The Annals of Applied Probability
2016, Vol. 26, No. 2, 1111–1146
DOI: 10.1214/15-AAP1113

ON THE CONVERGENCE OF ADAPTIVE SEQUENTIAL MONTE CARLO METHODS

BY ALEXANDROS BESKOS[*,1,2], AJAY JASRA[†,3], NIKOLAS KANTAS[‡] AND ALEXANDRE THIERY[†,1]

University College London[*], National University of Singapore[†] and Imperial College London[‡]

In several implementations of Sequential Monte Carlo (SMC) methods it is natural and important, in terms of algorithmic efficiency, to exploit the information of the history of the samples to optimally tune their subsequent propagations. In this article we provide a carefully formulated asymptotic theory for a class of such *adaptive* SMC methods. The theoretical framework developed here will cover, under assumptions, several commonly used SMC algorithms [Chopin, *Biometrika* **89** (2002) 539–551; Jasra et al., *Scand. J. Stat.* **38** (2011) 1–22; Schäfer and Chopin, *Stat. Comput.* **23** (2013) 163–184]. There are only limited results about the theoretical underpinning of such adaptive methods: we will bridge this gap by providing a weak law of large numbers (WLLN) and a central limit theorem (CLT) for some of these algorithms. The latter seems to be the first result of its kind in the literature and provides a formal justification of algorithms used in many real data contexts [Jasra et al. (2011); Schäfer and Chopin (2013)]. We establish that for a general class of adaptive SMC algorithms [Chopin (2002)], the asymptotic variance of the estimators from the adaptive SMC method is *identical* to a "limiting" SMC algorithm which uses ideal proposal kernels. Our results are supported by application on a complex high-dimensional posterior distribution associated with the Navier–Stokes model, where adapting high-dimensional parameters of the proposal kernels is critical for the efficiency of the algorithm.

1. Introduction. Sequential Monte Carlo (SMC) methods are amongst the most widely used computational techniques in statistics, engineering, physics, finance and many other disciplines; see [18] for a recent overview. They are designed to approximate a sequence $\{\eta_n\}_{n\geq 0}$ of probability distributions of increasing dimension. The method uses $N \geq 1$ samples (or particles) that are generated in parallel, and are propagated via importance sampling and resampling methods. Several convergence results, as N grows, have been proved (see, e.g., [6, 11, 12, 16])

Received February 2014; revised January 2015.
[1]Supported by a Singapore MOE grant.
[2]Supported by the Leverhulme Trust Prize.
[3]Supported by Singapore Ministry of Education Academic Research Fund Tier 1 Grant (R-155-000-156-112) and is also affiliated with the RMI and CQF at NUS.
MSC2010 subject classifications. Primary 82C80, 60K35; secondary 60F99, 62F15.
Key words and phrases. Adaptive sequential Monte Carlo, CLT, MCMC.

along with the stability in time of the error of the algorithm [17] in the context of filtering. SMC methods have also recently been proven to be stable in certain high-dimensional contexts [2]. Current state of the art CLTs in SMC algorithms can be found in [6, 8, 11, 12, 16]; in particular, Chan and Lai [6] provide a method for online, consistent estimation of the asymptotic variance.

In this article, we are concerned with *adaptive* SMC methods; in an effort to improve algorithmic efficiency, the weights and/or Markov proposal kernels can depend upon the history of the simulated process. Such procedures appear in a wealth of articles including [7, 14, 15, 21, 23] and have important applications in, for example, econometrics, population genetics and data assimilation. The underlying idea of these algorithms is that, using the particle approximation η_n^N of the distribution η_n, one can exploit the induced information to build effective proposals and/or to determine the next probability distribution in the sequence. This is often achieved by using expectations of some relevant summary statistics with respect to the current SMC approximation η_n^N. In other cases, one can use the particles to determine the next distribution in an artificial sequence of densities. Such approaches are expected to lead to algorithms that are more efficient than their "nonadaptive" counter-parts. Critically, such ideas also deliver more automated algorithms by reducing the number of user-specified tuning parameters.

While the literature on adaptive MCMC methods is by now well developed, (e.g., [1]), and sufficient conditions for an adaptive MCMC algorithm to be ergodic are well understood, the analysis of adaptive SMC algorithms is still in its infancy. To the best of our knowledge, a theoretical study of the consistency and fluctuation properties of adaptive SMC algorithms is lacking in the current literature. This article aims at filling this critical gap in the theory of SMC methods. Some preliminary results can be found, under exceptionally strong conditions, in [10, 21]. Proof sketches are given in [15], some more realistic but limited analysis can be found in [20]; it should be noted that, contrary to most of the results presented in this article, the analysis of [20] is nonasymptotic. Some more recent work of [5] focuses upon consistency and CLTs for a different class of adaptive SMC algorithms, and the work there is complementary to the results proved in this article.

1.1. *Contributions.* In this article we consider a sequence of target distributions $\{\eta_n\}_{n\geq 0}$ defined on a corresponding sequence of measurable spaces $(E_n, \mathscr{E}_n)_{n\geq 0}$. We write $\eta_n^N = (1/N)\sum_{i=1}^N \delta_{x_n^i}$ for the particle approximation of η_n, with δ_{x_n} the Dirac measure at $x_n \in E_n$ and $\{x_n^i\}_{i=1}^N \in E_n^N$ the collection of particles at time $n \geq 0$.

