SIAM J. CONTROL OPTIM.
Vol. 53, No. 3, pp. 1278–1304

UNIFORM STABILITY OF A PARTICLE APPROXIMATION OF THE OPTIMAL FILTER DERIVATIVE*

PIERRE DEL MORAL†, ARNAUD DOUCET‡, AND SUMEETPAL S. SINGH§

Abstract. Particle methods, also known as Sequential Monte Carlo methods, are a principled set of algorithms used to approximate numerically the optimal filter in nonlinear non-Gaussian state-space models. However, when performing maximum likelihood parameter inference in state-space models, it is also necessary to approximate the derivative of the optimal filter with respect to the parameter of the model. References [G. Poyiadjis, A. Doucet, and S. S. Singh, *Particle methods for optimal filter derivative: Application to parameter estimation*, in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 5, Philadelphia, 2005, pp. 925–928 and G. Poyiadjis, A. Doucet, and S. S. Singh, *Biometrika*, 98 (2011), pp. 65–80] present an original particle method to approximate this derivative, and it was shown in numerical examples to be numerically stable in the sense that it did not deteriorate over time. In this paper we theoretically substantiate this claim. $\mathbb{L}_p$ bounds and a central limit theorem for this particle approximation are presented. Under mixing conditions these $\mathbb{L}_p$ bounds and the asymptotic variance are uniformly bounded with respect to the time index.

Key words. hidden Markov models, state-space models, sequential Monte Carlo, smoothing, filter derivative, recursive maximum likelihood

AMS subject classifications. 60J05, 62M05, 80M31, 93E11, 93E12

DOI. 10.1137/140993703

1. Introduction. State-space models are a very popular class of nonlinear and non-Gaussian time series models in control, signal processing, and statistics; see, for example, [10] and [12]. A state-space model is comprised of a pair of discrete-time stochastic processes, $\{X_n\}_{n\geq 0}$ and $\{Y_n\}_{n\geq 0}$, where the former is an $\mathcal{X}$-valued unobserved process and the latter is a $\mathcal{Y}$-valued process, which is observed. The hidden process $\{X_n\}_{n\geq 0}$ is a Markov process with initial law $\pi_\theta(x)\,\mathrm{d}x$ and time homogeneous transition law $f_\theta(x'|\,x)\,\mathrm{d}x'$; i.e.,

$$(1.1)\quad X_0 \sim \pi_\theta(x_0)\,\mathrm{d}x_0 \quad\text{and}\quad X_n|\,(X_{n-1}=x_{n-1}) \sim f_\theta(x_n|\,x_{n-1})\,\mathrm{d}x_n, \qquad n\geq 1.$$

It is assumed that the observations $\{Y_n\}_{n\geq 0}$ conditioned upon $\{X_n\}_{n\geq 0}$ are statistically independent and have marginal laws

$$(1.2)\qquad Y_n|\left(\{X_k\}_{k\geq 0}=\{x_k\}_{k\geq 0}\right) \sim g_\theta(y_n|\,x_n)\,\mathrm{d}y_n.$$

Here $\pi_\theta(x)$, $f_\theta(x|\,x')$, and $g_\theta(y|\,x)$ are densities with respect to (w.r.t.) suitable dominating measures denoted generically as $\mathrm{d}x$ and $\mathrm{d}y$. For example, if $\mathcal{X}\subseteq\mathbb{R}^p$

*Received by the editors October 30, 2014; accepted for publication December 15, 2014; published electronically May 19, 2015.
http://www.siam.org/journals/sicon/53-3/99370.html
†School of Mathematics and Statistics, University of New South Wales, Sydney, NSW 2052, Australia (p.del-moral@unsw.edu.au).
‡Department of Statistics, Oxford University, Oxford OX1 3TG, UK (doucet@stats.ox.ac.uk). This author's research was supported by the Engineering and Physical Sciences Research Council (grant EP/K000276/1, EP/K009850/1) and by the Air Force Office of Scientific Research/Asian Office of Aerospace Research and Development (AFOSR/AOARD) (grant AOARD-144042).
§Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK (sss40@cam.ac.uk). This author's research was supported by the Engineering and Physical Sciences Research Council (grant EP/G037590/1).

and $\mathcal{Y} \subseteq \mathbb{R}^q$, then the dominating measures could be the Lebesgue measures. The variable θ in the densities are the particular parameters of the model. The set of possible values for θ, denoted Θ, is assumed to be an open subset of $\mathbb{R}^d$. The model (1.1)–(1.2) is also often referred to as a hidden Markov model in the literature.

For a sequence $\{z_n\}_{n\geq 0}$ and integers i, j, let $z_{i:j}$ denote the set $\{z_i, z_{i+1}, \ldots, z_j\}$, which is empty if $j < i$. Equations (1.1) and (1.2) define the law of $(X_{0:n}, Y_{0:n-1})$, which is given by the measure $p_\theta(\mathrm{d}x_{0:n}, \mathrm{d}y_{0:n-1})$ of density w.r.t. $\mathrm{d}x_{0:n}\mathrm{d}y_{0:n-1}$,

$$p_\theta(x_{0:n}, y_{0:n-1}) = \pi_\theta(x_0) \prod_{k=1}^{n} f_\theta(x_k | x_{k-1}) \prod_{k=0}^{n-1} g_\theta(y_k | x_k), \tag{1.3}$$

from which the probability density of the observed process is obtained:

$$p_\theta(y_{0:n-1}) = \int p_\theta(x_{0:n}, y_{0:n-1})\mathrm{d}x_{0:n}. \tag{1.4}$$

For a realization of observations $Y_{0:n-1} = y_{0:n-1}$, let $\mathbb{Q}_{\theta,n}$ denote the law of $X_{0:n}$ conditioned on this sequence of observed variables whose density is obtained by dividing (1.3) by (1.4). Let $\eta_{\theta,n}$ denote the time n marginal of $\mathbb{Q}_{\theta,n}$. This marginal, which we call the filter, may be computed recursively using Bayes' formula,

$$\eta_{\theta,n+1}(\mathrm{d}x_{n+1}) = \mathbb{Q}_{\theta,n+1}(\mathrm{d}x_{n+1}) = \frac{\mathrm{d}x_{n+1} \int \eta_{\theta,n}(\mathrm{d}x_n)\, g_\theta(y_n | x_n)\, f_\theta(x_{n+1} | x_n)}{\int \eta_{\theta,n}(\mathrm{d}x'_n)\, g_\theta(y_n | x'_n)}$$

for $n \geq 0$ and $\eta_{\theta,0} = \pi_\theta$ by convention. Except for simple models such as the linear Gaussian state-space model or when $\mathcal{X}$ is a finite set, it is impossible to compute $p_\theta(y_{0:n})$, $\mathbb{Q}_{\theta,n}(\mathrm{d}x_{0:n})$, or $\eta_{\theta,n}(\mathrm{d}x_n)$ exactly. Particle methods have been applied extensively to approximate these quantities for general state-space models of the form (1.1)–(1.2); see, for example, [3, 10, 12].

The particle approximation of $\mathbb{Q}_{\theta,n}(\mathrm{d}x_{0:n})$ is the empirical measure corresponding to a set of $N \geq 1$ random samples termed particles; that is,

$$\mathbb{Q}^{\mathrm{p},N}_{\theta,n}(\mathrm{d}x_{0:n}) = \frac{1}{N}\sum_{i=1}^{N} \delta_{X^{(i)}_{0:n}}(\mathrm{d}x_{0:n}), \tag{1.5}$$

where $\delta_z(\mathrm{d}z)$ denotes the Dirac measure located at z. This approximation is referred to as the *path space* approximation [3] and it is denoted by the superscript "p." The particle approximation of $\eta_{\theta,n}(\mathrm{d}x_n)$ is obtained from $\mathbb{Q}^{\mathrm{p},N}_{\theta,n}(\mathrm{d}x_{0:n})$ by marginalization,

$$\eta^N_{\theta,n}(\mathrm{d}x_n) = \frac{1}{N}\sum_{i=1}^{N} \delta_{X^{(i)}_n}(\mathrm{d}x_n).$$

These particles are propagated in time using importance sampling and resampling steps; see [12] and [10] for a review of the literature. Specifically, $\mathbb{Q}^{\mathrm{p},N}_{\theta,n+1}(\mathrm{d}x_{0:n+1})$ is the empirical measure constructed from N independent samples from

$$\frac{\mathbb{Q}^{\mathrm{p},N}_{\theta,n}(\mathrm{d}x_{0:n})\, f_\theta(x_{n+1} | x_n)\, \mathrm{d}x_{n+1} g_\theta(y_n | x_n)}{\int \mathbb{Q}^{\mathrm{p},N}_{\theta,n}(\mathrm{d}x_{0:n})\, g_\theta(y_n | x_n)}. \tag{1.6}$$

It is a well-known fact that the particle approximation of $\mathbb{Q}_{\theta,n}(\mathrm{d}x_{0:n})$ becomes progressively impoverished as n increases because of the successive resampling steps [4, 18].

That is, the number of distinct particles representing the marginal $\mathbb{Q}^{\mathrm{p},N}_{\theta,n}(\mathrm{d}x_{0:k})$ for any fixed $k < n$ diminishes as n increases until it collapses to a single particle—this is known as the *particle path degeneracy* problem.

The focus of this paper is on the convergence properties of particle methods which have been recently proposed to approximate the derivative of the measures $\{\eta_{\theta,n}(\mathrm{d}x_n)\}_{n\geq 0}$ w.r.t. $\theta = [\theta_1, \dots \theta_d]^{\mathrm{T}} \in \mathbb{R}^d$:

$$\zeta_{\theta,n} = \nabla\eta_{\theta,n} = \left[\frac{\partial\eta_{\theta,n}}{\partial\theta_1}, \dots, \frac{\partial\eta_{\theta,n}}{\partial\theta_d}\right]^{\mathrm{T}}.$$

See section 2 for a precise definition of the derivative. References [1] and [11] present particle methods which have a computational complexity that scales linearly with the number N of particles. It was shown in [23] that the performance of these $\mathcal{O}(N)$ methods, which inherently rely on the particle approximations of $\{\mathbb{Q}_{\theta,n}(\mathrm{d}x_{0:n})\}_{n\geq 0}$ constructed as in (1.6) above, degraded over time and it was conjectured that this may be attributed to the particle path degeneracy problem. In contrast, the alternative method of [22] was shown in numerical examples to be stable. The method of [22] is a nonstandard particle implementation that avoids the particle path degeneracy problem at the expense of a computational complexity per time step which is quadratic in the number of particles, i.e., $\mathcal{O}(N^2)$; see section 2 for more details. Supported by numerical examples, it was conjectured in [23] that even under strong mixing assumptions the variance of the estimate of the filter derivative computed with the $\mathcal{O}(N)$ methods increases at least linearly in time, while that of the $\mathcal{O}(N^2)$ is uniformly bounded w.r.t. the time index. This conjecture is confirmed in this paper. Specifically, we analyze the $\mathcal{O}(N^2)$ implementation of [22] in section 3 and obtain results on the errors of the approximation; in particular, $\mathbb{L}_p$ bounds and a central limit theorem (CLT) are presented. We show that these $\mathbb{L}_p$ bounds and asymptotic variances appearing in the CLT are uniformly bounded w.r.t. the time index when the state-space model satisfies certain mixing assumptions. In contrast, the asymptotic variance of the $\mathcal{O}(N)$ implementations, which is also captured through the CLT, is shown to increase linearly. To the best of our knowledge, these are the first results of this kind.

An important application of our results, which is discussed in detail in section 4, is to the problem of estimating the parameters of the model (1.1)–(1.2) from observed data. The estimates of the model parameters are found by maximizing the likelihood function $p_\theta(y_{0:n})$ with respect to θ using a gradient ascent algorithm which relies on the particle approximation of the filter derivative. The results we present in section 3 have bearing on the performance of an online parameter estimation algorithm, which we illustrate with numerical examples in section 4. The appendix contains the proofs of the main results as well as that of some supporting auxiliary results. As a final remark, although the algorithms and theoretical results are presented for a state-space model, they may be reinterpreted for Feynman–Kac models as well.

1.1. Notation and definitions. We give some basic definitions from probability and operator semigroup theory. For a measurable space $(E, \mathcal{E})$ let $\mathcal{M}(E)$ denote the set of all finite signed measures and let $\mathcal{P}(E)$ be the set of all probability measures on E. The n-fold product space $E \times \cdots \times E$ is denoted by E^n. Let $\mathcal{B}(E)$ denote the Banach space of all bounded real-valued and measurable functions $\varphi : E \to \mathbb{R}$ equipped with the uniform norm $\|\varphi\| = \sup_{x\in E}|\varphi(x)|$. For $\nu \in \mathcal{M}(E)$ and $\varphi \in \mathcal{B}(E)$, let $\nu(\varphi) = \int \nu(\mathrm{d}x)\ \varphi(x)$ be the Lebesgue integral of φ w.r.t. ν. If ν is a density w.r.t. some dominating measure $\mathrm{d}x$ on E, then $\nu(\varphi) = \int \mathrm{d}x\ \nu(x)\ \varphi(x)$. We recall that a bounded integral kernel $M(x, \mathrm{d}x')$ from a measurable space $(E, \mathcal{E})$ into an auxiliary measurable space $(E', \mathcal{E}')$ is an operator $\varphi \mapsto M(\varphi)$ from $\mathcal{B}(E')$ into $\mathcal{B}(E)$ such that the functions

$$x \mapsto M(\varphi)(x) := \int_{E'} M(x, \mathrm{d}x')\varphi(x')$$

are $\mathcal{E}$-measurable and bounded for any $\varphi \in \mathcal{B}(E')$. The kernel M also generates a dual operator $\nu \mapsto \nu M$ from $\mathcal{M}(E)$ into $\mathcal{M}(E')$ defined by

$$(\nu M)(\varphi) := \nu(M(\varphi)).$$

Given a pair of bounded integral operators (M_1, M_2), we let $(M_1 M_2)$ be the composition operator defined by $(M_1 M_2)(\varphi) = M_1(M_2(\varphi))$.

A Markov kernel is a positive and bounded integral operator M such that $M(1)(x) = 1$ for any $x \in E$. For $\varphi \in \mathcal{B}(E)$, let

$$\mathrm{osc}(\varphi) = \sup_{x,x' \in E} |\varphi(x) - \varphi(x')|,$$

and let

$$\mathrm{Osc}_1(E) = \{\varphi \in \mathcal{B}(E) : \mathrm{osc}(\varphi) \leq 1\}.$$

Let $\beta(M) \in [0, 1]$ denote the Dobrushin coefficient of the Markov kernel M which is defined by the formula [3, Proposition 4.2.1]:

$$\beta(M) := \sup \ \{\mathrm{osc}(M(\varphi)) \ ; \ \ \varphi \in \mathrm{Osc}_1(E')\}.$$

If there exists a positive constant ρ and a probability measure ν such that the Markov kernel M satisfies

$$M(x, \mathrm{d}z) \geq \rho\nu(\mathrm{d}z) \text{ for all } x \in E, \quad \text{then } \beta(M) \leq 1 - \rho.$$

For two Markov kernels M_1, M_2, $\beta(M_1 M_2) \leq \beta(M_1)\beta(M_2)$.

