resampling and proposal of particles. In this section we consider a family of APFs which includes the “fully-adapted” version of [27]. Our presentation of the APF is similar to that of [15, 20].

In addition to the ingredients of the HMM given in Section 1, introduce $r : \Omega \times \mathbf{X} \rightarrow \mathbb{R}_+$ such that for each ω , $r^\omega(x)$ is strictly positive and bounded in x . We have in mind choosing r^ω to be $d\Pi_\ell^\omega/d\pi_\ell^{\theta^{-\ell}\omega}$ or some approximation thereof. Then set

$$(4.11) \quad G^\omega(x) := \frac{g(x, Y_0(\omega)) \int_{\mathbf{X}} r^{\theta\omega}(z) f(x, dz)}{r^\omega(x)},$$

$$M^\omega(x, dx') \propto f(x, dx') r^{\theta\omega}(x').$$

In this case, sampling according to \mathbf{M}^ω given by (2.20) amounts to a form of APF. More specifically, let $\{\mu_0^\omega \in \mathcal{P}(\mathbf{X}); \omega \in \Omega\}$ be the family of probability measures such that $\mu_0^\omega(dx) \propto r^\omega(x)\mu_0(dx)$, where μ_0 is the initial distribution in the HMM, as in Section 1. Then sampling

$$(4.12) \quad \zeta_0 \sim (\mu_0^\omega)^{\otimes N}, \quad \zeta_n | \zeta_{n-1} \sim \mathbf{M}^{\theta^{n-1}\omega}(\zeta_{n-1}, \cdot)$$

(we leave it to the reader to write out the algorithmic details), it is straightforward to check using (2.5), (2.9) and (2.23) that

$$(4.13) \quad \check{Z}_n^{\omega, N} := \mu_0(r^\omega) \left(\frac{1}{N} \sum_{i=1}^N \frac{1}{r^{\theta^n \omega}(\zeta_n^i)} \right) \prod_{p=0}^{n-1} \mathbf{G}^{\theta^p \omega}(\zeta_p)$$

is an unbiased estimator of Z_n^ω . If r is bounded above and below away from zero, and (H1) and (H2) hold, then Proposition 4 may be applied to establish the existence of $\Upsilon_N(\mathbf{M}) \geq 0$ such that the following convergence holds \mathbb{P} -almost surely:

$$(4.14) \quad \frac{1}{n} \log \frac{\mathbb{E}_N^\omega [(\check{Z}_n^{\omega, N})^2]}{(Z_n^\omega)^2} \longrightarrow \Upsilon_N(\mathbf{M}),$$

since the $\mu_0(r^\omega)$ and $N^{-1} \sum_{i=1}^N [r^{\theta^n \omega}(\zeta_n^i)]^{-1}$ terms have no asymptotic contribution and since the convergence in Proposition 4 is independent of the distribution from which the particle system is initialized.

In the particular case of taking $r(\omega, x) := g(x, Y_0(\omega))$, inspection of (4.11) shows that we obtain the “fully adapted” APF [27]. Moreover, Lemma 3 then shows that $\Upsilon_N(\mathbf{M}) = 0$ if and only if for \mathbb{P} -almost all ω , $\int_X g(z, Y_1(\omega)) f(x, dz)$ is ν -almost everywhere a constant. Outside of this kind of degenerate scenario, our analysis does not reveal whether this constant $\Upsilon_N(\mathbf{M})$ for the APF is bigger or smaller than the counterpart constant for the bootstrap filter in (4.6) or the constant for the twisted bootstrap filter on the l.h.s. of (4.9) (with ℓ fixed and finite). Moreover, our analysis does not reveal whether such orderings are invariant to the ingredients of the underlying HMM or other elements of our setup such as the law of the observation process, \mathbb{P} . Exploratory numerical experiments suggest such invariance does not hold in general—see Section 4.4.