In Section 2, for each $n \geq 1$ we consider a parametric family, indexed by a parameter $\xi \in \mathbb{R}^d$, of Markov kernels $M_{n,\xi} : E_{n-1} \times \mathscr{E}_n \to \mathbb{R}_+$ and potential functions $G_{n-1,\xi} : E_{n-1} \to \mathbb{R}_+$. To construct the particle approximation η_n^N, the adaptive SMC algorithm exploits summary statistics $\xi_n : E_{n-1} \to \mathbb{R}^d$ by

reweighing and propagating the particle approximation η_{n-1}^N through the potential $G_{n-1,\eta_{n-1}^N(\xi_n)}$ and the Markov kernel $M_{n,\eta_{n-1}^N(\xi_n)}$. This is a substitute for the *limiting* algorithm which employs the Markov kernel $M_n \equiv M_{n,\eta_{n-1}(\xi_n)}$ and weight function $G_{n-1} \equiv G_{n-1,\eta_{n-1}(\xi_n)}$; in general, the exact value of $\eta_{n-1}(\xi_n)$ is not available, and the limiting algorithm cannot be implemented. This set-up is relevant, for example, in the context of sequential Bayesian parameter inference [7, 22] when $\{\eta_n\}_{n\geq 0}$ is a sequence of posterior distributions on a random variable(s) conditional upon an increasing amount of data. In this context, the parametric family of Markov kernels $M_{n,\xi}$ is user-specified and its role is to efficiently move the particles within the state space; in many situations $M_{n,\xi}$ is chosen to be reversible with respect to the distribution η_n for any value of ξ. A random walk Metropolis kernel that uses the estimated covariance structure of η_n for scaling its jump proposals is a popular choice; in this case, and when working in $\mathbb{R}^p$, the empirical covariance matrix can be constructed through the $d = p(p+1)/2$ dimensional summary statistics $\xi = (x_1, \ldots, x_p, x_1x_1, x_1x_2, \ldots, x_px_p)$. The case when the weight function $G_{n-1,\xi} : E_{n-1} \to \mathbb{R}_+$ also depends on a parameter $\xi \in \mathbb{R}^d$ is relevant, for instance, to particle filters [18], as described in Section 2.3.2.

Section 3 investigates situations when an additional layer of adaptivity appears through a tempering procedure. Standard MCMC methods can be inefficient for directly exploring complex probability distributions involving high-dimensional state spaces, multi-modality, greatly varying scales or a combination thereof. It is a standard approach to introduce a bridging sequence of distributions $\{\eta_n\}_{n=0}^{n=n_\star}$ between a distribution η_0 that is typically easy to sample from and the distribution of interest $\eta_{n_\star} \equiv \pi$. In accordance with the simulated tempering literature, the probability distribution of interest is written as $\pi(dx) = Z^{-1}\exp(-V(x))m(dx)$ for a potential V, dominating measure $m(dx)$ and normalization constant Z; the bridging sequence of distributions is constructed by introducing a ladder of temperature parameters $\beta_0 \leq \beta_1 \leq \cdots \leq \beta_{n_\star} = 1$ and setting $\eta_n(dx) = Z(\beta_n)^{-1}\exp(-\beta_n V(x))m(dx)$ for a normalization constant $Z(\beta_n)$. The choice of the bridging sequence of distributions is an important and complex problem; see, for example, [19]. To avoid the task of having to pre-specify a potentially large number of temperature parameters, an adaptive SMC method can compute them "on the fly" [21, 23], thus obtaining a random increasing sequence of tempering parameters $\{\beta_n^N\}_{n\geq 0}$. In this article, we adopt the following strategy: assuming the current particle approximation

$$\eta_{\beta_{n-1}^N}^N = (1/N)\sum_{i=1}^N \delta_{x_{n-1}^i},$$

the particles are assigned weights $\omega_{n-1}^i \propto \exp\{-(\beta_n^N - \beta_{n-1}^N)V(x_{n-1}^i)\}$ to represent the next distribution in the sequence; the tempering parameter β_n^N is chosen so that the effective sample size (ESS) of the weighted particle system

$\{(\omega_{n-1}^i, x_{n-1}^i)\}_{i=1}^N$ is larger than a pre-specified threshold. This strategy can be efficiently implemented using a bisection method [21]. Once the tempering parameter β_n^N has been determined, the weighted particle system $\{(\omega_{n-1}^i, x_{n-1}^i)\}_{i=1}^N$ is resampled to generate an equally weighted family $\{x_n^i\}_{i=1}^N$ such that $\eta_{\beta_n^N}^N = (1/N)\sum_{i=1}^N \delta_{x_n^i}$ approximates the distribution $\eta_{\beta_n^N}$; a Markov kernel $M_{n,\eta_{n-1}^N(\xi_n),\beta_n^N}$ is then used to explore $\eta_{\beta_n^N}$; as before, the algorithm makes use of summary statistics $\xi_n : E \to \mathbb{R}$ in order to automatically tune these Markov transition kernels.

We investigate in this article the consistency and fluctuation properties of these classes of adaptive SMC methods. We establish a WLLN for both algorithms; for instance, this implies that one can consistently approximate normalizing constants. We prove that central limit theorems hold at the usual $N^{-1/2}$ Monte Carlo rate and give explicit recursion equations for the asymptotic variances. These results establish that, in several situations of practical importance, the asymptotic variance of the adaptive SMC algorithm is identical to the asymptotic variance of the limiting algorithm. In these situations, this implies that the fluctuation analysis of the limiting algorithm can be used to describe the asymptotic properties of the adaptive algorithm. Related results have been obtained in [5] for adaptive multilevel splitting algorithms. The analysis of the SMC algorithm with adaptive scaling and tempering is more involved than the situation where only the scaling of the Markov kernel is adapted; in this case the fluctuations of the tempering parameters $\{\beta_n^N\}_{n\geq 0}$ and the correlation between these parameters and the particle approximations $\eta_{\beta_n^N}^N$ have to be simultaneously taken into account. For this reason, the analyses of the two algorithms are presented in two separated sections. In the last section we present two numerical applications: the first is a toy example and the second is a complex high-dimensional posterior distribution associated with the Navier–Stokes model (as in, e.g., [22]); in this setting, adapting the proposal kernels over hundreds of different directions is critical for the efficiency of the algorithm. The assumptions for the theoretical results of the paper to hold are not satisfied in the second example; these experiments thus provide some evidence that our theory could be relevant in more general scenarios.

2. Adaptive SMC via summary statistics. We begin with some notation and a description of the algorithm for the case where the Markov kernels and weight functions are adapted to summary statistics; we then establish asymptotic results. The description of the adaptive tempering procedure is postponed to Section 3.