Given a positive function G on E, let $\Psi_G : \nu \in \mathcal{P}(E) \mapsto \Psi_G(\nu) \in \mathcal{P}(E)$ be the probability distribution defined by

$$\Psi_G(\nu)(\mathrm{d}x) := \frac{\nu(\mathrm{d}x)G(x)}{\nu(G)}$$

provided $\infty > \nu(G) > 0$. The definitions above also apply if ν is a density and M is a transition density. In this case all instances of $\nu(\mathrm{d}x)$ should be replaced with $\nu(x)\mathrm{d}x$, and $M(x, \mathrm{d}x')$ replaced by $M(x, x')\mathrm{d}x'$ where $\mathrm{d}x$ and $\mathrm{d}x'$ is generic notation for the dominating measures.

It is convenient to introduce the following transition kernels:

$$Q_{\theta,n}(x_{n-1}, \mathrm{d}x_n) = g_\theta(y_{n-1}|x_{n-1}) f_\theta(x_n|x_{n-1})\mathrm{d}x_n = q_\theta(x_n|x_{n-1})\mathrm{d}x_n, \quad n > 0,$$
$$Q_{\theta,k,n}(x_k, \mathrm{d}x_n) = (Q_{\theta,k+1} Q_{\theta,k+2} \cdots Q_{\theta,n})(x_k, \mathrm{d}x_n), \quad 0 \leq k \leq n,$$

with the convention that $Q_{\theta,n,n} = Id$, the identity operator. Note that $Q_{\theta,k,n}(1)(x_k)$ is the density of the law of $Y_{k:n-1}$ given $X_k = x_k$. For $0 \leq p \leq n$, define the potential function $G_{\theta,p,n}$ on $\mathcal{X}$ to be

$$G_{\theta,p,n}(x_p) = Q_{\theta,p,n}(1)(x_p)/\eta_{\theta,p} Q_{\theta,p,n}(1). \tag{1.7}$$

Let the mapping $\Phi_{\theta,k,n} : \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{X})$, $0 \le k \le n$, be defined as follows:

$$\Phi_{\theta,k,n}(\nu)(\mathrm{d}x_n) = \frac{\nu Q_{\theta,k,n}(\mathrm{d}x_n)}{\nu Q_{\theta,k,n}(1)}.$$

It follows that $\eta_{\theta,n} = \Phi_{\theta,k,n}(\eta_{\theta,k})$. For conciseness, we also write $\Phi_{\theta,n-1,n}$ as $\Phi_{\theta,n}$.

A key quantity that facilitates the recursive computation of the derivative of $\eta_{\theta,n}$ is the following collection of backward Markov transition kernels:

$$M_{\theta,n}(x_n, \mathrm{d}x_{n-1}) = \frac{\eta_{\theta,n-1}(\mathrm{d}x_{n-1}) q_\theta(x_n|x_{n-1})}{\eta_{\theta,n-1}(q_\theta(x_n|\cdot))}, \quad n > 0. \tag{1.8}$$

Their particle approximations are

$$M^N_{\theta,n}(x_n, \mathrm{d}x_{n-1}) = \frac{\eta^N_{\theta,n-1}(\mathrm{d}x_{n-1}) q_\theta(x_n|x_{n-1})}{\eta^N_{\theta,n-1}(q_\theta(x_n|\cdot))}. \tag{1.9}$$

These backward Markov kernels are convenient for computing certain conditional expectations and probability measures. In particular, for $\varphi \in \mathcal{B}(\mathcal{X}^2)$, we have

$$\mathbb{E}_\theta \left[\varphi(X_{n-1}, X_n) \middle| y_{0:n-1}, x_n \right] = \int M_{\theta,n}(x_n, \mathrm{d}x_{n-1}) \varphi(x_{n-1}, x_n),$$

and the conditional law of $X_{0:n-1}$ given $X_n = x_n$ and $Y_{0:n-1} = y_{0:n-1}$ can be expressed as $M_{\theta,n}(x_n, \mathrm{d}x_{n-1}) \cdots M_{\theta,1}(x_1, \mathrm{d}x_0)$.

Finally, the following two definitions are needed for the CLT of the particle approximation of the derivative of $\eta_{\theta,n}$. The bounded integral operator $D_{\theta,k,n}$ from $\mathcal{X}$ into $\mathcal{X}^{n+1}$ is defined for any $F_n \in \mathcal{B}(\mathcal{X}^{n+1})$ and $0 \le k \le n$ by

$$D_{\theta,k,n}(F_n)(x_k) := \int \left(\prod_{j=k}^{1} M_{\theta,j}(x_j, \mathrm{d}x_{j-1}) \right) \left(\prod_{j=k}^{n-1} Q_{\theta,j+1}(x_j, \mathrm{d}x_{j+1}) \right) F_n(x_{0:n}), \tag{1.10}$$

with the convention that $\prod \emptyset = 1$. The particle approximation, $D^N_{\theta,k,n}$, is defined to be

$$D^N_{\theta,k,n}(F_n)(x_k) := \int \left(\prod_{j=k}^{1} M^N_{\theta,j}(x_j, \mathrm{d}x_{j-1}) \right) \left(\prod_{j=k}^{n-1} Q_{\theta,j+1}(x_j, \mathrm{d}x_{j+1}) \right) F_n(x_{0:n}). \tag{1.11}$$

To be concise we write

$$\eta_{\theta,k}(\mathrm{d}x_k) D_{\theta,k,n}(x_k, \mathrm{d}x_{0:k-1}, \mathrm{d}x_{k+1:n}) \quad \text{as} \quad \eta_{\theta,k} D_{\theta,k,n}(\mathrm{d}x_{0:n}),$$

and adopt similar notation for the particle versions. Although convention dictates that $\eta_{\theta,k} D_{\theta,k,n}$ should be understood as the measure $(\eta_{\theta,k} D_{\theta,k,n})(\mathrm{d}x_{0:k-1}, \mathrm{d}x_{k+1:n})$, when we mean otherwise it should be clear from the infinitesimal neighborhood under consideration.

2. Computing the filter derivative. For any $F_n \in \mathcal{B}(\mathcal{X}^{n+1})$, we have

$$
\begin{aligned}
\nabla \mathbb{Q}_{\theta,n}(F_n) &= \frac{1}{p_\theta\left(y_{0:n-1}\right)} \int F_n(x_{0:n}) \nabla p_\theta(x_{0:n}, y_{0:n-1}) \mathrm{d}x_{0:n} \\
&\quad - \frac{1}{p_\theta\left(y_{0:n-1}\right)} \mathbb{E}_\theta\left\{ F_n(X_{0:n}) \middle| y_{0:n-1} \right\} \int \nabla p_\theta(x_{0:n}, y_{0:n-1}) \mathrm{d}x_{0:n} \\
&= \mathbb{E}_\theta\left\{ F_n(X_{0:n}) T_{\theta,n}(X_{0:n}) \middle| y_{0:n-1} \right\} \\
&\quad - \mathbb{E}_\theta\left\{ F_n(X_{0:n}) \middle| y_{0:n-1} \right\} \mathbb{E}_\theta\left\{ T_{\theta,n}(X_{0:n}) \middle| y_{0:n-1} \right\}, \qquad (2.1)
\end{aligned}
$$

where

$$T_{\theta,n}(x_{0:n}) = \sum_{k=0}^{n} t_{\theta,k}(x_{k-1}, x_k), \qquad (2.2)$$

$$t_{\theta,k}(x_{k-1}, x_k) = \nabla \log \left(g_\theta\left(y_{k-1} \middle| x_{k-1} \right) f_\theta\left(x_k \middle| x_{k-1} \right) \right), \quad k > 0, \qquad (2.3)$$

$$t_{\theta,0}(x_{-1}, x_0) = t_{\theta,0}(x_0) = \nabla \log \pi_\theta\left(x_0\right). \qquad (2.4)$$

The first equality in (2.1) follows from the definition of $\mathbb{Q}_{\theta,n}$ and interchanging the order of differentiation and integration. The interchange is permissible under certain regularity conditions [20]; e.g., a sufficient condition would be the main assumption in section 3 under which the uniform stability results are proved. The second equality follows from a change of measure, which then permits an importance sampling-based estimator for the derivative of $\mathbb{Q}_{\theta,n}$; this is the well-known score method; e.g., see [20, section 4.2.1]. For any $\varphi_n \in \mathcal{B}(\mathcal{X})$, it follows by setting $F_n(x_{0:n}) = \varphi_n(x_n)$ in (2.1) that

$$
\begin{aligned}
&\nabla \int \eta_{\theta,n}(\mathrm{d}x_n) \varphi_n(x_n) \\
&= \mathbb{E}_\theta\left\{ \varphi_n(X_n) T_{\theta,n}(X_{0:n}) \middle| y_{0:n-1} \right\} - \mathbb{E}_\theta\left\{ \varphi_n(X_n) \middle| y_{0:n-1} \right\} \mathbb{E}_\theta\left\{ T_{\theta,n}(X_{0:n}) \middle| y_{0:n-1} \right\} \\
&= \int \zeta_{\theta,n}(\mathrm{d}x_n) \varphi_n(x_n),
\end{aligned}
$$

where

$$\zeta_{\theta,n}(\mathrm{d}x_n) = \eta_{\theta,n}(\mathrm{d}x_n) \left\{ \mathbb{E}_\theta\left[T_{\theta,n}\left(X_{0:n}\right) \middle| y_{0:n-1}, x_n \right] - \mathbb{E}_\theta\left[T_{\theta,n}\left(X_{0:n}\right) \middle| y_{0:n-1} \right] \right\}. \qquad (2.5)$$

We call $\zeta_{\theta,n}$ the derivative of $\eta_{\theta,n}$.

Given the particle approximation (1.5) of $\mathbb{Q}_{\theta,n}(\mathrm{d}x_{0:n})$, it is straightforward to construct a particle approximation of $\zeta_{\theta,n}(\mathrm{d}x_n)$:

$$\zeta_{\theta,n}^{\mathrm{p},N}(\mathrm{d}x_n) = \sum_{i=1}^{N} \frac{1}{N} \left\{ T_{\theta,n}(X_{0:n}^{(i)}) - \frac{1}{N} \sum_{j=1}^{N} T_{\theta,n}(X_{0:n}^{(j)}) \right\} \delta_{X_n^{(i)}}\left(\mathrm{d}x_n\right). \qquad (2.6)$$

This approximation is also referred to as the path space method. Such approximations were implicitly proposed in [1, 11], and there are several reasons why this estimate appears attractive. First, even with the resampling steps in the construction of $\mathbb{Q}_{\theta,n}^{\mathrm{p},N}$, $\zeta_{\theta,n}^{\mathrm{p},N}$ can be computed recursively. Second, there is no need to store the entire ancestry of each particle, i.e., $X_{0:n}^{(i)}$ for $1 \leq i \leq N$, and thus the memory requirement to construct $\zeta_{\theta,n}^{\mathrm{p},N}$ is constant over time. Third, the computational cost per time is

$\mathcal{O}(N)$. However, as $\mathbb{Q}^{\mathrm{p},N}_{\theta,n}$ suffers from the particle path degeneracy problem, we expect the approximation $\zeta^{\mathrm{p},N}_{\theta,n}$ to worsen over time. This was indeed observed in numerical examples in [23], and it was conjectured that the asymptotic variance (i.e., as $N \to \infty$) of $\zeta^{\mathrm{p},N}_{\theta,n}$ for bounded integrands would increase linearly with n even under strong mixing assumptions. This is now proven in this paper.

An alternative particle method to approximate $\{\zeta_{\theta,n}\}_{n\geq 0}$ has been proposed in [22, 23]. We now reinterpret this method using the representation in (2.5) and a different particle approximation of $\mathbb{Q}_{\theta,n}$ that avoids the path degeneracy problem.

The measure $\mathbb{Q}_{\theta,n}(\mathrm{d}x_{0:n})$ admits the backward representation

$$\mathbb{Q}_{\theta,n}(\mathrm{d}x_{0:n}) = \eta_{\theta,n}(\mathrm{d}x_n)\prod_{k=n}^{1} M_{\theta,k}(x_k, \mathrm{d}x_{k-1}),$$

and the corresponding particle approximation of $\mathbb{Q}_{\theta,n}$ is given by

$$\mathbb{Q}^N_{\theta,n}(\mathrm{d}x_{0:n}) = \eta^N_{\theta,n}(\mathrm{d}x_n)\prod_{k=n}^{1} M^N_{\theta,k}(x_k, \mathrm{d}x_{k-1}),$$

where $M^N_{\theta,k}(x_k, \mathrm{d}x_{k-1})$ was defined in (1.9). This now gives rise to the following particle approximation of $\zeta_{\theta,n}(\varphi_n)$ [22, 23]:

$$\zeta^N_{\theta,n}(\varphi_n) = \int \mathbb{Q}^N_{\theta,n}(\mathrm{d}x_{0:n})T_{\theta,n}(x_{0:n})\left(\varphi_n(x_n) - \eta^N_{\theta,n}(\varphi_n)\right)$$

as indeed $\eta^N_{\theta,n}(\varphi_n) = \int \mathbb{Q}^N_{\theta,n}(\mathrm{d}x_{0:n})\varphi_n(x_n)$. It is apparent that $\mathbb{Q}^N_{\theta,n}(\mathrm{d}x_{0:n})$ constructed using this backward method avoids the degeneracy in paths. It is even possible to compute $\zeta^N_{\theta,n}(\mathrm{d}x_n)$ recursively as detailed in Algorithm 1; since a recursion for $\eta_{\theta,n}(\mathrm{d}x_n)$ is already available, it is apparent from (2.5) that what remains is to specify a recursion for $\overline{T}_{\theta,n}(x_n) := \mathbb{E}_\theta\left[T_{\theta,n}\left(X_{0:n}\right)\middle| y_{0:n-1}, x_n\right]$. For $n \geq 1$, we have

$$\begin{aligned}\overline{T}_{\theta,n}(x_n) &= \mathbb{E}_\theta\left[T_{\theta,n-1}\left(X_{0:n-1}\right)\middle| y_{0:n-1}, x_n\right] + \mathbb{E}_\theta\left[t_{\theta,n}\left(X_{n-1}, X_n\right)\middle| y_{0:n-1}, x_n\right]\\ &= \int M_{\theta,n}(x_n, \mathrm{d}x_{n-1})\left(\mathbb{E}_\theta\left[T_{\theta,n-1}\left(X_{0:n-1}\right)\middle| y_{0:n-2}, x_{n-1}\right] + t_{\theta,n}\left(x_{n-1}, x_n\right)\right)\\ &= \int M_{\theta,n}(x_n, \mathrm{d}x_{n-1})\left(\overline{T}_{\theta,n-1}(x_{n-1}) + t_{\theta,n}\left(x_{n-1}, x_n\right)\right),\end{aligned}$$

where $\overline{T}_{\theta,0}(x_0) = t_{\theta,0}(x_0)$. Algorithm 1 computes $\zeta^N_{\theta,n}(\mathrm{d}x_n)$ recursively in time by approximating $\left(\overline{T}_{\theta,n}(x_n), \eta_{\theta,n}(\mathrm{d}x_n)\right)$ and is initialized with $\overline{T}^{(i)}_{\theta,0} = t_{\theta,0}(X^{(i)}_0)$ (see (2.2)) where $\{X^{(i)}_0\}_{1\leq i\leq N}$ are samples from $\pi_\theta(x_0)$.