We can say something, however, about an “ideal” APF, arising through a particular choice of r . We have seen in (4.9) that taking $\ell \rightarrow \infty$ in this twisted bootstrap filter, we can push the variance growth rate to zero. There is an APF which performs equally well in that sense; if we choose

$$(4.15) \quad r(\omega, x) := \lim_{n \rightarrow \infty} \frac{d\Pi_n^\omega}{d\pi_n^{\theta^{-n}\omega}}(x),$$

that is, the generalized eigenfunction for the kernel $g(x, Y_0(\omega)) f(x, dx')$, then

$$(4.16) \quad \int_X g(x, Y_0(\omega)) f(x, dx') r^{\theta \omega}(x') = \chi_\omega r^\omega(x)$$

for a nonnegative random variable χ . Applying (4.16) to (4.11), we find that G^ω appearing therein is constant in x . By Lemma 3, the constant on the r.h.s. of (4.14) then satisfies $\Upsilon_N(\mathbf{M}) = 0$.

We can also point out a difference in how the APF and the twisted bootstrap filter may be used to approximate integrals with respect to the prediction filters $\{\pi_n^\omega; n \geq 0\}$. If for some test function φ , one wishes to use $N^{-1} \sum_{i=1}^N \varphi(\zeta_n^i)$ to approximate $\pi_n^\omega(\varphi)$ in a $N \rightarrow \infty$ consistent manner, then, in general and in contrast to (4.10), some re-weighting must be applied to the particles. For example, assuming (H3) holds with G as in (4.11) and $r^\omega(x)$ is bounded below away from zero in x , Lemma 4 and some elementary manipulations involving Φ_n^ω show that for bounded measurable φ ,

$$(4.17) \quad \frac{\sum_{i=1}^N \varphi(\zeta_n^i) / r^{\theta^n \omega}(\zeta_n^i)}{\sum_{i=1}^N 1 / r^{\theta^n \omega}(\zeta_n^i)} - \pi_n^\omega(\varphi) \rightarrow 0$$

as $N \rightarrow \infty$, with probability 1 under the law of the particle system specified by (4.11)–(4.12). Numerical experiments (see Section 4.4, Figure 3) indicate that the variance of the APF estimator in (4.17) may be larger than that of the bootstrap estimator (4.10). This is perhaps attributable to the weighting of the particles in (4.17).

From a practical point of view, it should be noted that the computational cost of the twisted bootstrap filter and the APF are, in general, different: in the former, $N - 1$ of the particles are propagated using the HMM kernel f , whereas in the APF, all N particles are propagated using the generally more complicated kernel in (4.11). The difference in computational cost may, however, be rather dependent on the particular model treated and the specific techniques of simulation.

Last we note that, upon assuming the setting (4.11) and then following the generic structure of Section 3, twisted auxiliary particle filters can readily be devised. The idea of “twisting” is equally applicable to several other families of sequential Monte Carlo algorithms.

4.4. Numerical illustrations. In order to give some impression of the practical performance of the algorithms we have analyzed, we now present some numerical findings. (H2) is not satisfied for the M and G which specify the models below; in this section some of our theoretical results can only be used as guidelines for the design of practical algorithms. We note, however, that the much milder regularity condition (H3) is satisfied for the models we consider and, thus, Lemma 4 and Theorem 2 apply to the particle systems in question.

We shall first consider the influence of ψ on the variance growth behavior of the twisted bootstrap particle filter (henceforth TPF). The purpose of this example is to illustrate an idealized scenario in which $\psi^\omega := d\Pi_\ell^\omega/d\pi_\ell^{\theta^{-\ell}\omega}$ can be computed exactly. Consider a linear-Gaussian state-space model where $X_n = 0.9X_{n-1} + V_n$, $Y_n = X_n + W_n$, where $(V_n), (W_n)$ are i.i.d. zero mean, unit variance Gaussian sequences. Note that the TPF algorithm can be implemented with ψ^ω only known up to a constant of proportionality. Figure 1 shows variance growth behavior estimated empirically using 10,000 independent runs of the algorithm for a single observation sequence, which was drawn from the model and then fixed. Convergence of $n^{-1} \log \tilde{V}_{n,N}^\omega$ is apparent and the influence of ℓ on the rate of variance growth is substantial.