2.1. *Notation and definitions.* Let $(E_n, \mathscr{E}_n)_{n\geq 0}$ be a sequence of measurable spaces endowed with a countably generated σ-field $\mathscr{E}_n$. The notation $\mathcal{B}_b(E_n)$ denotes the class of bounded $\mathscr{E}_n/\mathbb{B}(\mathbb{R})$-measurable functions where $\mathbb{B}(\mathbb{R})$ is the Borel σ-algebra on $\mathbb{R}$. The supremum norm is written as $\|f\|_\infty = \sup_{x\in E_n} |f(x)|$, and $\mathcal{P}(E_n)$ is the set of probability measures on $(E_n, \mathscr{E}_n)$. We will consider nonnegative operators $K : E_{n-1} \times \mathscr{E}_n \to \mathbb{R}_+$ such that for each $x \in E_{n-1}$ the mapping

$A \mapsto K(x, A)$ is a finite nonnegative measure on $\mathscr{E}_n$, and for each $A \in \mathscr{E}_n$ the function $x \mapsto K(x, A)$ is $\mathscr{E}_{n-1}/\mathbb{B}(\mathbb{R})$-measurable; the kernel K is Markovian if $K(x, dy)$ is a probability measure for every $x \in E_{n-1}$. For a finite measure μ on $(E_{n-1}, \mathscr{E}_{n-1})$ and Borel test function $f \in \mathcal{B}_b(E_n)$, we define

$$\mu K : A \mapsto \int K(x, A)\mu(dx); \qquad Kf : x \mapsto \int f(y)K(x, dy).$$

We use the notation $o_{\mathrm{P}}(1)$ to indicate a quantity that converges to zero in probability. Finally, we will use the following notion of continuity in several places in this article.

DEFINITION 2.1. Let $\mathcal{X}$ be a set and $(\mathcal{Y}, d_{\mathcal{Y}})$, $(\mathcal{Z}, d_{\mathcal{Z}})$ two metric spaces. A function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ is continuous at $y_0 \in \mathcal{Y}$ uniformly on $\mathcal{X}$ if

$$\lim_{\delta \to 0^+} \sup\bigl\{d_{\mathcal{Z}}\bigl(f(x, y), f(x, y_0)\bigr) : x \in \mathcal{X}, d_{\mathcal{Y}}(y, y_0) < \delta\bigr\} = 0. \tag{2.1}$$

In words, a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ is continuous at $y_0 \in \mathcal{Y}$ uniformly on $\mathcal{X}$ if the quantities $f(x, y)$ and $f(x, y_0)$ can be made arbitrarily close, uniformly over $x \in \mathcal{X}$, by choosing y close enough to y_0. Finally, for a scalar function $(x, \mu) \mapsto f_\mu(x)$, we often use the shorthand notation $\partial_{\overline{\mu}} f_\mu(x)$ to designate the differential of f with respect to μ and evaluated at $(\overline{\mu}, x)$, that is, $\partial_\mu f_\mu(x)|_{\mu=\overline{\mu}}$. The Kronecker product $u \otimes v$ of two vectors $u, v \in \mathbb{R}^d$ is the matrix $u \cdot v^\top \in \mathbb{R}^{d \times d}$.

2.2. *SMC algorithm.* For each $n \geq 1$, we consider a family of Markov operators $M_{n,\xi} : E_{n-1} \times \mathscr{E}_n \to \mathbb{R}_+$ and potential functions $G_{n-1,\xi} : E_{n-1} \to \mathbb{R}_+$ parametrized by a vector $\xi \in \mathbb{R}^d$. The adaptive SMC algorithm to be described exploits summary statistics $\xi_n : E_{n-1} \to \mathbb{R}^d$ and aims at approximating a sequence of probability distributions $\{\eta_n\}_{n \geq 0}$ on $(E_n, \mathscr{E}_n)_{n \geq 0}$ defined via their operation on $\varphi_n \in \mathcal{B}_b(E_n)$,

$$\eta_n(\varphi_n) := \gamma_n(\varphi_n)/\gamma_n(1). \tag{2.2}$$

The unnormalized measure γ_n on $(E_n, \mathscr{E}_n)$ is

$$\gamma_n(\varphi_n) := \mathrm{E}\left[\left\{\prod_{p=0}^{n-1} G_p(X_p)\right\}\varphi_n(X_n)\right], \tag{2.3}$$

where $\{X_n\}_{n \geq 0}$ is a nonhomogeneous Markov chain with initial distribution $X_0 \sim \eta_0 \equiv \gamma_0$ and transition $\mathrm{P}(X_n \in A | X_{n-1} = x) = M_n(x, A)$. We have used the simplified notation

$$G_{p-1} \equiv G_{p-1,\overline{\xi_p}} \quad \text{and} \quad M_p \equiv M_{p,\overline{\xi_p}} \qquad \text{with } \overline{\xi_p} \equiv \eta_{p-1}(\xi_p).$$

In the sequel, the notation $\mathrm{Range}(\xi_n) \subset \mathbb{R}^d$ denotes an open convex set such that $\xi_n(x) \in \mathrm{Range}(\xi_n)$ for any $x \in E_{n-1}$, and we only require $G_{n-1,\xi}$ and $M_{n,\xi}$ to be

well-defined for $\xi \in \mathrm{Range}(\xi_n)$. In practice, the expectations $\overline{\xi}_n$ of the summary statistics are not analytically tractable, and it is thus impossible to sample from the Markov chain $\{X_n\}_{n\geq 0}$ or evaluate the weight functions $\{G_n\}_{n\geq 0}$. Nevertheless, for the purpose of analysis, we introduce the following idealized algorithm, referred to as the *limiting* SMC algorithm in the sequel, that propagates a set of $N \geq 1$ particles over n iterations; that is, $(x_0^{1:N}, x_1^{1:N}, \ldots, x_n^{1:N})$ has joint probability distribution given by

$$\prod_{i=1}^{N} \eta_0(dx_0^i) \prod_{p=1}^{n} \prod_{i=1}^{N} \Phi_p(\eta_{p-1}^N)(dx_p^i). \tag{2.4}$$