ALGORITHM 1. A Particle Method to Compute the Filter Derivative

- Assume at time $n-1$ that approximate samples $\{X^{(i)}_{n-1}\}_{1\leq i\leq N}$ from $\eta_{\theta,n-1}$ and approximations $\{\overline{T}^{(i)}_{\theta,n-1}\}_{1\leq i\leq N}$ of $\{\overline{T}_{\theta,n-1}(X^{(i)}_{n-1})\}_{1\leq i\leq N}$ are available.
- At time n, sample $\{X^{(i)}_n\}_{1\leq i\leq N}$ independently from the mixture

$$\frac{\sum_{j=1}^N f_\theta(x_n|X^{(j)}_{n-1})g_\theta(\,y_{n-1}|\,X^{(j)}_{n-1})}{\sum_{j=1}^N g_\theta(\,y_{n-1}|\,X^{(j)}_{n-1})} \tag{2.7}$$

and then compute $\{\overline{T}^{(i)}_{\theta,n}\}_{1\leq i\leq N}$ and $\zeta^N_{\theta,n}$ as follows:

$$\overline{T}^{(i)}_{\theta,n} = \frac{\sum_{j=1}^N \left(\overline{T}^{(j)}_{\theta,n-1} + t_{\theta,n}(X^{(j)}_{n-1}, X^{(i)}_n)\right) f_\theta(X^{(i)}_n | X^{(j)}_{n-1}) g_\theta(\, y_{n-1} |\, X^{(j)}_{n-1})}{\sum_{j=1}^N f_\theta\left(X^{(i)}_n | X^{(j)}_{n-1}\right) g_\theta(\, y_{n-1} |\, X^{(j)}_{n-1})}, \tag{2.8}$$

$$\zeta^N_{\theta,n}(\mathrm{d}x_n) = \frac{1}{N}\sum_{i=1}^N \left(\overline{T}^{(i)}_{\theta,n} - \frac{1}{N}\sum_{j=1}^N \overline{T}^{(j)}_{\theta,n}\right) \delta_{X^{(i)}_n}(\mathrm{d}x_n). \tag{2.9}$$

Algorithm 1 uses the bootstrap particle filter of [13]. Note that any SMC implementation of $\{\eta_{\theta,n}\}_{n\geq 0}$ may be used, e.g., the auxiliary SMC method of [21] or sequential importance resampling with a tailored proposal distribution [12]. It was conjectured in [23] that the asymptotic variance of $\zeta^N_{\theta,n}(\varphi)$ for bounded integrands φ is uniformly bounded w.r.t. n under mixing assumptions. This is established in this paper.

3. Stability of the particle estimates. The convergence analysis of $\zeta^N_{\theta,n}$ (and for performance comparison $\zeta^{\mathrm{p},N}_{\theta,n}$) will largely focus on the convergence analysis of the N-particle measures $\mathbb{Q}^N_{\theta,n}$ (and correspondingly $\mathbb{Q}^{\mathrm{p},N}_{\theta,n}$) towards their limiting values $\mathbb{Q}_{\theta,n}$, as $N \to \infty$, which is in turn intimately related to the convergence of the flow of particle measures $\{\eta^N_{\theta,n}\}_{n\geq 0}$ towards their limiting measures $\{\eta_{\theta,n}\}_{n\geq 0}$. The $\mathbb{L}_r$ error bounds and the CLT presented here have been derived using the techniques developed in [3] for the convergence analysis of the particle occupation measures $\eta^N_{\theta,n}$. One of the central objects in this analysis is the local sampling errors defined as

$$V^N_{\theta,n} = \sqrt{N}\left(\eta^N_{\theta,n} - \Phi_{\theta,n}(\eta^N_{\theta,n-1})\right). \tag{3.1}$$

The fluctuation and deviations of these centered random measures can be estimated using nonasymptotic Kintchine's type $\mathbb{L}_r$-inequalities, as well as Hoeffding's or Bernstein's type exponential deviations [3, 7]. In [5] it is proved that these random perturbations behave asymptotically as Gaussian random perturbations; see Lemma A.10 in the appendix for more details. In the proof of Theorem A.11 (a supporting theorem) in the appendix we provide some key decompositions expressing the deviation of the particle measures $\mathbb{Q}^N_{\theta,n}$ around its limiting value $\mathbb{Q}_{\theta,n}$ in terms of the local sampling errors $(V^N_{\theta,0},\ldots,V^N_{\theta,n})$. These decompositions are key to deriving the $\mathbb{L}_r$-mean error bounds and central limit theorems for the filter derivative.

The following regularity conditions are assumed.

Assumption (A). The dominating measures $\mathrm{d}x$ on $\mathcal{X}$ and $\mathrm{d}y$ on $\mathcal{Y}$ are finite, and there exist constants $0 < \rho, \delta, c < \infty$ such that for all $(x, x', y, \theta) \in \mathcal{X}^2 \times \mathcal{Y} \times \Theta$, the derivatives of $\pi_\theta(x)$, $f_\theta(x'|x)$, and $g_\theta(y|x)$ w.r.t. θ exist and

$$\rho^{-1} \leq f_\theta(x'|x) \leq \rho, \quad \delta^{-1} \leq g_\theta(y|x) \leq \delta, \tag{3.2}$$

$$|\nabla \log \pi_\theta(x)| \vee |\nabla \log f_\theta(x'|x)| \vee |\nabla \log g_\theta(y|x)| \leq c. \tag{3.3}$$

Admittedly, these conditions are restrictive and fail to hold for many models in practice. Exceptions would include applications with a compact state-space. However, they are typically made to establish the time uniform stability of particle approximations of the filter [3, 10] as they lead to simpler and more transparent proofs. Also, we observe that the behaviors predicted by the following theorems seem to hold in practice even in cases where the state-space models do not satisfy these assumptions; see section 4. Thus the results in this paper can be seen to provide a qualitative guide to the behavior of the particle approximation even in the more general setting.

For each parameter vector $\theta \in \Theta$, realization of observations $y = \{y_n\}_{n\geq 0}$ and particle number N, let $(\Omega, \mathcal{F}, \mathbb{P}_\theta^y)$ be the underlying probability space of the random process $\{(X_n^{(1)}, \dots, X_n^{(N)})\}_{n\geq 0}$ comprised of the particle system only. Let $\mathbb{E}_\theta^y$ be the corresponding expectation operator computed with respect to $\mathbb{P}_\theta^y$. The first of the two main results in this section is a time uniform nonasymptotic error bound.

THEOREM 3.1. *Assume Assumption (A). For any $r \geq 1$, there exists a constant C_r such that for all $\theta \in \Theta$, $y = \{y_n\}_{n\geq 0}$, $n \geq 0$, $N \geq 1$, and $\varphi_n \in Osc_1(\mathcal{X})$,*

$$\sqrt{N}\mathbb{E}_\theta^y \left\{ \left| \zeta_{\theta,n}^N(\varphi_n) - \zeta_{\theta,n}(\varphi_n) \right|^r \right\}^{\frac{1}{r}} \leq C_r.$$

Let $\{V_{\theta,n}\}_{n\geq 0}$ be a sequence of independent centered Gaussian random fields defined as follows. For any sequence $\{\varphi_n\}_{n\geq 0}$ in $\mathcal{B}(\mathcal{X})$ and any $p \geq 0$, $\{V_{\theta,n}(\varphi_n)\}_{n=0}^p$ is a collection of independent zero-mean Gaussian random variables with variances given by

$$\eta_{\theta,n}(\varphi_n^2) - \eta_{\theta,n}(\varphi_n)^2. \tag{3.4}$$

THEOREM 3.2. *Assume (A) . There exists a constant $C < \infty$ such that for any $\theta \in \Theta$, $y = \{y_n\}_{n\geq 0}$, $n \geq 0$ and $\varphi_n \in Osc_1(\mathcal{X})$, $\sqrt{N}(\zeta_{\theta,n}^N - \zeta_{\theta,n})(\varphi_n)$ converges in law, as $N \to \infty$, to the centered Gaussian random variable*

$$\sum_{p=0}^{n} V_{\theta,p}\left(G_{\theta,p,n}\ \frac{D_{\theta,p,n}(F_{\theta,n} - \mathbb{Q}_{\theta,n}(F_{\theta,n}))}{D_{\theta,p,n}(1)} \right) \tag{3.5}$$

of variance bounded above by C, where

$$F_{\theta,n} = (\varphi_n - \mathbb{Q}_{\theta,n}(\varphi_n))\,(T_{\theta,n} - \mathbb{Q}_{\theta,n}(T_{\theta,n}))\,.$$

The proofs of both these results are in the appendix.

As a comparison, we quantify the variance of the particle estimate of the filter derivative computed using the path-based method (see (2.6).) Consider the following simplified example that serves to illustrate the point. Let $g_\theta\,(y|\,x) = g\,(y|\,x)$ (that is θ-independent), $f_\theta\,(x_n|\,x_{n-1}) = \pi_\theta(x_n)$, where π_θ is the initial distribution. (Note that f_θ in this case satisfies a rephrased version of (3.2) under which the conclusion of Theorem 3.2 also holds.) Also, consider the sequence of repeated observations $y_0 = y_1 = y_2 \cdots$ where y_0 is arbitrary. Applying Lemma A.12 (in the appendix) that characterizes the limiting distribution of $\sqrt{N}(\mathbb{Q}_{\theta,n}^{\mathrm{p},N} - \mathbb{Q}_{\theta,n})$ to this special case results in $\sqrt{N}(\zeta_{\theta,n}^{\mathrm{p},N} - \zeta_{\theta,n})(\varphi)$ (see (2.6)) having an asymptotic distribution which is Gaussian with mean zero and variance

$$n\pi_\theta(\overline{\varphi}^2)\pi_\theta'\left[(\nabla \log \pi_\theta)^2\right] + \pi_\theta\left[\overline{\varphi}^2(\nabla \log \pi_\theta)^2\right] - \nabla\pi_\theta(\varphi)^2,$$

where $\overline{\varphi} = \varphi - \pi_\theta(\varphi)$, $\pi_\theta'(x) = \pi_\theta(x)g\,(y_0|\,x)\,/\pi_\theta\,(g\,(y_0|\,\cdot))$. This variance increases linearly with time in contrast to the time bounded variance of Theorem 3.2.

4. Application to recursive parameter estimation. Being able to compute $\{\zeta_{\theta,n}(\mathrm{d}x_n)\}_{n\geq 0}$ is particularly useful when performing online static parameter estimation for state-space models using recursive maximum likelihood (RML) techniques [17, 22, 23]; see also [16] for a general review of available particle methods based solutions, including Bayesian ones, for this problem. The filter derivative may also be useful in other areas; e.g., see [2] for an application in control.

4.1. Recursive maximum likelihood. Let θ^* be the true static parameter generating the observed data $\{y_n\}_{n\geq 0}$. Given a finite record of observations $y_{0:T}$, the log-likelihood may be maximized with the following steepest ascent algorithm:

$$\theta_k = \theta_{k-1} + \gamma_k \nabla \log p_\theta(y_{0:T})|_{\theta=\theta_{k-1}}, \quad k \geq 1, \tag{4.1}$$

where θ_0 is some arbitrary initial guess of θ^*, $\nabla \log p_\theta(y_{0:T})|_{\theta=\theta_{k-1}}$ denotes the gradient of the log-likelihood evaluated at the current parameter estimate, and $\{\gamma_k\}_{k\geq 1}$ is a decreasing positive real-valued step-size sequence, which should satisfy the following conditions:

$$\sum_{k=1}^{\infty} \gamma_k = \infty, \qquad \sum_{k=1}^{\infty} \gamma_k^2 < \infty.$$

Although $\nabla \log p_\theta(y_{0:T})$ can be computed using (4.3), the computational cost can be prohibitive for a long data record since each iteration of (4.1) would require a complete browse through the $T+1$ data points. A more attractive alternative would be a recursive procedure in which the data is run through once only sequentially. For example, consider the following update scheme:

$$\theta_n = \theta_{n-1} + \gamma_n \nabla \log p_\theta(y_n|y_{0:n-1})|_{\theta=\theta_{n-1}}, \tag{4.2}$$

where $\nabla \log p_\theta(y_n|y_{0:n-1})|_{\theta=\theta_{n-1}}$ denotes the gradient of $\log p_\theta(y_n|y_{0:n-1})$ evaluated at the current parameter estimate; that is, upon receiving y_n, θ_{n-1} is updated in the direction of ascent of the conditional density of this new observation. Since we have $\nabla \log p_\theta(y_n|y_{0:n-1})|_{\theta=\theta_{n-1}}$ equal to

$$\frac{\int \eta_{\theta_{n-1},n}(\mathrm{d}x_n) \nabla g_\theta(y_n|x_n)|_{\theta_{n-1}} + \int \zeta_{\theta_{n-1},n}(\mathrm{d}x_n) g_{\theta_{n-1}}(y_n|x_n)}{\int \eta_{\theta_{n-1},n}(\mathrm{d}x_n) g_{\theta_{n-1}}(y_n|x_n)}, \tag{4.3}$$

this clearly requires the filter derivative $\zeta_{\theta,n}(\mathrm{d}x_n)$. The algorithm in the present form is not suitable for online implementation as it requires recomputing the filter and its derivative at the value $\theta = \theta_{n-1}$ from time zero. The RML procedure uses an approximation of (4.3) which is obtained by updating the filter and its derivative using the parameter value θ_{n-1} at time n; we refer the reader to [17] for details. The asymptotic properties of the RML algorithm, i.e., the behavior of θ_n in the limit as n goes to infinity, have been studied in the case of an independently and identically distributed (i.i.d.) hidden process by [24] and [17] for a finite state-space hidden Markov model. It is shown in [17] that under regularity conditions this algorithm converges towards a local maximum of the average log-likelihood and that this average log-likelihood is maximized at θ^*. A particle version of the RML algorithm of [17] that uses Algorithm 1's estimate of $\eta_{\theta,n}(\mathrm{d}x_n)$ is presented as Algorithm 2.