We now turn to a standard stochastic volatility model, in which $d\Pi_\ell^\omega/d\pi_\ell^{\theta^{-\ell}\omega}$ is unavailable in closed form, but for which a standard deterministic (henceforth, “the”) approximation is available. For details of the model, the approximation and the real data set of daily returns on pound/dollar exchange rates, see [17] and the references therein. We tested the TPF and APF using this data set and the same model parameter settings as in the aforementioned paper. We took both ψ^ω (for the TPF) and r^ω [for the APF as in (4.11)] to both be the approximation of $d\Pi_\ell^\omega/d\pi_\ell^{\theta^{-\ell}\omega}$.

Figure 2 shows empirical variance growth behavior for a range of values of ℓ , estimated from 10,000 independent runs of each algorithm. For both algorithms, increasing ℓ appears to generally yield a decrease in variance. The figures indicate that, apart from occasional fluctuations, the APF mostly exhibits lower variance

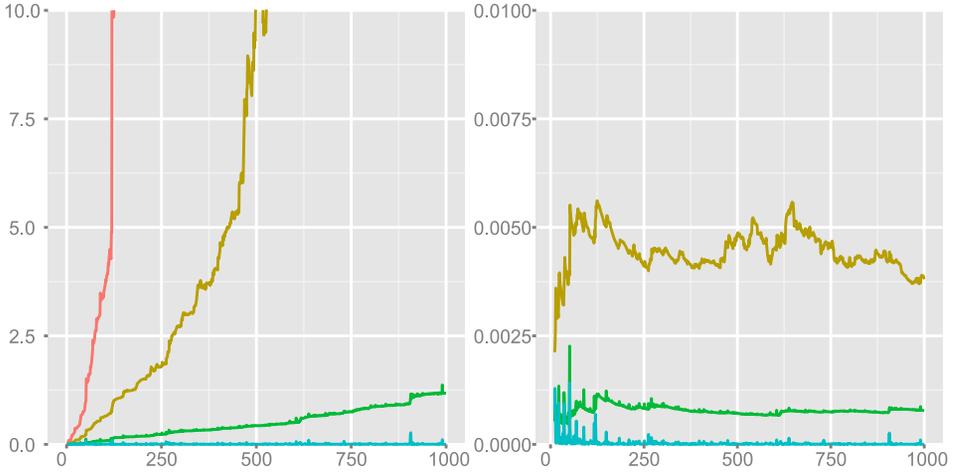


FIG. 1. *Linear-Gaussian model and the TPF. Estimated values of $\tilde{V}_{n,N}^\omega - 1$ (left) and $n^{-1} \log \tilde{V}_{n,N}^\omega$ (right) against n ; with $N = 100$ and $\ell = 0$ (red), $\ell = 1$ (yellow), $\ell = 2$ (green) and $\ell = 5$ (cyan). The $\ell = 0$ plot is omitted from the right-hand figure due to scale constraints. To the precision of these figures, increasing the lag beyond $\ell = 5$ had no noticeable influence on the variance.*

than the TPF, however, we found this phenomenon to be dependent on model parameter settings, for other parameter values we found the TPF exhibited lower variance than the APF (not shown).

The left plot in Figure 3 illustrates how the variance of the estimates from the TPF varies with N . An increase in variance growth rate is evident as N is de-

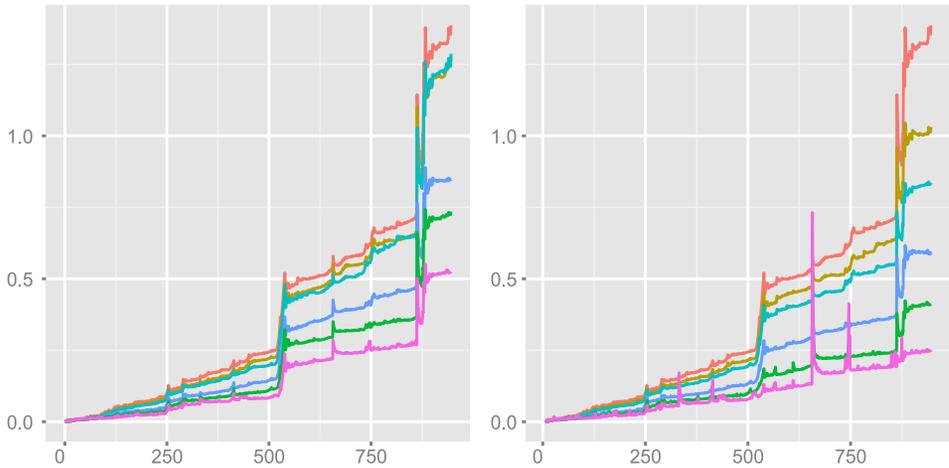


FIG. 2. *Stochastic volatility model. Estimated values of $\tilde{V}_{n,N}^\omega - 1$ against n , for the TPF (left) and APF (right); with $N = 1000$ and $\ell = 0$ (red), $\ell = 1$ (yellow), $\ell = 2$ (cyan), $\ell = 5$ (blue), $\ell = 10$ (green) and $\ell = 50$ (violet).*