The empirical distribution $\eta_n^N = (1/N)\sum_{i=1}^N \delta_{x_n^i}$ is an approximation of (2.2), and the operator $\Phi_n : \mathcal{P}(E_{n-1}) \to \mathcal{P}(E_n)$ is defined as

$$\Phi_n(\mu)(dy) = \frac{\mu(G_{n-1}M_n)(dy)}{\mu(G_{n-1})}.$$

Expression (2.4) is a mathematically concise way to describe a standard particle method that begins by sampling N i.i.d. particles from the distribution η_0 and, given particles $\{x_{n-1}^i\}_{i=1}^N$, performs multinomial resampling according to the (unnormalized) weights $G_{n-1}(x_{n-1}^i)$ before propagating the particles via the Markov kernel M_n. The particle genealogy $(x_0^{1:N}, x_1^{1:N}, \ldots, x_n^{1:N})$ simulated in practice has a joint probability law given by

$$\prod_{i=1}^{N} \eta_0(dx_0^i) \prod_{p=1}^{n} \prod_{i=1}^{N} \Phi_{p,N}(\eta_{p-1}^N)(dx_p^i), \tag{2.5}$$

for an approximate operator $\Phi_{n,N}$, which is defined as follows: $\Phi_{n,N}(\mu)(dy) = \mu(G_{n-1,N}M_{n,N})(dy)/\mu(G_{n-1,N})$ with

$$G_{n-1,N} \equiv G_{n-1,\eta_{n-1}^N(\xi_n)} \quad \text{and} \quad M_{n,N} \equiv M_{n,\eta_{n-1}^N(\xi_n)}.$$

Throughout this article we assume that the potentials are strictly positive, that is, $G_{n,\xi}(x) > 0$ for all $x \in E_n$ and $\xi \in \mathrm{Range}(\xi_n)$, so that there is no possibility for the algorithm to collapse. The particle approximation of the unnormalized distribution (2.3) is defined as

$$\gamma_n^N(\varphi_n) = \left\{\prod_{p=0}^{n-1} \eta_p^N(G_{p,N})\right\} \eta_n^N(\varphi_n). \tag{2.6}$$

Contrarily to the nonadaptive case [11], the quantity $\gamma_n^N(1) = \prod_{p=0}^{n-1} \eta_p^N(G_{p,N})$ is *not*, in general, an unbiased estimate of the normalization constant $\gamma_n(1) = \prod_{p=0}^{n-1} \eta_p(G_p)$. It is useful to introduce the nonnegative operator

$$Q_{n,N}(x, dy) = G_{n-1,N}(x) M_{n,N}(x, dy) \tag{2.7}$$

and the limiting version $Q_n(x, dy) = G_{n-1}(x)M_n(x, dy)$. To emphasize the dependency on the parameter $\xi \in \mathbb{R}^d$, we will sometimes write $Q_{n,\xi}(x, dy) := G_{n-1,\xi}(x)M_{n,\xi}(x, dy)$ when no ambiguity is possible. With these definitions, the following identities hold:

$$\eta_n(\varphi_n) = \Phi_n(\eta_{n-1})(\varphi_n) = \frac{\eta_{n-1}(Q_n\varphi_n)}{\eta_{n-1}(G_{n-1})}; \qquad \gamma_n(\varphi_n) = \gamma_{n-1}(Q_n\varphi_n). \tag{2.8}$$

Similar formulas are available for the N-particle approximations. Let $\mathscr{F}_n^N$ denote the filtration generated by the particle system up-to (and including) time n; we have

$$\mathrm{E}[\eta_n^N(\varphi_n)|\mathscr{F}_{n-1}^N] = \Phi_{n,N}(\eta_{n-1}^N)(\varphi_n),$$
$$\mathrm{E}[\gamma_n^N(\varphi_n)|\mathscr{F}_{n-1}^N] = \gamma_{n-1}^N(Q_{n,N}\varphi_n).$$

In the sequel, we use the expressions $\mathrm{E}_n[\cdot]$ and $\mathrm{Var}_n[\cdot]$ to denote the conditional expectation $\mathrm{E}[\cdot|\mathscr{F}_n^N]$ and conditional variance $\mathrm{Var}(\cdot|\mathscr{F}_n^N)$, respectively. We use the standard notation $\langle u, v\rangle$ to denote the Euclidean scalar product between two vectors u and v.

2.3. *Motivating examples.* We motivate the above structure for $\{\eta_n\}_{n\geq 0}$ via two examples.

2.3.1. *Sequential Bayesian parameter inference.* We consider Bayesian inference for a parameter $x \in E$, associated to observations $y_i \in \mathcal{Y}$ and prior measure $\eta_0(dx)$. Assuming that all the distributions have a density with respect to a relevant dominating measure, the posterior density η_n given $y_{1:n} \in \mathcal{Y}^n$ is $\eta_n(x) \propto \eta_0(x)\,\mathrm{p}(y_{1:n}|x)$. The approach in [7] fits into the framework described in Section 2.2 with state spaces $E_n = E$ and potential functions $G_n(x) = \mathrm{p}(y_{n+1}|y_{1:n}, x)$. For an MCMC kernel $M_n \equiv M_{n,\eta_{n-1}(\xi_n)}$ with invariant measure η_n, the posterior distribution η_n is given by $\eta_n(\varphi_n) = \gamma_n(\varphi_n)/\gamma_n(1)$ where the unnormalized measure γ_n is defined as in (2.3). As described in Section 1.1, a popular choice for $M_{n,\eta_{n-1}(\xi_n)}$ corresponds to a random walk Metropolis kernel that is reversible with respect to η_n and a jump covariance structure matching the one of η_{n-1}.

While such an example is quite simple, it is indicative of more complex applications in the literature. Kantas, Beskos and Jasra [22] consider a state-space with dimension $p \approx 10^4$ and dimension of adapted statistic of about $d \approx 5 \cdot 10^2$. In such a setting, pre-specifying the covariance structure of the random walk Metropolis proposals is impractical; the adaptive SMC strategy of Section 2.2 provides a principled framework for automatically setting this covariance structure; see also Section 4.2.