ALGORITHM 2. Particle Recursive Maximum Likelihood

- At time $n-1$ we are given $y_{0:n-1}$, the estimate θ_{n-1} of θ^* and $\{(X_{n-1}^{(i)}, \overline{T}_{n-1}^{(i)})\}_{i=1}^N$.
- At time n, upon receiving y_n, sample $\{X_n^{(i)}\}_{1\leq i\leq N}$ independently from (2.7) using parameter $\theta = \theta_{n-1}$ to obtain

$$\eta_n^N(\mathrm{d}x_n) = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_n^{(i)}}(\mathrm{d}x_n)$$

and then compute

(4.4)

$$\overline{T}_n^{(i)} = \frac{\sum_{j=1}^N (\overline{T}_{n-1}^{(j)} + t_{\theta_{n-1},n}(X_{n-1}^{(j)}, X_n^{(i)})) f_{\theta_{n-1}}(X_n^{(i)}|X_{n-1}^{(j)}) g_{\theta_{n-1}}(\, y_{n-1}|\, X_{n-1}^{(j)})}{\sum_{j=1}^N f_{\theta_{n-1}}(X_n^{(i)}|X_{n-1}^{(j)}) g_{\theta_{n-1}}(\, y_{n-1}|\, X_{n-1}^{(j)})},$$

(4.5)

$$\zeta_n^N(\mathrm{d}x_n) = \frac{1}{N}\sum_{i=1}^N \left(\overline{T}_n^{(i)} - \frac{1}{N}\sum_{j=1}^N \overline{T}_n^{(j)} \right) \delta_{X_n^{(i)}}(\mathrm{d}x_n),$$

and

$$\widehat{\nabla} \log p\,(y_n|\, y_{0:n-1}) = \frac{\int \eta_n^N(\mathrm{d}x_n)\, \nabla g_\theta\,(y_n|\, x_n)|_{\theta_{n-1}} + \int \zeta_n^N(\mathrm{d}x_n) g_{\theta_{n-1}}\,(y_n|\, x_n)}{\int \eta_n^N(\mathrm{d}x_n) g_{\theta_{n-1}}\,(y_n|\, x_n)}.$$

Finally, we update the parameter:

$$\theta_n = \theta_{n-1} + \gamma_n \widehat{\nabla} \log p\,(y_n|\, y_{0:n-1})\,.$$

Omitting step (4.1), i.e., $\theta_n = \theta_0$ for all n, then the particle approximation of the filter is stable under Assumption (A) [3]; see also Lemma A.4 in the appendix. This combined with the proven stability of the particle approximation of the filter derivative implies that the particle estimate of the derivative of $\log p_\theta\,(y_n|\, y_{0:n-1})$ at $\theta = \theta_0$ is also stable.

4.2. Simulations. The RML algorithm is applied to the following stochastic volatility model [21]:

$$X_0 \sim \mathcal{N}\left(0, \frac{\sigma^2}{1-\phi^2}\right), \quad X_{n+1} = \phi X_n + \sigma V_{n+1},$$
$$Y_n = \beta \exp\left(X_n/2\right) W_n,$$

where $\mathcal{N}(m, s)$ denotes a Gaussian random variable with mean m and variance s, and $V_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ and $W_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ are two mutually independent sequences, both independent of the initial state X_0. The model parameters, $\theta = (\phi, \sigma, \beta)$, are to be estimated.

Our first example demonstrates the theoretical results in section 3. The estimate of $\partial/\partial\sigma \log p\,(y_{n:n+L-1}|\, y_{0:n-1})$ at $\theta^* = (0.8, \sqrt{0.1}, 1)$ was computed using Algorithm 1 with 500 particles and using the path-space method (see (2.6)) with 2.5×10^5 particles for the stochastic volatility model. The block size L was 500. Shown in Figure 1 is the variance of these particle estimates for various values of n derived from many independent random replications of the simulation. The linear increase of the variance of the path-space method as predicted by theory is evident although Assumption A is not satisfied.

For the path-space method, because the (asymptotic) variance of the estimate of the filter derivative grows linearly in time, the eventual high variance in the gradient estimate can result in the divergence of the parameter estimates. To illustrate this point, (4.1) was implemented with the path-space estimate of the filter derivative (2.6) computed with 10,000 particles and constant step-size sequence $\gamma_n = 10^{-4}$ for all n.

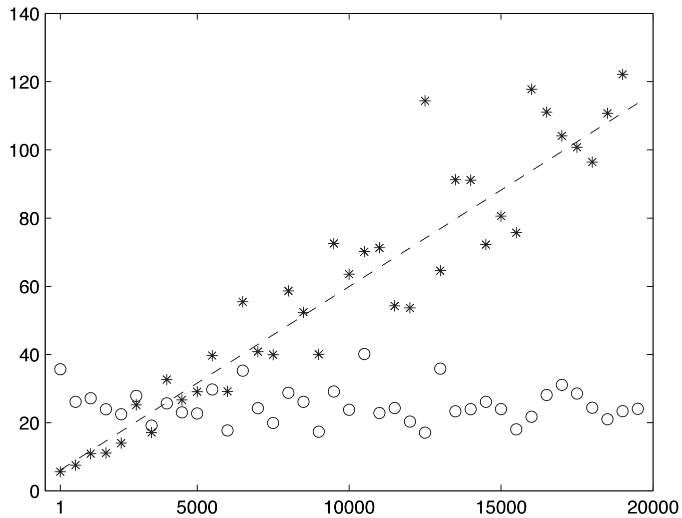

FIG. 1. *Variance of the particle estimates of* $\partial/\partial\sigma \log p\,(y_{n:n+500-1}|\, y_{0:n-1})$ *for various values of* n *for the stochastic volatility model. Circles are the variance of Algorithm* 1*'s estimate with* 500 *particles. Stars indicate the variance of the estimate of the path-space method with* 2.5×10^5 *particles. The dotted line is the best fitting straight line to path-space method's variance to indicate trend.*

θ_0 was initialized at the true parameter value. A sequence of two million observations was simulated with $\theta^* = (0.8, \sqrt{0.1}, 1)$. The results are shown in Figure 2.

For the same value of θ^* and sequence of observations used in the previous example, Algorithm 2 was executed with 500 particles and $\gamma_n = 0.01$, $n \le 10^5$, $\gamma_n = (n - 5\times 10^4)^{-0.6}$, $n > 10^5$. As can be seen from the results in Figure 3, the estimate converges to a value in the neighborhood of the true parameter.

5. Conclusion. We have presented theoretical results establishing the uniform stability of the particle approximation of the optimal filter derivative proposed in [22, 23]. While these results have been presented in the context of state-space models, they can also be applied to Feynman–Kac models [3], which could potentially enlarge the range of applications. For example, if $f_\theta\,(x'|\,x)\,\mathrm{d}x'$ is reversible w.r.t. to some probability measure $\mu_\theta(\mathrm{d}x')$, and if we replace $g_\theta\,(y_n|\,x_n)$ with a time-homogeneous potential function $g_\theta\,(x_n)$, then $\eta_{\theta,n}(\mathrm{d}x)$ converges, as $n\to\infty$, to the probability measure $\mu_{\theta,h}(\mathrm{d}x)$ defined as

$$\mu_{\theta,h}(\mathrm{d}x) := \frac{1}{\mu_\theta(h_\theta\ \int f_\theta\,(x'|\,\cdot)\,h_\theta(x')\mathrm{d}x')}\mu_\theta(\mathrm{d}x)\ h_\theta(x)\ \int f_\theta\,(x'|\,x)\,h_\theta(x')\mathrm{d}x',$$

where h_θ is a positive eigenmeasure associated with the top eigenvalue of the integral operator $Q_\theta(x,\mathrm{d}x') = g_\theta(x) f_\theta\,(x'|\,x)\,\mathrm{d}x'$ (see section 12.4 of [3]). The measure $\mu_{\theta,h}(\mathrm{d}x)$ is the invariant measure of the h-process defined as the Markov chain with transition kernel $M_\theta\,(x,\mathrm{d}x') \propto f_\theta\,(x'|\,x)\,h_\theta(x')\mathrm{d}x'$. The particle algorithm described here can be directly used to approximate the derivative of this invariant measure w.r.t. to θ. It would also be of interest to weaken Assumption (A) and there are several ways this might be approached. For example, one could use ideas in [19, 14] or via

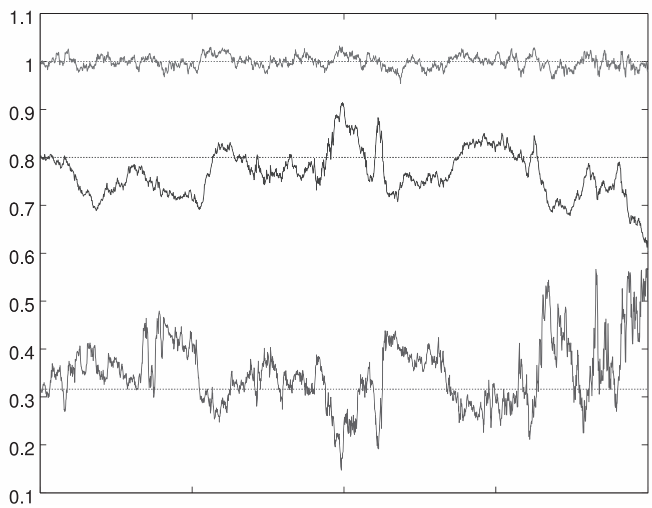

FIG. 2. *RML for stochastic volatility with path-space gradient estimate with* $10,000$ *particles, constant step-size, and initialized at the true parameter values which are indicated by the dashed lines. From top to bottom:* β*,* ϕ*, and* σ*.*

Lyapunov conditions as in [26] and [15].

Appendix. The statements of the results in this section hold for any θ and any sequence of observations $y = \{y_n\}_{n\geq 0}$. All mathematical expectations are taken with respect to the law of the particle system only for the specific θ and y under consideration. While θ is retained in the statement of the results, it is omitted in the proofs. The superscript y of the expectation operator is also omitted in the proofs.

This section commences with some essential definitions in addition to those in section 1.1. Let

$$P_{\theta,k,n}(x_k, \mathrm{d}x_n) = \frac{Q_{\theta,k,n}(x_k, \mathrm{d}x_n)}{Q_{\theta,k,n}(1)(x_k)},$$

and

$$\mathcal{M}_{\theta,p}(x_p, \mathrm{d}x_{0:p-1}) = \prod_{k=p}^{1} M_{\theta,k}(x_k, \mathrm{d}x_{k-1}), \quad p > 0,$$

and its corresponding particle approximation is

$$\mathcal{M}^N_{\theta,p}(x_p, \mathrm{d}x_{0:p-1}) = \prod_{k=p}^{1} M^N_{\theta,k}(x_k, \mathrm{d}x_{k-1}).$$

To make the subsequent expressions more terse, let

$$\widetilde{\eta}^N_{\theta,n} = \Phi_{\theta,n}(\eta^N_{\theta,n-1}), \quad n \geq 0, \tag{A.1}$$

where $\widetilde{\eta}^N_{\theta,0} = \Phi_{\theta,0}(\eta^N_{-1}) = \eta_{\theta,0} = \pi_\theta$ by convention. (Recall $\Phi_{\theta,n} = \Phi_{\theta,n-1,n}$.) Let

$$\mathcal{F}^N_n = \sigma\left(\left\{X^{(i)}_k; 0 \leq k \leq n, 1 \leq i \leq N\right\}\right), \quad n \geq 0,$$

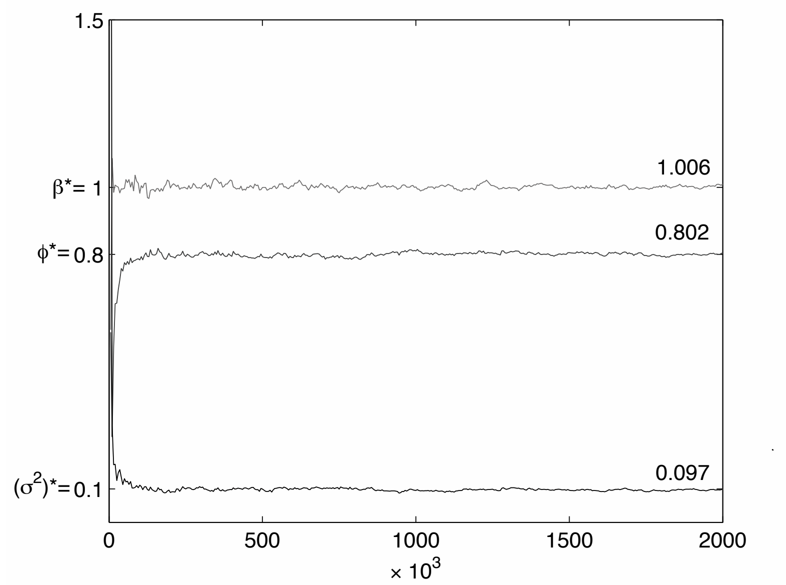

FIG. 3. *Sequence of recursive parameter estimates,* $\theta_n = (\sigma_n, \phi_n, \beta_n)$*, computed using* (4.1) *with* $N = 500$*. From top to bottom:* β_n*,* ϕ_n*, and* σ_n *and marked on the right are the "converged values" which were taken to be the empirical average of the last* 1,000 *values.*

be the natural filtration associated with the N-particle approximation model, and let $\mathcal{F}_{-1}^N$ be the trivial sigma field.

The following estimates are a straightforward consequence of Assumption (A). For all θ and time indices $0 \le k < q \le n$,

(A.2)
$$b_{\theta,k,n} = \sup_{x_k, x_k'} \frac{Q_{\theta,k,n}(1)(x_k)}{Q_{\theta,k,n}(1)(x_k')} \le \rho^2\delta^2, \quad \beta\left(\frac{Q_{\theta,k,q}(x_k, \mathrm{d}x_q)Q_{\theta,q,n}(1)(x_q)}{Q_{\theta,k,q}(Q_{\theta,q,n}(1))(x_k)}\right) \le \overline{\rho}^{q-k},$$

where $\overline{\rho} = \left(1 - \rho^{-4}\right)$ and for θ, $0 < k \le q$,

$$M_{\theta,k}^N(x, \mathrm{d}z) \le \rho^4 \; M_{\theta,k}^N(x', \mathrm{d}z) \Longrightarrow \beta\left(M_{\theta,q}^N \cdots M_{\theta,k}^N\right) \le \left(1 - \rho^{-4}\right)^{q-k+1}. \tag{A.3}$$

Note that setting $q = n$ in (A.2) yields an estimate for $\beta(P_{\theta,k,n})$.

Several auxiliary results are now presented, all of which hinge on the following Kintchine type moment bound proved in [3, Lemma 7.3.3].