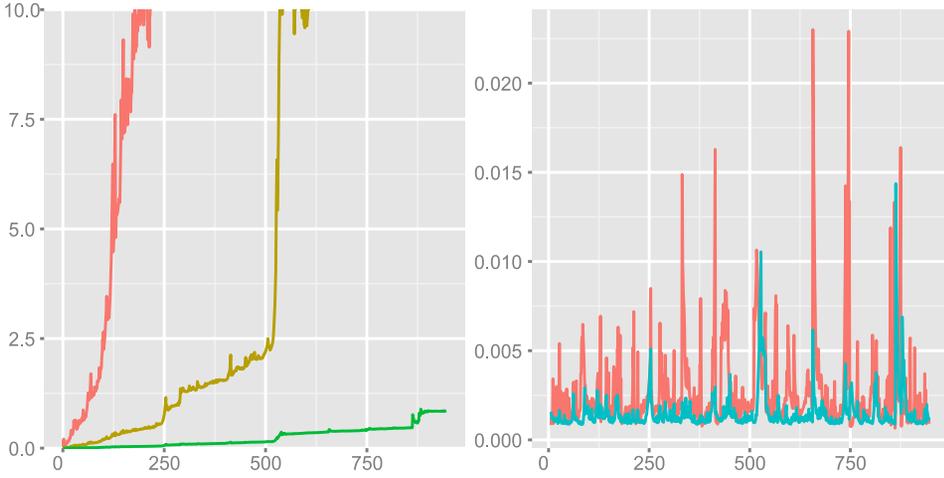


FIG. 3. Stochastic volatility model. Left: estimated values of $\tilde{V}_{n,N}^\omega - 1$ against n for the TPF for $\ell = 5$ and $N = 10$ (red), $N = 100$ (yellow) and $N = 1000$ (green). Right: empirical variance of particle approximations of the mean of π_n^ω against n , with $N = 1000$ and $\ell = 5$, for the TPF (cyan) and APF (red).

creased. The right plot of Figure 3 shows the empirical variance of particle estimates of the mean of the prediction filter π_n^ω against n , obtained from the TPF and APF both with $\ell = 5$. It is notable that here the TPF generally exhibits lower variance than the APF. Results for the standard bootstrap particle filter were found to be identical to those for the TPF on the scale of this figure, which is in agreement with the conclusions of Theorem 2 applied to the TPF, that is, that the asymptotic variance of prediction filter approximations is independent of ψ .

4.5. *Generalizations and extensions.* We have only mentioned the multinomial resampling scheme, appearing implicitly in the definition of \mathbf{M} given in (2.20). Several alternative schemes are popular in practice. In order to develop extensions of Proposition 4 and Theorem 1 to alternative schemes (assuming resampling is applied at every time step and with a fixed number of particles), it suffices to redefine \mathbf{M} appropriately so that it incorporates the resampling scheme of interest, to check that the conditions in the statement of Lemma 1 are satisfied, and to check that Lemma 2 holds with \mathbf{Q} redefined in terms of this new \mathbf{M} . Of course, the new \mathbf{M} will influence the form of the corresponding twisted algorithms.