2.3.2. *Filtering*. This section illustrates the case of an adaptive weight function. Consider a state-space model with observations $y_{1:n} \in \mathcal{Y}^n$, unobserved Markov chain $U_{0:n} \in \mathcal{U}^{n+1}$ and joint density with respect to a dominating measure $\lambda_{\mathcal{Y}}^{\otimes n} \otimes \lambda_{\mathcal{U}}^{\otimes n+1}$ given by

$$\eta_0(u_0) \prod_{p=1}^{n} g_p(u_p, y_p) f_p(u_{p-1}, u_p).$$

The density η_0 describes the prior distribution for the initial state of the unobserved Markov chain, $g_p(u_p, y_p)$, the conditional observation density at time p and $f_p(u_{p-1}, u_p)$, the dynamics of the unobserved Markov process. A standard particle filter with proposal at time p distributed according to the Markov kernel $m_p(u_{p-1}, u_p)\lambda_{\mathcal{U}}(du_p)$ uses importance weights of the form

$$G_p(x_p) = \frac{g_p(u_p, y_p) f_p(u_{p-1}, u_p)}{m_p(u_{p-1}, u_p)},$$

where here $x_p \equiv (x_p^{(1)}, x_p^{(2)}) \equiv (u_{p-1}, u_p)$. The process $\{X_p\}_{p=1}^n$ is Markovian with transition

$$M_p(x_{p-1}, dx_p) = \delta_{x_{p-1}^{(2)}}(dx_p^{(1)}) m_p(x_{p-1}^{(2)}, x_p^{(2)}) \lambda_{\mathcal{U}}(dx_p^{(2)}).$$

In this setting the marginal of the distribution η_p on the $x_p^{(2)}$-component is the filtering distribution. In practice, the choice of the proposal kernel m_n is critical to the efficiency of the algorithm, and one may want to exploit the information contained in the distribution η_{n-1} in order to build efficient proposal kernels; approximating the filter mean is a standard strategy. In these cases, both the Markov kernel M_n and the weight function G_{n-1} depend upon the distribution η_{n-1}; this is covered by the framework of Section 2. See [18] and the references therein for ideas associated to such approaches.

2.4. *Assumptions*. Our results make use of Assumptions 1 and 2 below [written (A1) and (A2) below as shorthand]. Recall that $\text{Range}(\xi_n) \subset \mathbb{R}^d$ denotes an open convex set that contains the range of the statistic $\xi_n : E_{n-1} \to \mathbb{R}^d$.

ASSUMPTION 1. For each $n \geq 1$ the statistic $\xi_n : E_{n-1} \to \mathbb{R}^d$ is bounded, and for any test function $\varphi \in \mathcal{B}_b(E_n)$ the functions $(x, \xi) \mapsto G_{n-1,\xi}(x)$ and $(x, \xi) \mapsto Q_{n,\xi}\varphi(x)$ are bounded and continuous at $\overline{\xi}_n = \eta_{n-1}(\xi_n)$ uniformly over $x \in E_{n-1}$.

ASSUMPTION 2. For each $n \geq 1$ and $\varphi \in \mathcal{B}_b(E_n)$, the function $(x, \xi) \mapsto \partial_\xi Q_{n,\xi}\varphi(x)$ is well defined on $E_{n-1} \times \text{Range}(\xi_n)$ and is bounded and continuous at $\overline{\xi}_n$ uniformly over $x \in E_{n-1}$.

Assumptions 1 and 2 are reasonably weak in comparison to some assumptions frequently used in the SMC literature, such as those in [11], but are certainly not the weakest adopted for the WLLN and CLTs; see, for example, [8]. The continuity assumptions in Assumption 2 are associated to the use of a first-order Taylor expansion. We have defined $\mathrm{Range}(\xi_p)$ as a convex set because we need to compute integrals along segments between elements of $\mathrm{Range}(\xi_p)$. We expect that these assumptions can be relaxed to accommodate unbounded test functions at the cost of increased length and complexity of the proofs.

2.5. *Weak law of large numbers.* In this section we establish a WLLN. We first state a slightly stronger result that will be repeatedly used in the fluctuation analysis presented in Section 2.6.

THEOREM 2.1 (WLLN for SMC with adaptive scaling). *Assume* (A1). *Let* V *be a Polish space and* $\{V_N\}_{N\geq 0}$ *a sequence of* V*-valued random variables that converges in probability to* $\mathsf{v}\in\mathsf{V}$. *Let* $n\geq 0$ *and* $\varphi_n\in\mathcal{B}_b(E_n\times\mathsf{V})$ *be a bounded function, continuous at* $\mathsf{v}\in\mathsf{V}$ *uniformly on* E_n. *We have*

$$\eta_n^N\big[\varphi_n(\cdot,V_N)\big]\xrightarrow{\mathrm{P}}\eta_n\big[\varphi_n(\cdot,\mathsf{v})\big].$$

COROLLARY 1. *Assume* (A1). *Let* $n\geq 0$ *and* $\varphi_n\in\mathcal{B}_b(E_n)$. *We have*

$$\eta_n^N(\varphi_n)\xrightarrow{\mathrm{P}}\eta_n(\varphi_n).$$

PROOF OF THEOREM 2.1. The proof is by induction on n. The initial case $n=0$ is a direct consequence of the WLLN for i.i.d. random variables and Definition 2.1. For notational convenience, in the rest of the proof we write $\overline{\varphi}_n(\cdot)$ instead of $\varphi_n(\cdot,\mathsf{v})$. We assume the result at rank $(n-1)$ and proceed to the induction step. Since $V_N\xrightarrow{\mathrm{P}}\mathsf{v}\in\mathsf{V}$, Definition 2.1 implies that it suffices to prove that $[\eta_n^N-\eta_n](\overline{\varphi}_n)\xrightarrow{\mathrm{P}}0$. We use the decomposition $[\eta_n^N-\eta_n](\overline{\varphi}_n)=A(N)+B(N)$ with

$$\begin{aligned}A(N)&=\big[\eta_n^N-\eta_n\big](\overline{\varphi}_n)-\mathrm{E}\big[\big[\eta_n^N-\eta_n\big](\overline{\varphi}_n)|\mathscr{F}_{n-1}^N\big]\\&=\big[\eta_n^N-\Phi_{n,N}\big(\eta_{n-1}^N\big)\big](\overline{\varphi}_n),\\B(N)&=\mathrm{E}\big[\big[\eta_n^N-\eta_n\big](\overline{\varphi}_n)|\mathscr{F}_{n-1}^N\big]=\big[\Phi_{n,N}\big(\eta_{n-1}^N\big)-\eta_n\big](\overline{\varphi}_n).\end{aligned}$$

The proof consists of showing that each of these terms converges to zero in probability.