LEMMA A.1 (see [3, Lemma 7.3.3]). *Let* μ *be a probability measure on the measurable space* $(E, \mathcal{E})$*. Let* G *and* h *be* $\mathcal{E}$*-measurable functions satisfying* $G(x) \ge cG(x') > 0$ *for all* $x, x' \in E$ *where* c *is some finite positive constant. Let* $\{X^{(i)}\}_{1 \le i \le N}$ *be a collection of independent random samples from* μ*. If* h *has finite oscillation, then for any integer* $r \ge 1$ *there exists a finite constant* a_r*, independent of* N*,* G*, and* h*, such that*

$$\sqrt{N}\mathbb{E}\left\{\left|\frac{\sum_{i=1}^N G(X^{(i)})h(X^{(i)})}{\sum_{i=1}^N G(X^{(i)})} - \frac{\mu(Gh)}{\mu(G)}\right|^r\right\}^{\frac{1}{r}} \le c^{-1}\mathrm{osc}(h)a_r.$$

Proof. The result for $G = 1$ and $c = 1$ is proved in [3]. The case stated here can be established using the representation

$$\frac{\mu^N(Gh)}{\mu^N(G)} - \frac{\mu(Gh)}{\mu(G)} = \frac{\mu(G)}{\mu^N(G)} \left(\mu^N - \mu\right) \left[\frac{G}{\mu(G)} \left(h - \frac{\mu(Gh)}{\mu(G)}\right)\right],$$

where $\mu^N(\mathrm{d}x) = N^{-1}\sum_{i=1}^N \delta_{X^{(i)}}(\mathrm{d}x)$. ∎

Remark A.2. For $k \geq 0$, let h_{k-1}^N be a $\mathcal{F}_{k-1}^N$ measurable function satisfying $h_{k-1}^N \in \mathrm{Osc}_1(\mathcal{X})$ almost surely. Then Lemma A.1 can be invoked to establish

$$\sqrt{N}\mathbb{E}_\theta^y \left\{\left|\frac{\eta_{\theta,k}^N(Gh_{k-1}^N)}{\eta_{\theta,k}^N(G)} - \frac{\Phi_{\theta,k}(\eta_{\theta,k-1}^N)(Gh_{k-1}^N)}{\Phi_{\theta,k}(\eta_{\theta,k-1}^N)(G)}\right|^r\right\}^{\frac{1}{r}} \leq c^{-1}a_r,$$

where G is defined as in Lemma A.1.

Lemmas A.3 to A.6 are consequences of Lemma A.1 and the estimates in (A.2).

LEMMA A.3. *For any $r \geq 1$ there exist a finite constant a_r such that the following inequality holds for all θ, y, $0 \leq k \leq n$, and $\mathcal{F}_{k-1}^N$ measurable function φ_n^N satisfying $\varphi_n^N \in Osc_1(\mathcal{X})$ almost surely;*

$$\sqrt{N}\mathbb{E}_\theta^y \left(\left|\Phi_{\theta,k,n}(\eta_{\theta,k}^N)(\varphi_n^N) - \Phi_{\theta,k-1,n}(\eta_{\theta,k-1}^N)(\varphi_n^N)\right|^r \quad\right)^{\frac{1}{r}} \leq a_r \; b_{\theta,k,n} \; \beta\left(P_{\theta,k,n}\right),$$

where, by convention $\Phi_{\theta,-1,n}(\eta_{\theta,-1}^N) = \eta_{\theta,n}$, and the constants $b_{\theta,k,n}$ and $\beta\left(P_{\theta,k,n}\right)$ were defined in (A.2).

Proof. The following equality holds:

$$\begin{aligned}&\Phi_{k,n}(\eta_k^N)(\varphi_n^N) - \Phi_{k-1,n}(\eta_{k-1}^N)(\varphi_n^N)\\ &= \int \left(\frac{\eta_k^N(\mathrm{d}x_k)Q_{k,n}(1)(x_k)}{\eta_k^N Q_{k,n}(1)} - \frac{\Phi_k(\eta_{k-1}^N)(\mathrm{d}x_k)Q_{k,n}(1)(x_k)}{\Phi_k(\eta_{k-1}^N)Q_{k,n}(1)}\right) P_{k,n}(\varphi_n^N)(x_k),\end{aligned}$$

where $\Phi_0(\eta_{-1}^N) = \eta_0$ by convention. Applying Lemma A.1 with the estimates in (A.2) we have

$$\sqrt{N}\mathbb{E}\left(\left|\Phi_{k,n}(\eta_k^N)(\varphi_n^N) - \Phi_{k-1,n}(\eta_{k-1}^N)(\varphi_n^N)\right|^r \;\middle|\; \mathcal{F}_{k-1}^N\right)^{\frac{1}{r}} \leq a_r \; b_{k,n} \; \beta\left(P_{k,n}\right)$$

almost surely. ∎

Lemma A.3 may be used to derive the following error estimate [3, Theorem 7.4.4].

LEMMA A.4. *For any $r \geq 1$, there exists a constant c_r such that the following inequality holds for all θ, y, $n \geq 0$, and $\varphi \in Osc_1(\mathcal{X})$;*

$$\sqrt{N}\mathbb{E}_\theta^y \left(\left|[\eta_{\theta,n}^N - \eta_{\theta,n}](\varphi)\right|^r\right)^{\frac{1}{r}} \leq c_r \sum_{k=0}^n b_{\theta,k,n} \; \beta\left(P_{\theta,k,n}\right). \tag{A.4}$$

Assume Assumption (A). For any $r \geq 1$, there exists a constant c_r' such that for all θ, y, $n \geq 0$, $\varphi \in Osc_1(\mathcal{X})$, $G \in \mathcal{B}(\mathcal{X})$ such that G is positive and satisfies $G(x) \geq c_G G(x')$ for all $x, x' \in \mathcal{X}$ for some positive constant c_G,

$$\sqrt{N}\mathbb{E}_\theta^y \left(\left|\left[\frac{\eta_{\theta,n}^N(\mathrm{d}x_n)G(x_n)}{\eta_{\theta,n}^N(G)} - \frac{\eta_{\theta,n}(\mathrm{d}x_n)G(x_n)}{\eta_{\theta,n}(G)}\right](\varphi)\right|^r\right)^{\frac{1}{r}} \leq c_r'(1 + c_G^{-1}). \tag{A.5}$$

Proof. The first part follows from applying Lemma A.3 to the telescopic sum [3, Theorem 7.4.4]:

$$\left(\eta_n^N - \eta_n\right)(\varphi) = \sum_{k=0}^{n} \Phi_{k,n}(\eta_k^N)(\varphi) - \Phi_{k-1,n}(\eta_{k-1}^N)(\varphi)$$

with the convention that $\Phi_{-1,n}(\eta_{-1}^N) = \eta_n$. For the second part, use the same telescopic sum but with the kth term being

$$\begin{aligned}&\frac{\Phi_{k,n}(\eta_k^N)(\varphi G)}{\Phi_{k,n}(\eta_k^N)(G)} - \frac{\Phi_{k-1,n}(\eta_{k-1}^N)(\varphi G)}{\Phi_{k-1,n}(\eta_{k-1}^N)(G)}\\ &= \int \left(\frac{\eta_k^N(\mathrm{d}x_k)Q_{k,n}(G)(x_k)}{\eta_k^N Q_{k,n}(G)} - \frac{\Phi_k(\eta_{k-1}^N)(\mathrm{d}x_k)Q_{k,n}(G)(x_k)}{\Phi_k(\eta_{k-1}^N)Q_{k,n}(G)} \right) \frac{Q_{k,n}(G\varphi)(x_k)}{Q_{k,n}(G)(x_k)}.\end{aligned}$$

Apply Lemma A.1 using the same estimates in (A.2), i.e., the same estimates hold with G replacing 1 in the definition of $b_{k,n}$ and with G replacing $Q_{q,n}(1)$ in the argument of β. ∎

The following result is a consequence of Lemma A.4.

LEMMA A.5. *Assume Assumption (A). For any $r \geq 1$, there exists a constant c_r such that the following inequality holds for all θ, y, $0 \leq k \leq n$, $N \geq 1$, and $\varphi_n \in Osc_1(\mathcal{X})$:*

$$\sqrt{N}\mathbb{E}_\theta^y \left(\left| \left[\Phi_{\theta,k,n}(\eta_{\theta,k}^N) - \Phi_{\theta,k,n}(\eta_{\theta,k}) \right](\varphi_n) \right|^r \right)^{\frac{1}{r}} \leq c_r \overline{\rho}^{n-k}.$$

Proof. The result is established by expressing $\Phi_{k,n}(\eta_k^N)$ as

$$\Phi_{k,n}(\eta_k^N)(\mathrm{d}x_n) = \int \frac{\eta_k^N(\mathrm{d}x_k)Q_{k,n}(1)(x_k)}{\eta_k^N Q_{k,n}(1)} P_{k,n}(x_k, \mathrm{d}x_n),$$

expressing $\Phi_{k,n}(\eta_k)$ similarly, setting G in (A.5) to $Q_{k,n}(1)$, $\varphi = P_{k,n}(\varphi_n)$, and using the estimates in (A.2). ∎

LEMMA A.6. *For each $r \geq 1$, there exists a finite constant c_r such that for all θ, y, $0 \leq k \leq q \leq n$, and $\mathcal{F}_{k-1}^N$ measurable functions φ_q^N satisfying $\varphi_q \in Osc_1(\mathcal{X})$ almost surely,*

$$\begin{aligned}\mathbb{E}_\theta^y &\left(\left| \int \left(\frac{\Phi_{\theta,k,q}(\eta_{\theta,k}^N)(\mathrm{d}x_q)Q_{\theta,q,n}(1)(x_q)}{\Phi_{\theta,k,q}(\eta_{\theta,k}^N)Q_{\theta,q,n}(1)} \right.\right.\right.\\ &\qquad \left.\left.\left. - \frac{\Phi_{\theta,k-1,q}(\eta_{\theta,k-1}^N)(\mathrm{d}x_q)Q_{\theta,q,n}(1)(x_q)}{\Phi_{\theta,k-1,q}(\eta_{\theta,k-1}^N)Q_{\theta,q,n}(1)} \right) \varphi_q^N(x_q) \right|^r \right)^{\frac{1}{r}}\\ &\leq \frac{1}{\sqrt{N}} c_r\ b_{\theta,k,n}\ \beta\left(\frac{Q_{\theta,k,q}(x_k, \mathrm{d}x_q)Q_{\theta,q,n}(1)(x_q)}{Q_{\theta,k,q}(Q_{\theta,q,n}(1))(x_k)} \right).\end{aligned}$$

Proof. This result is established by noting that

$$\begin{aligned}&\frac{\Phi_{k,q}(\eta_k^N)(\mathrm{d}x_q)Q_{q,n}(1)(x_q)}{\Phi_{k,q}(\eta_k^N)Q_{q,n}(1)} - \frac{\Phi_{k-1,q}(\eta_{k-1}^N)(\mathrm{d}x_q)Q_{q,n}(1)(x_q)}{\Phi_{k-1,q}(\eta_{k-1}^N)Q_{q,n}(1)}\\ &\quad = \int \left(\frac{\eta_k^N(\mathrm{d}x_k)Q_{k,n}(1)(x_k)}{\eta_k^N Q_{k,n}(1)} \right.\\ &\qquad \left. - \frac{\Phi_k(\eta_{k-1}^N)(\mathrm{d}x_k)Q_{k,n}(1)(x_k)}{\Phi_k(\eta_{k-1}^N)Q_{k,n}(1)} \right) \frac{Q_{k,q}(x_k, \mathrm{d}x_q)Q_{q,n}(1)(x_q)}{Q_{k,n}(1)(x_k)}.\end{aligned}$$

Now Lemma A.1 is applied using the estimates in (A.2). ∎

Lemma A.7. *Assume Assumption (A). There exists a collection of a pair of finite positive constants, a_i, c_i, $i \geq 1$, such that the following bounds hold for all $r \geq 1$, θ, y, $0 \leq p \leq n$, $N \geq 1$, $x_p \in \mathcal{X}$, $F_p \in \mathcal{B}(\mathcal{X}^{p+1})$, $F_n \in \mathcal{B}(\mathcal{X}^{n+1})$:*

$$\sqrt{N}\mathbb{E}_\theta^y \left(\left|\mathcal{M}_{\theta,p}^N \left(F_p(.,x_p)\right)(x_p) - \mathcal{M}_{\theta,p}\left(F_p(.,x_p)\right)(x_p)\right|^r \right)^{\frac{1}{r}} \leq \|F_p\|\, a_r p,$$

$$\sqrt{N}\mathbb{E}_\theta^y \left(\left|D_{\theta,p,n}^N(F_n)(x_p) - D_{\theta,p,n}(F_n)(x_p)\right|^r \right)^{\frac{1}{r}} \leq a_r c_n \|F_n\|.$$

Proof. For each x_p, let $x_{0:p-1} \to G_{p-1,x_p}(x_{0:p-1}) = F_p(x_{0:p})q(x_p|x_{p-1})$. Adopting the convention $\widetilde{\eta}_0^N = \eta_0$,

$$\begin{aligned}
&\mathcal{M}_p^N\left(F_p(.,x_p)\right)(x_p) - \mathcal{M}_p\left(F_p(.,x_p)\right)(x_p)\\
&\quad= \sum_{k=1}^{p}\int \left(\frac{\eta_{p-k}^N D_{p-k,p-1}^N(\mathrm{d}x_{0:p-1})q(x_p|x_{p-1})}{\eta_{p-k}^N D_{p-k,p-1}^N(q(x_p|.))} \right.\\
&\quad\quad \left. - \frac{\widetilde{\eta}_{p-k}^N D_{p-k,p-1}^N(\mathrm{d}x_{0:p-1})q(x_p|x_{p-1})}{\widetilde{\eta}_{p-k}^N D_{p-k,p-1}^N(q(x_p|.))}\right) F_p(x_{0:p})\\
&\quad= \sum_{k=1}^{p}\int \left(\frac{\eta_{p-k}^N(\mathrm{d}x_{p-k})Q_{p-k,p-1}(q(x_p|.))(x_{p-k})}{\eta_{p-k}^N Q_{p-k,p-1}(q(x_p|.))} \right.\\
&\quad\quad \left. - \frac{\widetilde{\eta}_{p-k}^N(\mathrm{d}x_{p-k})Q_{p-k,p-1}(q(x_p|.))(x_{p-k})}{\widetilde{\eta}_{p-k}^N Q_{p-k,p-1}(q(x_p|.))}\right)\\
&\quad\quad\quad \times \frac{G_{p-k,p-1,x_p}^N(x_{p-k})}{Q_{p-k,p-1}(q(x_p|.))(x_{p-k})},
\end{aligned}$$

where $G_{p-k,p-1,x_p}^N(x_{p-k}) = D_{p-k,p-1}^N(G_{p-1,x_p})(x_{p-k})$, which is an $\mathcal{F}_{p-k-1}^N$-measurable function with norm

$$\sup_{x_{p-k}} \left| \frac{G_{p-k,p-1,x_p}^N(x_{p-k})}{Q_{p-k,p-1}(q(x_p|.))(x_{p-k})} \right| \leq \|F_p\|.$$

The result is established upon applying Lemma A.1 (see Remark A.2) to each term in the sum separately and using the estimates in (A.2).

To establish the second result, let

$$F_{p,n}(x_{0:p}) = \int Q_{p+1}(x_p, \mathrm{d}x_{p+1}) \cdots Q_n(x_{n-1}, \mathrm{d}x_n) F_n(x_{0:n}).$$

Then,

$$D_{p,n}^N(F_n)(x_p) - D_{p,n}(F_n)(x_p) = \mathcal{M}_p^N\left(F_{p,n}(.,x_p)\right)(x_p) - \mathcal{M}_p\left(F_{p,n}(.,x_p)\right)(x_p).$$

The result follows by setting $c_n = p \sup_\theta \|Q_{\theta,p,n}(1)\|$, and it follows from Assumption (A) that c_n is finite. ∎

Lemmas A.8 and A.9 both build on the previous results and are needed for the proof of Theorem 3.1.