Some types of standard particle algorithms and variants of the APF resample according to weights which depend on two or more historical components of the trajectory of each particle. Such algorithms can be incorporated into the framework presented here by a simple state-space augmentation. For example, starting from each Markov kernel M^ω on $(\mathbf{X}, \mathcal{X})$ (according to which particles are sampled in the algorithm of interest), one builds a kernel, $\overline{M}^\omega(x, dz) := \delta_{x_2}(dz_1)M^\omega(z_1, dz_2)$ on $(\mathbf{X}^2, \mathcal{X}^{\otimes 2})$, where $x = (x_1, x_2), z = (z_1, z_2)$ are points in \mathbf{X}^2 , and introduces

the appropriate incremental weight $\overline{G}^\omega(x)$. Then the analyses of Section 2 can be repeated with mostly superficial differences: when M^ω satisfies (2.4), then \overline{M}^ω satisfies a 2-step version of the same condition; one then works on $(\mathcal{X}^2, \mathcal{X}^{\otimes 2})$ instead of $(\mathcal{X}, \mathcal{X})$, dealing with the kernel $\overline{Q}^\omega(x, dz) := \overline{G}^\omega(x)\overline{M}^\omega(x, dz)$ instead of Q^ω .

Last, we note that Proposition 4 can be easily generalized from dealing with the second moment to any $1 + \delta$ moment ($\delta \geq 0$), subject to suitable redefinition of \mathbb{M} .

Acknowledgment. The authors thank the Associate Editor and referees for helpful comments and suggestions.

SUPPLEMENTARY MATERIAL

Twisted particle filters (DOI: [10.1214/13-AOS1167SUPP](https://doi.org/10.1214/13-AOS1167SUPP); .pdf). This supplement contains proofs of Lemmas 3–4, Propositions 1–5 and Theorems 1–2.

REFERENCES

- [1] ANDRIEU, C., DOUCET, A. and HOLENSTEIN, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** 269–342. [MR2758115](#)
- [2] ANDRIEU, C. and ROBERTS, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.* **37** 697–725. [MR2502648](#)
- [3] ANDRIEU, C. and VIHOLA, M. (2012). Convergence properties of pseudo-marginal Markov chain Monte Carlo algorithms. Available at [arXiv:1210.1484v1](https://arxiv.org/abs/1210.1484v1) [math.PR].
- [4] ATHREYA, K. B. (2000). Change of measures for Markov chains and the $L \log L$ theorem for branching processes. *Bernoulli* **6** 323–338. [MR1748724](#)
- [5] BUCKLEW, J. A., NEY, P. and SADOWSKY, J. S. (1990). Monte Carlo simulation and large deviations theory for uniformly recurrent Markov chains. *J. Appl. Probab.* **27** 44–59. [MR1039183](#)
- [6] CÉROU, F., DEL MORAL, P. and GUYADER, A. (2011). A nonasymptotic theorem for unnormalized Feynman–Kac particle models. *Ann. Inst. Henri Poincaré Probab. Stat.* **47** 629–649. [MR2841068](#)
- [7] CHOPIN, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.* **32** 2385–2411. [MR2153989](#)
- [8] DEL MORAL, P. (2004). *Feynman–Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer, New York. [MR2044973](#)
- [9] DEL MORAL, P. and GUIONNET, A. (1999). Central limit theorem for nonlinear filtering and interacting particle systems. *Ann. Appl. Probab.* **9** 275–297. [MR1687359](#)
- [10] DEL MORAL, P. and GUIONNET, A. (2001). On the stability of interacting processes with applications to filtering and genetic algorithms. *Ann. Inst. Henri Poincaré Probab. Stat.* **37** 155–194. [MR1819122](#)
- [11] DEL MORAL, P. and JACOD, J. (2001). Interacting particle filtering with discrete time observations: Asymptotic behaviour in the Gaussian case. In *Stochastics in Finite and Infinite Dimensions: In Honor of Gopinath Kallianpur* (T. Hida, R. L. Karandikar, H. Kunita, B. S. Rajput, S. Watanabe and J. Xiong, eds.). Birkhäuser, Basel.
- [12] DOOB, J. L. (1994). *Measure Theory. Graduate Texts in Mathematics* **143**. Springer, New York. [MR1253752](#)
- [13] DOUC, R. and MOULINES, E. (2008). Limit theorems for weighted samples with applications to sequential Monte Carlo methods. *Ann. Statist.* **36** 2344–2376. [MR2458190](#)