- *Proof that* $A(N)\xrightarrow{\mathrm{P}}0$.
  Since the expected value of $A(N)$ is zero, it suffices to prove that its moment of order two also converges to zero as N goes to infinity. This follows as

$$\mathrm{E}_{n-1}\big[A(N)^2\big]=\frac{1}{N}\mathrm{E}_{n-1}\big[\big(\overline{\varphi}\big(x_n^1\big)-\mathrm{E}_{n-1}\big[\overline{\varphi}\big(x_n^1\big)\big]\big)^2\big]\leq\frac{4\|\overline{\varphi}\|_\infty^2}{N}.$$

- *Proof that* $B(N) \xrightarrow{\mathrm{P}} 0$.
The definition of $\Phi_{n,N}(\eta_{n-1}^N)$ yields $B(N) = B_1(N) + B_2(N) + B_3(N)$ with

$$\begin{cases} B_1(N) = \eta_{n-1}^N\{[Q_{n,N} - Q_n](\overline{\varphi}_n)\}/\eta_{n-1}^N(G_{n-1,N}), \\ B_2(N) = [\eta_{n-1}^N - \eta_{n-1}](Q_n\overline{\varphi}_n)/\eta_{n-1}^N(G_{n-1,N}), \\ B_3(N) = \eta_{n-1}[Q_n\overline{\varphi}_n] \times \{1/\eta_{n-1}^N(G_{n-1,N}) - 1/\eta_{n-1}(G_{n-1})\}. \end{cases}$$

We prove that $B_i(N) \to 0$, in probability, for $i = 1, 2, 3$. (All limits below are in probability.) From the induction hypothesis, we directly get that $\eta_{n-1}^N(\xi_n) \to \overline{\xi}_n$. By (A1), the bounded function $(x, \xi) \mapsto G_{n-1,\xi}(x)$ is continuous at $\xi = \overline{\xi}_n$ uniformly on E_{n-1}; therefore, the induction hypothesis applies, and we have that $\eta_{n-1}^N(G_{n-1,N}) \to \eta_{n-1}(G_{n-1})$. Similarly, since $Q_n(\overline{\varphi}_n) \in \mathcal{B}_b(E_{n-1})$ is bounded by the boundedness of $\overline{\varphi}_n$, we get $\eta_{n-1}^N[Q_n(\overline{\varphi}_n)] \to \eta_{n-1}[Q_n(\overline{\varphi}_n)]$. By Slutsky's lemma, both $B_2(N)$ and $B_3(N)$ converge to zero in probability. By (A1) the bounded function $(x, \xi) \mapsto Q_{n,\xi}\overline{\varphi}_n(x)$ is continuous at $\xi = \overline{\xi}_n$ uniformly on E_{n-1} so that the induction yields that $\eta_{n-1}^N[Q_{n,N}(\overline{\varphi}_n)] \to \eta_{n-1}[Q_n(\overline{\varphi}_n)]$ and $\eta_{n-1}^N[Q_n(\overline{\varphi}_n)] \to \eta_{n-1}[Q_n(\overline{\varphi}_n)]$; it follows that

$$\eta_{n-1}^N\{[Q_{n,N} - Q_n](\overline{\varphi}_n)\} \xrightarrow{\mathrm{P}} 0.$$

Consequently, the quantity $B_1(N)$ also converges to zero in probability. □

As a corollary, one can establish a similar consistency result for the sequence of unnormalized approximations $\gamma_n^N(\varphi_n)$ defined in equation (2.6).

COROLLARY 2. *Assume* (A1). *For* $n \geq 0$ *and test function* $\varphi_n \in \mathcal{B}_b(E_n)$, *we have*

$$\gamma_n^N(\varphi_n) \xrightarrow{\mathrm{P}} \gamma_n(\varphi_n).$$

PROOF. Since $\gamma_n^N(\varphi_n) = \gamma_n^N(1)\eta_n^N(\varphi_n)$ and $\gamma_n(\varphi_n) = \gamma_n(1)\eta_n(\varphi_n)$, by Corollary 1 it suffices to prove that $\gamma_n^N(1) = \prod_{p=0}^{n-1} \eta_p^N(G_{p,N})$ converges in probability to $\gamma_n(1) = \prod_{p=0}^{n-1} \eta_p(G_p)$. Due to the regularity conditions for $(x, \xi) \mapsto G_{n,\xi}(x)$ in (A1), Theorem 2.1 applies so that $\eta_p^N(G_{p,N}) \to \eta_p(G_p)$, in probability, for any index $p \geq 0$. The conclusion follows. □

2.6. *Central limit theorems.* In this section, for $\varphi_n : E_n \to \mathbb{R}$, we carry out a fluctuation analysis of the particle approximations $\gamma_n^N(\varphi_n)$, $\eta_n^N(\varphi_n)$ around their limiting values. We establish that CLTs hold at the usual Monte Carlo rate and give explicit expressions for the asymptotic variances. More precisely, we show that

there exist asymptotic variance functionals $\mathbb{V}_n^\gamma : \mathcal{B}_b(E_n) \to \mathbb{R}^+$ and $\mathbb{V}_n^\eta : \mathcal{B}_b(E_n) \to \mathbb{R}^+$ such that for any test function $\varphi_n \in \mathcal{B}_b(E_n)$, we have

$$N^{1/2}[\gamma_n^N - \gamma_n](\varphi_n) \xrightarrow{\mathcal{D}} \mathrm{N}\{0, \mathbb{V}_n^\gamma(\varphi_n)\}, \tag{2.9}$$

$$N^{1/2}[\eta_n^N - \eta_n](\varphi_n) \xrightarrow{\mathcal{D}} \mathrm{N}\{0, \mathbb{V}_n^\eta(\varphi_n)\}. \tag{2.10}$$