Lemma A.8. *Assume Assumption (A). For any $r \geq 1$ there exists a constant C_r*

such that for all θ, y, $0 \le k < n$, $N \ge 1$, $\varphi_n \in Osc_1(\mathcal{X})$,

$$\sqrt{N}\mathbb{E}_\theta^y \left\{ \left| \int \mathbb{Q}_{\theta,n}^N(\mathrm{d}x_{0:n}) t_{\theta,k}(x_{k-1}, x_k) \left(\varphi_n(x_n) - \eta_{\theta,n}^N(\varphi_n)\right) \right.\right.$$
$$\left.\left. - \int \frac{\eta_{\theta,k}^N D_{\theta,k,n}^N(\mathrm{d}x_{0:n})}{\eta_{\theta,k}^N D_{\theta,k,n}^N(1)} t_{\theta,k}(x_{k-1}, x_k) \left(\varphi_n(x_n) - \frac{\eta_{\theta,k}^N D_{\theta,k,n}^N(\varphi_n)}{\eta_{\theta,k}^N D_{\theta,k,n}^N(1)}\right) \right|^r \right\}^{\frac{1}{r}}$$
$$\text{(A.6)} \qquad \le 2(n-k) C_r \overline{\rho}^{n-k}.$$

Proof. The term (A.6) can be further expanded as

$$\int \frac{\eta_k^N D_{k,n}^N(\mathrm{d}x_{0:n})}{\eta_k^N D_{k,n}^N(1)} t_k(x_{k-1}, x_k) \left(\varphi_n(x_n) - \frac{\eta_k^N D_{k,n}^N(\varphi_n)}{\eta_k^N D_{k,n}^N(1)}\right)$$
$$- \int \mathbb{Q}_n^N(\mathrm{d}x_{0:n}) t_k(x_{k-1}, x_k) \left(\varphi_n(x_n) - \eta_n^N(\varphi_n)\right)$$
$$= \sum_{p=k}^{n-1} \int \frac{\eta_p^N D_{p,n}^N(\mathrm{d}x_{0:n})}{\eta_p^N D_{p,n}^N(1)} t_k(x_{k-1}, x_k) \left(\varphi_n(x_n) - \frac{\eta_p^N D_{p,n}^N(\varphi_n)}{\eta_p^N D_{p,n}^N(1)}\right)$$
$$- \sum_{p=k}^{n-1} \int \frac{\eta_{p+1}^N D_{p+1,n}^N(\mathrm{d}x_{0:n})}{\eta_{p+1}^N D_{p+1,n}^N(1)} t_k(x_{k-1}, x_k) \left(\varphi_n(x_n) - \frac{\eta_{p+1}^N D_{p+1,n}^N(\varphi_n)}{\eta_{p+1}^N D_{p+1,n}^N(1)}\right)$$
$$= \sum_{p=k}^{n-1} \int \left(\frac{\eta_p^N D_{p,n}^N(\mathrm{d}x_{0:n})}{\eta_p^N D_{p,n}^N(1)} \right.$$
$$\left. - \frac{\eta_{p+1}^N D_{p+1,n}^N(\mathrm{d}x_{0:n})}{\eta_{p+1}^N D_{p+1,n}^N(1)} \right) t_k(x_{k-1}, x_k) \left(\varphi_n(x_n) - \frac{\eta_p^N D_{p,n}^N(\varphi_n)}{\eta_p^N D_{p,n}^N(1)}\right)$$
$$- \sum_{p=k}^{n-1} \left(\frac{\eta_p^N D_{p,n}^N(\varphi_n)}{\eta_p^N D_{p,n}^N(1)} - \frac{\eta_{p+1}^N D_{p+1,n}^N(\varphi_n)}{\eta_{p+1}^N D_{p+1,n}^N(1)} \right) \left(\frac{\eta_{p+1}^N D_{p+1,n}^N(t_k)}{\eta_{p+1}^N D_{p+1,n}^N(1)} - \frac{\eta_p^N D_{p,n}^N(t_k)}{\eta_p^N D_{p,n}^N(1)} \right)$$
$$- \sum_{p=k}^{n-1} \left(\frac{\eta_p^N D_{p,n}^N(\varphi_n)}{\eta_p^N D_{p,n}^N(1)} - \frac{\eta_{p+1}^N D_{p+1,n}^N(\varphi_n)}{\eta_{p+1}^N D_{p+1,n}^N(1)} \right) \frac{\eta_p^N D_{p,n}^N(t_k)}{\eta_p^N D_{p,n}^N(1)}$$

(A.7)
$$= \sum_{p=k}^{n-1} \int \left(\frac{\eta_p^N D_{p,n}^N(\mathrm{d}x_{0:n})}{\eta_p^N D_{p,n}^N(1)} - \frac{\eta_{p+1}^N D_{p+1,n}^N(\mathrm{d}x_{0:n})}{\eta_{p+1}^N D_{p+1,n}^N(1)} \right)$$
(A.8)
$$\times \left(t_k(x_{k-1}, x_k) - \frac{\eta_p^N D_{p,n}^N(t_k)}{\eta_p^N D_{p,n}^N(1)} \right) \left(\varphi_n(x_n) - \frac{\eta_p^N D_{p,n}^N(\varphi_n)}{\eta_p^N D_{p,n}^N(1)}\right)$$
(A.9)
$$- \sum_{p=k}^{n-1} \left(\frac{\eta_p^N D_{p,n}^N(\varphi_n)}{\eta_p^N D_{p,n}^N(1)} - \frac{\eta_{p+1}^N D_{p+1,n}^N(\varphi_n)}{\eta_{p+1}^N D_{p+1,n}^N(1)} \right) \left(\frac{\eta_{p+1}^N D_{p+1,n}^N(t_k)}{\eta_{p+1}^N D_{p+1,n}^N(1)} - \frac{\eta_p^N D_{p,n}^N(t_k)}{\eta_p^N D_{p,n}^N(1)} \right).$$

For the first equality, note that $\eta_n^N D_{n,n}^N(\mathrm{d}x_{0:n}) = \mathbb{Q}_n^N(\mathrm{d}x_{0:n})$. It is straightforward to establish that

$$\text{(A.10)} \qquad \eta_p^N D_{p,n}^N(\mathrm{d}x_{0:n}) / \eta_p^N\left(g(y_p|\cdot)\right) = \widetilde{\eta}_{p+1}^N D_{p+1,n}^N(\mathrm{d}x_{0:n}),$$

which is due to

$$\begin{aligned}
&\frac{\eta_p^N(\mathrm{d}x_p)}{\eta_p^N(g(y_p|\cdot))}\prod_{j=p}^{n-1}Q_{j+1}(x_j,\mathrm{d}x_{j+1})\\
&=\frac{\eta_p^N(\mathrm{d}x_p)g(y_p|x_p)f(x_{p+1}|x_p)}{\eta_p^N\left(g(y_p|\cdot)f(x_{p+1}|\cdot)\right)}\frac{\mathrm{d}x_{p+1}\eta_p^N\left(g(y_p|\cdot)f(x_{p+1}|\cdot)\right)}{\eta_p^N\left(g(y_p|\cdot)\right)}\prod_{j=p+1}^{n-1}Q_{j+1}(x_j,\mathrm{d}x_{j+1})\\
&=M_{p+1}^N(x_{p+1},\mathrm{d}x_p)\widetilde{\eta}_{p+1}^N(\mathrm{d}x_{p+1})\prod_{j=p+1}^{n-1}Q_{j+1}(x_j,\mathrm{d}x_{j+1}).
\end{aligned}$$

Thus

$$\begin{aligned}
&\frac{\eta_p^N D_{p,n}^N(\mathrm{d}x_{0:p+1},\mathrm{d}x_n)}{\eta_p^N D_{p,n}^N(1)}-\frac{\eta_{p+1}^N D_{p+1,n}^N(\mathrm{d}x_{0:p+1},\mathrm{d}x_n)}{\eta_{p+1}^N D_{p+1,n}^N(1)}\\
&\quad=\frac{\widetilde{\eta}_{p+1}^N D_{p+1,n}^N(\mathrm{d}x_{0:p+1},\mathrm{d}x_n)}{\widetilde{\eta}_{p+1}^N D_{p+1,n}^N(1)}-\frac{\eta_{p+1}^N D_{p+1,n}^N(\mathrm{d}x_{0:p+1},\mathrm{d}x_n)}{\eta_{p+1}^N D_{p+1,n}^N(1)}\\
&\quad=\left(\frac{\widetilde{\eta}_{p+1}^N(\mathrm{d}x_{p+1})Q_{p+1,n}(1)(x_{p+1})}{\widetilde{\eta}_{p+1}^N Q_{p+1,n}(1)}\right.\\
&\qquad\left.-\frac{\eta_{p+1}^N(\mathrm{d}x_{p+1})Q_{p+1,n}(1)(x_{p+1})}{\eta_{p+1}^N Q_{p+1,n}(1)}\right)\mathcal{M}_{p+1}^N(x_{p+1},\mathrm{d}x_{0:p})\frac{Q_{p+1,n}(x_{p+1},\mathrm{d}x_n)}{Q_{p+1,n}(1)(x_{p+1})}.
\end{aligned}\tag{A.11}$$

In the first line, variables $x_{p+2:n-1}$ of the measures $\eta_p D_{p,n}^N(\mathrm{d}x_{0:n})$ and $\eta_{p+1}D_{p+1,n}^N(\mathrm{d}x_{0:n})$ are integrated out while the second line follows from (A.10). Using (A.11), the term (A.8) can be expressed as

$$\begin{aligned}
&\sum_{p=k}^{n-1}\int\left(\frac{\widetilde{\eta}_{p+1}^N(\mathrm{d}x_{p+1})Q_{p+1,n}(1)(x_{p+1})}{\widetilde{\eta}_{p+1}^N Q_{p+1,n}(1)}-\frac{\eta_{p+1}^N(\mathrm{d}x_{p+1})Q_{p+1,n}(1)(x_{p+1})}{\eta_{p+1}^N Q_{p+1,n}(1)}\right)\\
&\quad\times P_{p+1,n}\left(\varphi_n-\frac{\eta_p^N D_{p,n}^N(\varphi_n)}{\eta_p^N D_{p,n}^N(1)}\right)(x_{p+1})\mathcal{M}_{p+1}^N\left(t_k-\frac{\widetilde{\eta}_{p+1}^N D_{p+1,n}^N(t_k)}{\widetilde{\eta}_{p+1}^N D_{p+1,n}^N(1)}\right)(x_{p+1}).
\end{aligned}$$

Note that by (3.3), (A.2), and (A.3),

$$\left|P_{p+1,n}\left(\varphi_n-\frac{\eta_p^N D_{p,n}^N(\varphi_n)}{\eta_p^N D_{p,n}^N(1)}\right)(x_{p+1})\right|\le\beta\left(\frac{Q_{p+1,n}(x_{p+1},\mathrm{d}x_n)}{Q_{p+1,n}(1)(x_{p+1})}\right),$$

$$\left|\mathcal{M}_{p+1}^N\left(t_k-\frac{\widetilde{\eta}_{p+1}^N D_{p+1,n}^N(t_k)}{\widetilde{\eta}_{p+1}^N D_{p+1,n}^N(1)}\right)(x_{p+1})\right|\le C\beta\left(M_{p+1}^N\dots M_{k+1}^N\right).$$

Thus by (A.2) and Lemma A.6, we conclude that there exists a finite constant C_r (depending only on r),

$$\begin{aligned}
&\sum_{p=k}^{n-1}\sqrt{N}\mathbb{E}\left\{\left|\int\left(t_k\left(x_{k-1},x_k\right)-\frac{\eta_p^N D_{p,n}^N(t_k)}{\eta_p^N D_{p,n}^N(1)}\right)\left(\varphi_n(x_n)-\frac{\eta_p^N D_{p,n}^N(\varphi_n)}{\eta_p^N D_{p,n}^N(1)}\right)\right.\right.\\
&\qquad\left.\left.\times\left(\frac{\eta_p^N D_{p,n}^N(\mathrm{d}x_{0:n})}{\eta_p^N D_{p,n}^N(1)}-\frac{\eta_{p+1}^N D_{p+1,n}^N(\mathrm{d}x_{0:n})}{\eta_{p+1}^N D_{p+1,n}^N(1)}\right)\right|^r\right\}^{\frac{1}{r}}\le(n-k)C_r\overline{\rho}^{n-k}.
\end{aligned}\tag{A.12}$$

For the term (A.9), it follows from (A.11) that

$$\begin{aligned}&\frac{\eta_{p+1}^N D_{p+1,n}^N(t_k)}{\eta_{p+1}^N D_{p+1,n}^N(1)} - \frac{\eta_p^N D_{p,n}^N(t_k)}{\eta_p^N D_{p,n}^N(1)}\\ &\qquad = \int \frac{\eta_{p+1}^N(\mathrm{d}x_{p+1})Q_{p+1,n}(1)(x_{p+1})}{\eta_{p+1}^N Q_{p+1,n}(1)}\Bigg(\mathcal{M}_{p+1}^N(t_k)(x_{p+1})\\ &\qquad\quad - \frac{\widetilde{\eta}_{p+1}^N\left(Q_{p+1,n}(1)\mathcal{M}_{p+1}^N(t_k)\right)}{\widetilde{\eta}_{p+1}^N Q_{p+1,n}(1)}\Bigg).\end{aligned}$$

Thus, using (3.3) and (A.3), there exists some nonrandom constant C such that the following bound holds almost surely for all integers $k \le p < n$, N:

$$\left|\frac{\eta_{p+1}^N D_{p+1,n}^N(t_k)}{\eta_{p+1}^N D_{p+1,n}^N(1)} - \frac{\eta_p^N D_{p,n}^N(t_k)}{\eta_p^N D_{p,n}^N(1)}\right| \le C\overline{\rho}^{p-k+1}.$$

Combine this bound with Lemma A.3 to conclude that there exists a finite (nonrandom) constant C_r (depending only on r) such that for all integers $k \le p < n$, N:

(A.13)
$$\begin{aligned}&\sqrt{N}\mathbb{E}\left\{\left|\left(\frac{\eta_p^N D_{p,n}^N(\varphi_n)}{\eta_p^N D_{p,n}^N(1)} - \frac{\eta_{p+1}^N D_{p+1,n}^N(\varphi_n)}{\eta_{p+1}^N D_{p+1,n}^N(1)}\right)\left(\frac{\eta_{p+1}^N D_{p+1,n}^N(t_k)}{\eta_{p+1}^N D_{p+1,n}^N(1)} - \frac{\eta_p^N D_{p,n}^N(t_k)}{\eta_p^N D_{p,n}^N(1)}\right)\right|^r\right\}^{\frac{1}{r}}\\ &\quad \le C_r\overline{\rho}^{n-k}.\end{aligned}$$