- [14] DOUC, R. and MOULINES, E. (2011). Asymptotic properties of the maximum likelihood estimation in misspecified Hidden Markov models. Available at [arXiv:1110.0356](https://arxiv.org/abs/1110.0356).
- [15] DOUC, R., MOULINES, É. and OLSSON, J. (2009). Optimality of the auxiliary particle filter. *Probab. Math. Statist.* **29** 1–28. [MR2552996](https://doi.org/10.1007/s11464-009-9296-6)
- [16] DOUC, R., MOULINES, E. and OLSSON, J. (2012). Long-term stability of sequential Monte Carlo methods under verifiable conditions. Available at [arXiv:1203.6898](https://arxiv.org/abs/1203.6898).
- [17] DOUCET, A., BRIERS, M. and SÉNÉCAL, S. (2006). Efficient block sampling strategies for sequential Monte Carlo methods. *J. Comput. Graph. Statist.* **15** 693–711. [MR2280152](https://doi.org/10.1198/106186006000000152)
- [18] FAVETTO, B. (2012). On the asymptotic variance in the central limit theorem for particle filters. *ESAIM Probab. Stat.* **16** 151–164. [MR2946125](https://doi.org/10.1051/ps/2012/16151)
- [19] GORDON, N. J., SALMOND, D. J. and SMITH, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *Radar and Signal Processing, IEE Proceedings F* **140** 107–113.
- [20] JOHANSEN, A. M. and DOUCET, A. (2008). A note on auxiliary particle filters. *Statist. Probab. Lett.* **78** 1498–1504. [MR2528342](https://doi.org/10.1080/00107170802528342)
- [21] KIFER, Y. (1996). Perron–Frobenius theorem, large deviations, and random perturbations in random environments. *Math. Z.* **222** 677–698. [MR1406273](https://doi.org/10.1007/BF01171406)
- [22] KÜNSCH, H. R. (2005). Recursive Monte Carlo filters: Algorithms and theoretical analysis. *Ann. Statist.* **33** 1983–2021. [MR2211077](https://doi.org/10.1214/00905360500000011077)
- [23] LE GLAND, F. and OUDJANE, N. (2004). Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters. *Ann. Appl. Probab.* **14** 144–187. [MR2023019](https://doi.org/10.1214/00905650400000023019)
- [24] LEGLAND, F. and OUDJANE, N. (2003). A robustification approach to stability and to uniform particle approximation of nonlinear filters: The example of pseudo-mixing signals. *Stochastic Process. Appl.* **106** 279–316. [MR1989630](https://doi.org/10.1016/S0304-4149(03)00163-0)
- [25] LEROUX, B. G. (1992). Maximum-likelihood estimation for hidden Markov models. *Stochastic Process. Appl.* **40** 127–143. [MR1145463](https://doi.org/10.1016/0304-4149(92)90046-3)
- [26] OUDJANE, N. and RUBENTHALER, S. (2005). Stability and uniform particle approximation of nonlinear filters in case of non ergodic signals. *Stoch. Anal. Appl.* **23** 421–448. [MR2140972](https://doi.org/10.1016/j.sna.2004.10.072)
- [27] PITT, M. K. and SHEPHARD, N. (1999). Filtering via simulation: Auxiliary particle filters. *J. Amer. Statist. Assoc.* **94** 590–599. [MR1702328](https://doi.org/10.1198/0162145990000002328)
- [28] VAN HANDEL, R. (2009). Uniform time average consistency of Monte Carlo particle filters. *Stochastic Process. Appl.* **119** 3835–3861. [MR2552307](https://doi.org/10.1016/j.spsa.2009.05.007)
- [29] WHITELEY, N. (2013). Stability properties of some particle filters. *Ann. Appl. Probab.* **23** 2500–2537. [MR3127943](https://doi.org/10.1214/12-AOS1167)
- [30] WHITELEY, N. and LEE, A. (2014). Supplement to “Twisted particle filters.” DOI:10.1214/13-AOS1167SUPP.

SCHOOL OF MATHEMATICS
UNIVERSITY OF BRISTOL
UNIVERSITY WALK
BRISTOL BS8 1TW
UNITED KINGDOM
E-MAIL: nick.whiteley@bristol.ac.uk

DEPARTMENT OF STATISTICS
UNIVERSITY OF WARWICK
COVENTRY CV4 7AL
UNITED KINGDOM
E-MAIL: anthony.lee@warwick.ac.uk