The variance functionals are such that $\mathbb{V}_0^\gamma(\varphi) = \mathbb{V}_0^\eta(\varphi) = \mathrm{Var}_{\eta_0}(\varphi)$ and

$$\begin{cases} \mathbb{V}_n^\gamma(\varphi_n) = \gamma_n(1)^2 \times \mathrm{Var}_{\eta_n}(\varphi_n) + \mathbb{V}_{n-1}^\gamma(\mathscr{L}_n\varphi_n), \\ \mathbb{V}_n^\eta(\varphi_n) = \mathrm{Var}_{\eta_n}(\varphi_n) + \mathbb{V}_{n-1}^\eta\big(\mathscr{L}_n(\varphi_n - \eta_n(\varphi_n))\big)/\eta_{n-1}^2(G_{n-1}) \end{cases} \tag{2.11}$$

for $n \geq 1$; the linear operator $\mathscr{L}_n : \mathcal{B}_b(E_n) \to \mathcal{B}_b(E_{n-1})$ is given by

$$\mathscr{L}_n\varphi_n = \langle \eta_{n-1}[\partial_{\overline{\xi}_n} Q_{n,\xi}\varphi_n], \xi_n - \overline{\xi}_n \rangle + Q_n\varphi_n.$$

Indeed, replacing $\mathscr{L}_n$ by $\mathscr{L}_n^\infty \equiv Q_n$ in (2.11) yields the standard expression for the asymptotic variances of nonadaptive SMC methods [11].

THEOREM 2.2 (CLTs for SMC with adaptive scaling). *Assume* (A1)–(A2). *For $n \geq 0$ and $\varphi_n \in \mathcal{B}_b(E_n)$, equations* (2.9) *and* (2.10) *hold, and the asymptotic variance functionals $\mathbb{V}_n^\gamma$ and $\mathbb{V}_n^\eta$ satisfy* (2.11).

PROOF. We first establish equation (2.9); we proceed by induction and prove that equation (2.9) holds for any index $n \geq 0$ and any test function $\varphi_n \in \mathcal{B}_b(E_n)$. The case $n = 0$ follows from the usual CLT for i.i.d. random variables. To prove the induction step we use the decomposition $[\gamma_n^N - \gamma_n](\varphi_n) = P(N) + R(N)$ with

$$P(N) = \mathrm{E}\big[[\gamma_n^N - \gamma_n](\varphi_n) | \mathscr{F}_{n-1}^N\big], \qquad R(N) = [\gamma_n^N - \gamma_n](\varphi_n) - P(N).$$

It then suffices to prove that for any $t \in \mathbb{R}$ we have

$$\mathrm{E}\big[\exp\{iN^{1/2}tR(N)\} | \mathscr{F}_{n-1}^N\big] \xrightarrow{\mathrm{P}} \exp\left\{-\frac{t^2}{2}\gamma_n(1)^2\,\mathrm{Var}_{\eta_n}(\varphi_n)\right\} \tag{2.12}$$

and that the quantity $P(N)$ is such that

$$N^{1/2}P(N) = N^{1/2}[\gamma_{n-1}^N - \gamma_{n-1}](\mathscr{L}_n\varphi_n) + o_{\mathrm{P}}(1). \tag{2.13}$$

Equations (2.12), (2.13) and the induction hypothesis imply that the characteristic function of $N^{1/2}[P(N) + R(N)]$ converges point-wise to the characteristic function of a Gaussian variable with variance $\gamma_n(1)^2\,\mathrm{Var}_{\eta_n}(\varphi_n) + \mathbb{V}_{n-1}^\gamma(\mathscr{L}_n\varphi_n)$. The conclusion follows from Levy's characterization of convergence in distribution. We finish the proof of equation (2.9) by establishing these two results:

- *Proof of equation* (2.12). We have that

$$R(N) = \gamma_n^N(1)\left(\sum_{i=1}^N U_{N,i} - \mathrm{E}\big[U_{N,i} | \mathscr{F}_{n-1}^N\big]\right)\Big/N,$$

where the variables $\{U_{N,i}\}_{i=1}^N$ are, conditionally upon $\mathscr{F}_{n-1}^N$, independent and identically distributed as $\varphi_n(X_{N,n})$ with

$$\mathrm{P}(X_{N,n} \in dx | \mathscr{F}_{n-1}^N) = \sum_{i=1}^N \frac{G_{n-1,N}(x_{n-1}^i)}{\sum_{j=1}^N G_{n-1,N}(x_{n-1}^j)} M_{n,N}(x_{n-1}^i, dx). \tag{2.14}$$

By Theorem A.3 of [16] and since $\gamma_n^N(1) \xrightarrow{\mathrm{P}} \gamma_n(1)$, to prove equation (2.12) it suffices to verify that for any $\varepsilon > 0$, the following two conditions hold:

$$\begin{cases} \mathrm{Var}(U_{N,i} | \mathscr{F}_{n-1}^N) \xrightarrow{\mathrm{P}} \mathrm{Var}_{\eta_n}(\varphi_n), \\ \mathrm{E}[U_{N,i}^2 \mathbb{I}_{|U_{N,i}^2| > N\varepsilon} | \mathscr{F}_{n-1}^N] \xrightarrow{\mathrm{P}} 0. \end{cases} \tag{2.15}$$

Let us start by proving the first condition. Standard algebraic manipulations show that the conditional variance is

$$\mathrm{Var}(U_{N,i} | \mathscr{F}_{n-1}^N) = \frac{\eta_{n-1}^N[Q_{n,\eta_{n-1}^N(\xi_n)}\varphi_n^2]}{\eta_{n-1}^N[G_{n-1,\eta_{n-1}^N(\xi_n)}]} - \left\{ \frac{\eta_{n-1}^N[Q_{n,\eta_{n-1}^N(\xi_n)}\varphi_n]}{\eta_{n-1}^N[G_{n-1,\eta_{n-1}^N(\xi_n)}]} \right\}^2.$$

By (A1), the functions $(x,\xi) \mapsto G_{n-1,\xi}(x)$, $(x,\xi) \mapsto Q_{n,\xi}\varphi_n(x)$ and $(x,\xi) \mapsto Q_{n,\xi}\varphi_n^2(x)$ are bounded and continuous at $\overline{\xi}_n \equiv \eta_{n-1}(\xi_n)$ uniformly on E_{n-1}; it thus follows from Theorem 2.1 and Slutsky's Lemma that

$$\mathrm{Var}(U_{N,i} | \mathscr{F}_{n-1}^N) \xrightarrow{\mathrm{P}} \frac{\eta_{n-1}[Q_n\varphi_n^2]}{\eta_{n-1}(G_n)} - \left\{ \frac{\eta_{n-1}[Q_n\varphi_n]}{\eta_{n-1}(G_n)} \right\}^2 = \mathrm{Var}_{\eta_n}(\varphi_n).$$

This completes the proof of the first condition of (2.15). The second condition directly follows from the boundedness of the function φ_n.