The result now follows from (A.12) and (A.13). ∎

LEMMA A.9. *Assume Assumption (A). For any $r \ge 1$ there exists a constant C_r such that for all θ, y, $0 \le k < n$, $N \ge 1$, $\varphi_n \in Osc_1(\mathcal{X})$,*

$$\begin{aligned}\text{(A.14)}\quad &\sqrt{N}\mathbb{E}_\theta^y\left\{\left|\int \frac{\eta_{\theta,k}^N D_{\theta,k,n}^N(\mathrm{d}x_{0:n})}{\eta_{\theta,k}^N D_{\theta,k,n}^N(1)} t_{\theta,k}(x_{k-1},x_k)\left(\varphi_n(x_n) - \frac{\eta_{\theta,k}^N D_{\theta,k,n}^N(\varphi_n)}{\eta_{\theta,k}^N D_{\theta,k,n}^N(1)}\right)\right.\right.\\ &\qquad\left.\left. - \int \mathbb{Q}_{\theta,n}(\mathrm{d}x_{0:n})t_{\theta,k}(x_{k-1},x_k)\left(\varphi_n(x_n) - \eta_{\theta,n}(\varphi_n)\right)\right|^r\right\}^{\frac{1}{r}} \le C_r\overline{\rho}^{n-k}.\end{aligned}$$

Proof. The following decomposition holds:

$$\begin{aligned}&\int \frac{\eta_k^N D_{k,n}^N(\mathrm{d}x_{0:n})}{\eta_k^N D_{k,n}^N(1)} t_k(x_{k-1},x_k)\left(\varphi_n(x_n) - \frac{\eta_k^N D_{k,n}^N(\varphi_n)}{\eta_k^N D_{k,n}^N(1)}\right)\\ &= \int \mathbb{Q}_n(\mathrm{d}x_{0:n})t_k(x_{k-1},x_k)\left(\varphi_n(x_n) - \eta_n(\varphi_n)\right)\\ \text{(A.15)}\quad &+ \int\left(\frac{\eta_k^N D_{k,n}^N(\mathrm{d}x_{0:n})}{\eta_k^N D_{k,n}^N(1)} - \mathbb{Q}_n(\mathrm{d}x_{0:n})\right)t_k(x_{k-1},x_k)\left(\varphi_n(x_n) - \eta_n(\varphi_n)\right)\\ \text{(A.16)}\quad &+ \left(\eta_n(\varphi_n) - \frac{\eta_k^N D_{k,n}^N(\varphi_n)}{\eta_k^N D_{k,n}^N(1)}\right)\frac{\eta_k^N D_{k,n}^N(t_k)}{\eta_k^N D_{k,n}^N(1)}.\end{aligned}$$

To study the errors, term (A.15) may be decomposed as

$$\int \left(\frac{\eta_k^N D_{k,n}^N(\mathrm{d}x_{0:n})}{\eta_k^N D_{k,n}^N(1)} - \mathbb{Q}_n(\mathrm{d}x_{0:n}) \right) t_k\left(x_{k-1}, x_k\right)\left(\varphi_n(x_n) - \eta_n(\varphi_n)\right)$$
$$= \sum_{p=0}^{k} \int \left(\frac{\eta_p^N D_{p,n}^N(\mathrm{d}x_{0:n})}{\eta_p^N D_{p,n}^N(1)} - \frac{\widetilde{\eta}_p^N D_{p,n}^N(\mathrm{d}x_{0:n})}{\widetilde{\eta}_p^N D_{p,n}^N(1)} \right) t_k\left(x_{k-1}, x_k\right)\left(\varphi_n(x_n) - \eta_n(\varphi_n)\right)$$

with the convention that $\widetilde{\eta}_0^N = \Phi_0\left(\eta_{-1}^N\right) = \eta_0$. The term corresponding to $p = k$ can be expressed as

$$\int \left(\frac{\eta_k^N(\mathrm{d}x_k) Q_{k,n}(1)(x_k)}{\eta_k^N Q_{k,n}(1)} - \frac{\widetilde{\eta}_k^N(\mathrm{d}x_k) Q_{k,n}(1)(x_k)}{\widetilde{\eta}_k^N Q_{k,n}(1)} \right)$$
$$\times\ M_k^N(x_k, \mathrm{d}x_{k-1}) t_k\left(x_{k-1}, x_k\right) P_{k,n}(\varphi_n - \eta_n(\varphi_n))(x_k).$$

Using Lemma A.1 and Remark A.2,

$$\sqrt{N}\mathbb{E}\left\{ \left| \int \left(\frac{\eta_k^N(\mathrm{d}x_k) Q_{k,n}(1)(x_k)}{\eta_k^N Q_{k,n}(1)} - \frac{\widetilde{\eta}_k^N(\mathrm{d}x_k) Q_{k,n}(1)(x_k)}{\widetilde{\eta}_k^N Q_{k,n}(1)} \right) M_k^N\left(t_k\right)\left(x_k\right) P_{k,n}(\varphi_n \right.\right.$$
$$\left.\left. - \eta_n(\varphi_n))(x_k) \right|^r \right\}^{\frac{1}{r}}$$
$$\leq C_r \overline{\rho}^{n-k}.$$

Similarly, the pth term when $p < k$ can be expressed as

$$\int \left(\frac{\eta_p^N D_{p,n}^N(\mathrm{d}x_{0:n})}{\eta_p^N D_{p,n}^N(1)} - \frac{\widetilde{\eta}_p^N D_{p,n}^N(\mathrm{d}x_{0:n})}{\widetilde{\eta}_p^N D_{p,n}^N(1)} \right) t_k\left(x_{k-1}, x_k\right)\left(\varphi_n(x_n) - \eta_n(\varphi_n)\right)$$
$$= \int \left(\frac{\Phi_{p,k-1}(\eta_p^N)(\mathrm{d}x_{k-1}) Q_{k-1,n}(1)(x_{k-1})}{\Phi_{p,k-1}(\eta_p^N) Q_{k-1,n}(1)} - \frac{\Phi_{p,k-1}(\widetilde{\eta}_p^N)(\mathrm{d}x_{k-1}) Q_{k-1,n}(1)(x_{k-1})}{\Phi_{p,k-1}(\widetilde{\eta}_p^N) Q_{k-1,n}(1)} \right)$$
$$\times \int \frac{Q_k(x_{k-1}, \mathrm{d}x_k) Q_{k,n}(1)(x_k)}{Q_{k-1,n}(1)(x_{k-1})} t_k\left(x_{k-1}, x_k\right) P_{k,n}\left(\varphi_n - \eta_n(\varphi_n)\right)(x_k).$$

Using Lemma A.6 for the outer integral (recall $\Phi_{p,k-1}(\widetilde{\eta}_p^N) = \Phi_{p-1,k-1}(\eta_{p-1}^N)$),

$$\sqrt{N}\mathbb{E}$$
$$\times \left\{ \left| \int \left(\frac{\eta_p^N D_{p,n}^N(\mathrm{d}x_{0:n})}{\eta_p^N D_{p,n}^N(1)} - \frac{\Phi_p\left(\eta_{p-1}^N\right) D_{p,n}^N(\mathrm{d}x_{0:n})}{\Phi_p\left(\eta_{p-1}^N\right) D_{p,n}^N(1)} \right) t_k\left(x_{k-1}, x_k\right)\left(\varphi_n(x_n) - \eta_n(\varphi_n)\right) \right|^r \right\}^{\frac{1}{r}}$$
$$\leq C_r \overline{\rho}^{n-k} \overline{\rho}^{k-1-p}.$$

Combining both cases for p yields

(A.17)
$$\sqrt{N}\mathbb{E}\left\{ \left| \int \left(\frac{\eta_k^N D_{k,n}^N(\mathrm{d}x_{0:n})}{\eta_k^N D_{k,n}^N(1)} - \mathbb{Q}_n(\mathrm{d}x_{0:n}) \right) t_k\left(x_{k-1}, x_k\right)\left(\varphi_n(x_n) - \eta_n(\varphi_n)\right) \right|^r \right\}^{\frac{1}{r}}$$
$$\leq C_r \overline{\rho}^{n-k} \sum_{p=0}^{k-1} \overline{\rho}^{k-1-p} + C_r \overline{\rho}^{n-k}$$
$$\leq C_r \overline{\rho}^{n-k} \left(1 + \frac{1}{1 - \overline{\rho}} \right).$$

For (A.16), Lemma A.5 yields the following estimate:

$$\sqrt{N}\mathbb{E}\left\{\left|\left(\eta_n(\varphi_n) - \frac{\eta_k^N D_{k,n}^N(\varphi_n)}{\eta_k^N D_{k,n}^N(1)}\right)\frac{\eta_k^N D_{k,n}^N(t_k)}{\eta_k^N D_{k,n}^N(1)}\right|^r\right\}^{\frac{1}{r}} \leq C_r\overline{\rho}^{n-k}. \tag{A.18}$$

The proof is completed by summing the bounds in (A.17), (A.18) and inflating constant C_r appropriately. ∎

A.1. Proof of Theorem 3.1. We may write

$$\begin{aligned}\zeta_n^N(\varphi_n) - \zeta_n(\varphi_n) = \sum_{k=0}^{n} \int \mathbb{Q}_n^N(\mathrm{d}x_{0:n})t_k(x_{k-1},x_k)\left(\varphi_n(x_n) - \eta_n^N(\varphi_n)\right)\\ - \int \mathbb{Q}_n(\mathrm{d}x_{0:n})t_k(x_{k-1},x_k)\left(\varphi_n(x_n) - \eta_n(\varphi_n)\right).\end{aligned}$$

To prove Theorem 3.1, it will be shown that the error due to the kth term in this expression is

$$\begin{aligned}&\sqrt{N}\mathbb{E}\left\{\left|\int \mathbb{Q}_n^N(\mathrm{d}x_{0:n})t_k(x_{k-1},x_k)\left(\varphi_n(x_n) - \eta_n^N(\varphi_n)\right)\right.\right.\\ &\quad\left.\left.- \int \mathbb{Q}_n(\mathrm{d}x_{0:n})t_k(x_{k-1},x_k)\left(\varphi_n(x_n) - \eta_n(\varphi_n)\right)\right|^r\right\}^{\frac{1}{r}} \leq (n-k+1)C_r\overline{\rho}^{n-k},\end{aligned}$$

where constant C_r depends only on r and the bounds in Assumption (A) (through the estimates $\overline{\rho}$ and $\rho^2\delta^2$ in (A.2) as well as the bounds on the score):

$$\begin{aligned}&\int \mathbb{Q}_n^N(\mathrm{d}x_{0:n})t_k(x_{k-1},x_k)\left(\varphi_n(x_n) - \eta_n^N(\varphi_n)\right)\\ &= \int \mathbb{Q}_n(\mathrm{d}x_{0:n})t_k(x_{k-1},x_k)\left(\varphi_n(x_n) - \eta_n(\varphi_n)\right)\\ &\quad+ \int \mathbb{Q}_n^N(\mathrm{d}x_{0:n})t_k(x_{k-1},x_k)\left(\varphi_n(x_n) - \eta_n^N(\varphi_n)\right)\\ &\qquad - \int \frac{\eta_k^N D_{k,n}^N(\mathrm{d}x_{0:n})}{\eta_k^N D_{k,n}^N(1)}t_k(x_{k-1},x_k)\left(\varphi_n(x_n) - \frac{\eta_k^N D_{k,n}^N(\varphi_n)}{\eta_k^N D_{k,n}^N(1)}\right) &\text{(A.19)}\\ &\quad+ \int \frac{\eta_k^N D_{k,n}^N(\mathrm{d}x_{0:n})}{\eta_k^N D_{k,n}^N(1)}t_k(x_{k-1},x_k)\left(\varphi_n(x_n) - \frac{\eta_k^N D_{k,n}^N(\varphi_n)}{\eta_k^N D_{k,n}^N(1)}\right) &\text{(A.20)}\\ &\qquad - \int \mathbb{Q}_n(\mathrm{d}x_{0:n})t_k(x_{k-1},x_k)\left(\varphi_n(x_n) - \eta_n(\varphi_n)\right).\end{aligned}$$

The proof is completed by summing the bounds in Lemma A.8 for (A.19) and Lemma A.9 for (A.20), and inflating constant C_r appropriately. ∎

A.2. Proof of Theorem 3.2. The following result, which characterizes the asymptotic behavior of the local sampling errors defined in (3.1), is proved in [3, Theorem 9.3.1].

LEMMA A.10. *Let* $\{\varphi_n\}_{n\geq 0} \subset \mathcal{B}(\mathcal{X})$. *For any* θ, y, $n \geq 0$, *the random vector* $(V_{\theta,0}^N(\varphi_0),\ldots,V_{\theta,n}^N(\varphi_n))$ *converges in law, as* $N \to \infty$, *to* $(V_{\theta,0}(\varphi_0),\ldots,V_{\theta,n}(\varphi_n))$, *where* $V_{\theta,i}$ *is defined in* (3.4).

The following multivariate fluctuation theorem, first proved under slightly different assumptions in [8], is needed. See also [9] for a related study.

THEOREM A.11. *Assume Assumption (A). For any θ, y, $n \geq 0$, $F_n \in \mathcal{B}(\mathcal{X}^{n+1})$, it follows that $\sqrt{N}(\mathbb{Q}_{\theta,n}^N - \mathbb{Q}_{\theta,n})(F_n)$ converges in law, as $N \to \infty$, to the centered Gaussian random variable*

$$\sum_{p=0}^{n} V_{\theta,p}\left(G_{\theta,p,n}\ \frac{D_{\theta,p,n}(F_n - \mathbb{Q}_{\theta,n}(F_n))}{D_{\theta,p,n}(1)}\right),$$

where $V_{\theta,p}$ is defined in (3.4).

Proof. Let

$$\gamma_n = \prod_{k=0}^{n-1} \eta_k(g(\,y_k|\,.))$$

and define the unnormalized measure

$$\Gamma_n = \gamma_n \mathbb{Q}_n.$$

The corresponding particle approximation is $\Gamma_n^N = \gamma_n^N \mathbb{Q}_n^N$, where $\gamma_n^N = \prod_{k=0}^{n-1} \eta_k^N(g(\,y_k|\,.))$. The result is proven by studying the limit of $\sqrt{N}\left(\Gamma_n^N - \Gamma_n\right)$ since

$$[\mathbb{Q}_n^N - \mathbb{Q}_n](F_n) = \frac{1}{\gamma_n^N}\left[\Gamma_n^N - \Gamma_n\right](F_n - \mathbb{Q}_n(F_n)).$$

Note that Lemma A.4 implies γ_n^N converges almost surely to γ_n. The key to studying the limit of $\sqrt{N}\left(\Gamma_n^N - \Gamma_n\right)$ is the decomposition

$$\sqrt{N}\left[\Gamma_n^N - \Gamma_n\right](F_n) = \sum_{p=0}^{n} \gamma_p^N\ V_p^N\left(D_{p,n}(F_n)\right) + R_n^N(F_n),$$

where the remainder term is

$$R_n^N(F_n) := \sum_{p=0}^{n} \gamma_p^N\ V_p^N\left(F_{p,n}^N\right) \quad \text{and} \quad F_{p,n}^N := [D_{p,n}^N - D_{p,n}](F_n).$$

By Slutsky's lemma and by the continuous mapping theorem (see [25]) it suffices to show that $R_n^N(F_n)$ converges to 0, in probability, as $N \to \infty$. To prove this, it will be established that $\mathbb{E}\left(R_n^N(F_n)^2\right)$ is $\mathcal{O}(N^{-1})$. Since

$$\mathbb{E}\left\{R_n^N(F_n)^2\right\} = \sum_{p=0}^{n} \mathbb{E}\left\{\left(\gamma_p^N\ V_p^N\left(F_{p,n}^N\right)\right)^2\right\},$$

and $\left|\gamma_p^N\right| \leq c_p$ almost surely, where c_p is some nonrandom constant which can be derived using Assumption (A), it suffices to prove that $\mathbb{E}(V_p^N(F_{p,n}^N)^2)$ is $\mathcal{O}(N^{-1})$. By expanding the square one arrives at

$$\mathbb{E}\left(V_p^N\left(F_{p,n}^N\right)^2\ \mid \mathcal{F}_{p-1}^N\right) \leq \Phi_p\left(\eta_{p-1}^N\right)\left(\left(F_{p,n}^N\right)^2\right).$$

By Assumption (A), for any $x_{p-1} \in \mathcal{X}$,

$$\Phi_p\left(\eta_{p-1}^N\right)\left(\left(F_{p,n}^N\right)^2\right) \leq \rho^2 \int \mathrm{d}x_p \; f(x_p|\, x_{p-1}) \; F_{p,n}^N(x_p)^2.$$

By Lemma A.7, $\mathbb{E}(V_p^N\left(F_{p,n}^N\right)^2)$ is $\mathcal{O}(N^{-1})$. □

The next lemma is needed to quantify the variance of the particle estimate of the filter gradient computed using the path-based method. Note that this lemma does not require the mixing of the hidden chain. We refer the reader to [6] for a propagation of chaos analysis.

For any θ, $y = \{y_n\}_{n\geq 0}$, let $\{\mathcal{V}_{\theta,n}\}_{n\geq 0}$ be a sequence of independent centered Gaussian random fields defined as follows. For any sequence of functions $\{F_n \in \mathcal{B}(\mathcal{X}^{n+1})\}_{n\geq 0}$ and any $p \geq 0$, $\{\mathcal{V}_{\theta,n}(F_n)\}_{n=0}^p$ is a collection of independent zero-mean Gaussian random variables with variances given by

$$\mathbb{E}_\theta(F_n(X_{0:n})^2|y_{0:n-1}) - \mathbb{E}_\theta(F_n(X_{0:n})|y_{0:n-1})^2. \tag{A.21}$$

Lemma A.12. *Let $\{\delta_\theta\}_{\theta\in\Theta} \subset [1,\infty)$ and assume $\delta_\theta^{-1} \leq g_\theta(y|x) \leq \delta_\theta$ for all $(x,y,\theta) \in \mathcal{X}\times\mathcal{Y}\times\Theta$. For any θ, y, $n \geq 0$, $F_n \in \mathcal{B}(\mathcal{X}^{n+1})$, we have that $\sqrt{N}\left(p_\theta^N(\mathrm{d}x_{0:n}|y_{0:n-1}) - \mathbb{Q}_{\theta,n}\right)(F_n)$ converges in law, as $N \to \infty$, to the centered Gaussian random variable*

$$\sum_{p=0}^{n} \mathcal{V}_{\theta,p}\left(G_{\theta,p,n}\; F_{\theta,p,n}\right),$$

where $G_{\theta,p,n}$ was defined in (1.7) *and*

$$F_{\theta,p,n} = \mathbb{E}_\theta(F(X_{0:n})|x_{0:p}, y_{p+1:n-1}) - \mathbb{Q}_{\theta,n}(F_n).$$

A.2.1. Proof of Theorem 3.2. It follows from Algorithm 1 that

$$\begin{aligned}&\left(\zeta_n^N - \zeta_n\right)(\varphi_n)\\&= \mathbb{Q}_n^N(\varphi_n T_n) - \mathbb{Q}_n(\varphi_n T_n) + \mathbb{Q}_n(\varphi_n)\mathbb{Q}_n(T_n) - \mathbb{Q}_n^N(\varphi_n)\mathbb{Q}_n^N(T_n).\end{aligned} \tag{A.22}$$

The second term on the right-hand side of the equality can be expressed as

$$\begin{aligned}&\mathbb{Q}_n(\varphi_n)\mathbb{Q}_n(T_n) - \mathbb{Q}_n^N(\varphi_n)\mathbb{Q}_n^N(T_n)\\&= \mathbb{Q}_n(\varphi_n\mathbb{Q}_n(T_n) + \mathbb{Q}_n(\varphi_n)T_n) - \mathbb{Q}_n^N(\varphi_n\mathbb{Q}_n(T_n) + \mathbb{Q}_n(\varphi_n)T_n)\\&\quad + \left(\mathbb{Q}_n^N(\varphi_n) - \mathbb{Q}_n(\varphi_n)\right)\left(\mathbb{Q}_n(T_n) - \mathbb{Q}_n^N(T_n)\right).\end{aligned} \tag{A.23}$$

Combining the two expressions in (A.22) and (A.23) gives

$$\begin{aligned}&\left(\zeta_n^N - \zeta_n\right)(\varphi_n)\\&= \mathbb{Q}_n^N\left(\left(\varphi_n - \mathbb{Q}_n(\varphi_n)\right)\left(T_n - \mathbb{Q}_n(T_n)\right)\right)\\&- \mathbb{Q}_n\left(\left(\varphi_n - \mathbb{Q}_n(\varphi_n)\right)\left(T_n - \mathbb{Q}_n(T_n)\right)\right)\\&+ \left(\mathbb{Q}_n^N(\varphi_n) - \mathbb{Q}_n(\varphi_n)\right)\left(\mathbb{Q}_n(T_n) - \mathbb{Q}_n^N(T_n)\right).\end{aligned}$$

Using Lemma A.4 with $r = 2$ and Chebyshev's inequality, we see that $\left(\mathbb{Q}_n^N(\varphi_n) - \mathbb{Q}_n(\varphi_n)\right)$ converges in probability to 0. Theorem A.11 can now be invoked with Slutsky's theorem to arrive at the stated result in (3.5).

Moving on to the uniform bound on the variance, let

$$T_n - \mathbb{Q}_n(T_n) = \sum_{k=0}^{n} \widetilde{t}_k,$$
$$\widetilde{t}_k = t_k - \mathbb{Q}_n(t_k),$$
$$\widetilde{\varphi}_n = \varphi_n - \mathbb{Q}_n(\varphi_n).$$

Also, the argument of V_p can be expressed as

$$\phi_p(x_p) = \frac{Q_{p,n}(1)(x_p)}{\eta_p Q_{p,n}(1)} \sum_{k=0}^{n} \frac{D_{p,n}\left(\widetilde{\varphi}_n \widetilde{t}_k - \mathbb{Q}_n\left(\widetilde{\varphi}_n \widetilde{t}_k\right)\right)(x_p)}{D_{p,n}(1)(x_p)}.$$

It is straightforward to see that $\eta_p(\phi_p) = 0$. Therefore, the variance (see (3.4)) now simplifies to

$$\text{var} \sum_{p=0}^{n} V_p\left(G_{p,n}\ \frac{D_{p,n}(F_n - \mathbb{Q}_n(F_n))}{D_{p,n}(1)}\right) = \sum_{p=0}^{n} \eta_p(\phi_p^2). \tag{A.24}$$

Consider the function ϕ_p. For $p \le k-1$,

$$\begin{aligned}
&\frac{D_{p,n}\left(\widetilde{\varphi}_n \widetilde{t}_k - \mathbb{Q}_n\left(\widetilde{\varphi}_n \widetilde{t}_k\right)\right)(x_p)}{D_{p,n}(1)(x_p)} \\
&\quad = \int \frac{\eta_p(\mathrm{d}x'_p) Q_{p,n}(1)(x'_p)}{\eta_p Q_{p,n}(1)} \\
&\qquad \times \int \left(\frac{Q_{p,k-1}(x_p, \mathrm{d}x_{k-1}) Q_{k-1,n}(1)(x_{k-1})}{Q_{p,n}(1)(x_p)} - \frac{Q_{p,k-1}(x'_p, \mathrm{d}x_{k-1}) Q_{k-1,n}(1)(x_{k-1})}{Q_{p,n}(1)(x'_p)}\right) \\
&\qquad \times \int \frac{Q_k(x_{k-1}, \mathrm{d}x_k) Q_{k,n}(1)(x_k)}{Q_{k-1,n}(1)(x_{k-1})} \widetilde{t}_k(x_{k-1}, x_k) P_{k,n}(\widetilde{\varphi}_n)(x_k).
\end{aligned}$$

Using the estimates in (3.3) and (A.2), this function is bounded by

$$\sup_{x_p} \left| \frac{D_{p,n}\left(\widetilde{\varphi}_n \widetilde{t}_k - \mathbb{Q}_n\left(\widetilde{\varphi}_n \widetilde{t}_k\right)\right)(x_p)}{D_{p,n}(1)(x_p)} \right| \le C \overline{\rho}^{\,n-1-p} \tag{A.25}$$

for some constant C. When $p \ge k$,

$$\begin{aligned}
&\frac{D_{p,n}\left(\widetilde{\varphi}_n \widetilde{t}_k - \mathbb{Q}_n\left(\widetilde{\varphi}_n \widetilde{t}_k\right)\right)(x_p)}{D_{p,n}(1)(x_p)} \\
&\quad = \int \frac{\eta_p(\mathrm{d}x'_p) Q_{p,n}(1)(x'_p)}{\eta_p Q_{p,n}(1)} \left(\mathcal{M}_p(\widetilde{t}_k)(x_p) P_{p,n}(\widetilde{\varphi}_n)(x_p) - \mathcal{M}_p(\widetilde{t}_k)(x'_p) P_{p,n}(\widetilde{\varphi}_n)(x'_p)\right).
\end{aligned}$$

Again using the estimates in (3.3), (A.2), and (A.3),

$$\sup_{x_p} \left| \frac{D_{p,n}\left(\widetilde{\varphi}_n \widetilde{t}_k - \mathbb{Q}_n\left(\widetilde{\varphi}_n \widetilde{t}_k\right)\right)(x_p)}{D_{p,n}(1)(x_p)} \right| \le C \overline{\rho}^{\,n-k}. \tag{A.26}$$

Combining (A.25) and (A.26),

$$\sup_{x_p} |\phi_p(x_p)| \le \frac{C \overline{\rho}^{\,n-p}}{1 - \overline{\rho}} + C \overline{\rho}^{\,n-p-1}(n-p),$$

for $0 \le p \le n$. Combining this bound with (A.24) will establish the result. ∎

Acknowledgment. We are grateful to Sinan Yildirim for carefully reading this paper.

REFERENCES

[1] F. Cérou, F. Le Gland, and N. J. Newton, *Stochastic particle methods for linear tangent filtering equations*, in Optimal Control and PDE's - Innovations and Applications, IOS Press, Amsterdam, 2001, pp. 231–240.

[2] P. Coquelin, R. Deguest, and R. Munos, *Particle filter based policy gradient in POMDPS*, in Advances in Neural Information Processing Systems 21, 2009, MIT Press, Cambridge, MA, pp. 337–344.

[3] P. Del Moral, *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*, Springer-Verlag, New York, 2004.

[4] P. Del Moral and A. Doucet, *On a class of genealogical and interacting Metropolis models*, in Lecture Notes in Math. 1832, Springer, Berlin, 2003, pp. 415–446.

[5] P. Del Moral and L. Miclo, *Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering*, in Lecture Notes in Math. 1729, J. Azéma, M. Ledoux, M. Emery, and M. Yor, eds., Springer, Berlin, 2000, pp. 1–145.

[6] P. Del Moral and L. Miclo, *Genealogies and increasing propagations of chaos for Feynman-Kac and genetic models*, Ann. Appl. Probab., 11 (2001), pp. 1166–1198.

[7] P. Del Moral and E. Rio, *Concentration inequalities for mean field particle models*, Ann. Appl. Probab., 21 (2011), pp. 1017–1052.

[8] P. Del Moral, A. Doucet, and S. S. Singh, *A backward particle interpretation of Feynman-Kac formulae*, M2AN Math. Model. Numer. Anal., 44 (2010), pp. 947–975.

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

[10] R. Douc, E. Moulines, and D. S. Stoffer, *Nonlinear Time Series*, CRC Press, Boca Raton, FL, 2014.

[11] A. Doucet and V. B. Tadić, *Parameter estimation in general state-space models using particle methods*, Ann. Inst. Statist. Math., 55 (2003), pp. 409–422.

[12] A. Doucet, N. de Freitas, and N. Gordon, eds., *Sequential Monte Carlo Methods in Practice*, Springer, New York, 2001.

[13] N. J. Gordon, D. Salmond, and A. F. M. Smith, *Novel approach to nonlinear/non-Gaussian Bayesian state estimation*, IEE Proc. Radar Sig. Proces., 140 (1993), pp. 107–113.

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

[15] A. Jasra, *On the behaviour of the backward interpretation of Feynman-Kac formulae under verifiable conditions*, J. Appl. Probab., to appear.

[16] N. Kantas, A. Doucet, S. S. Singh, J. M. Maciejowski and N. Chopin, *On particle methods for parameter estimation in state-space models*, Stat. Sci., to appear.

[17] F. Le Gland and M. Mevel, *Recursive estimation in hidden Markov models*, in Proceedings of the 36th IEEE Conference on Decision and Control, 1997, pp. 3468–3473.

[18] J. Olsson, R. Douc, O. Cappé, and É. Moulines, *Sequential Monte Carlo smoothing with application to parameter estimation in nonlinear state space models*, Bernoulli, 14 (2008), pp. 155–179.

[19] N. Oudjane and S. Rubenthaler, *Stability and uniform particle approximation of nonlinear filters in case of non ergodic signals*, Stoch. Anal. Appl., 23 (2005), pp. 421–448.

[20] G. Pflug, *Optimization of Stochastic Models: The Interface Between Simulation and Optimization*, Kluwer, Boston, 1996.

[21] M. K. Pitt and N. Shephard, *Filtering via simulation: auxiliary particle filter*, J. Amer. Statist. Assoc., 94 (1999), pp. 590–599.

[22] G. Poyiadjis, A. Doucet, and S. S. Singh, *Particle methods for optimal filter derivative: Application to parameter estimation*, in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 5, Philadelphia, 2005, pp. 925–928.

[23] G. Poyiadjis, A. Doucet, and S. S. Singh, *Particle approximations of the score and observed information matrix in state space models with application to parameter estimation*, Biometrika, 98 (2011), pp. 65–80.

[24] D. M. Titterington, *Recursive parameter estimation using incomplete data*, J. Roy. Statist. Soc. Ser. B, 46 (1984), pp. 257–267.

[25] A. W. van der Vaart, *Asymptotic Statistics*, Cambridge University Press, Cambridge, MA, 1998.

[26] N. Whiteley, *Stability properties of some particle filters*, Ann. Appl. Probab., 23 (2013), pp. 2500–2537.