- *Proof of equation* (2.13). We have

$$P(N) = \gamma_{n-1}^N(1) \times \eta_{n-1}^N[Q_{n,N} - Q_n](\varphi_n) + [\gamma_{n-1}^N - \gamma_{n-1}](Q_n\varphi_n). \tag{2.16}$$

The fundamental theorem of calculus yields that

$$\begin{aligned} &\eta_{n-1}^N[Q_{n,N} - Q_n](\varphi_n) \\ &\quad = \langle \eta_{n-1}^N[\omega(\cdot, \eta_{n-1}^N(\xi_n))], [\eta_{n-1}^N - \eta_{n-1}](\xi_n) \rangle, \end{aligned} \tag{2.17}$$

where the function ω is given by

$$\omega(x,v) = \int_0^1 \partial_\xi Q_{n,\xi}\varphi_n(x)\Big|_{\xi=\xi(\lambda)} d\lambda \qquad \text{with } \xi(\lambda) \equiv \overline{\xi}_n + \lambda\{v - \overline{\xi}_n\}.$$

Under (A2), the function $(x,v) \mapsto \omega(x,v)$ is bounded and continuous at $v = \overline{\xi}_n$, uniformly over $x \in E_{n-1}$; by Theorem 2.1, it then follows that the quantity $\eta_{n-1}^N[\omega(\cdot, \eta_{n-1}^N(\xi_n))]$ converges in probability to $\eta_{n-1}[\partial_{\overline{\xi}_n} Q_{n,\xi}(\varphi_n)]$. The induction hypothesis, Slutsky's Lemma and standard manipulations yield that

$$N^{1/2} P(N) = N^{1/2}[\gamma_{n-1}^N - \gamma_{n-1}](\mathscr{L}_n\varphi_n) + o_{\mathrm{P}}(1),$$

as required.

This completes the proof of equation (2.9). The proof of equation (2.10) is a consequence of equation (2.9). Indeed, one can verify that the normalized measure η_n^N is related to the unnormalized measure γ_n^N through the identity ([11], page 301),

$$[\eta_n^N - \eta_n](\varphi_n) = \frac{1}{\gamma_n^N(1)}[\gamma_n^N - \gamma_n]\big(\varphi_n - \eta_n(\varphi_n)\big). \tag{2.18}$$

By Corollary 2 we have $\gamma_n^N(1) \xrightarrow{\mathrm{P}} \gamma_n(1)$, so Slutsky's Lemma and equation (2.9) yield that

$$N^{1/2}[\eta_n^N - \eta_n](\varphi_n) \xrightarrow{\mathcal{D}} \mathrm{N}\big\{0, \mathbb{V}_n^\eta(\varphi_n)\big\} \qquad \text{with } \mathbb{V}_n^\eta(\varphi_n) = \frac{\mathbb{V}_n^\gamma(\varphi_n - \eta_n(\varphi_n))}{\gamma_n(1)^2}.$$

Since $\eta_{n-1}[\mathscr{L}_n(\varphi_n - \eta_n(\varphi_n))] = 0$, the second part of (2.11) follows. □

Note that a straightforward induction argument based on equation (2.11) yields that the asymptotic variances can also be expressed as

$$\begin{cases} \mathbb{V}_n^\gamma(\varphi_n) = \displaystyle\sum_{p=0}^{n} \gamma_p(1)^2 \operatorname{Var}_{\eta_p}(\mathscr{L}_{p:n}\varphi_n), \\ \mathbb{V}_n^\eta(\varphi_n) = \displaystyle\sum_{p=0}^{n} \big\{\gamma_p(1)/\gamma_n(1)\big\}^2 \operatorname{Var}_{\eta_p}\big(\mathscr{L}_{p:n}[\varphi_n - \eta_n\varphi_n]\big) \end{cases}$$

with the shorthand notation $\mathscr{L}_{n:n} = I_d$ and $\mathscr{L}_{p:n} = \mathscr{L}_{p+1} \circ \cdots \circ \mathscr{L}_n$.

2.7. *Stability.* Theorem 2.2 shows that the asymptotic variance of the SMC with adaptive scaling is described by the linear operator

$$\mathscr{L}_n\varphi_n = \langle \eta_{n-1}[\partial_{\overline{\xi}_n} Q_{n,\xi}\varphi_n], \xi_n - \overline{\xi}_n \rangle + Q_n\varphi_n.$$

It is a standard result [11] that the asymptotic variances, $\mathbb{V}_n^{\gamma,\infty}$ and $\mathbb{V}_n^{\eta,\infty}$, of the limiting SMC algorithm that does not use summary statistics, satisfy the same recursive equation (2.11) with the only difference that the operator $\mathscr{L}_n$ has to be replaced by its nonadaptive version $\mathscr{L}_n^\infty \equiv Q_n$; in other words, the effect of adaptive scaling is incorporated into the term $\langle \eta_{n-1}[\partial_{\overline{\xi}_n} Q_{n,\xi}\varphi_n], \xi_n - \overline{\xi}_n \rangle$. It should be noted that, in general, the asymptotic variance functionals $(\mathbb{V}_n^\gamma, \mathbb{V}_n^\eta)$ and $(\mathbb{V}_n^{\gamma,\infty}, \mathbb{V}_n^{\eta,\infty})$ cannot be compared; for example, for some test function $\varphi_n \in \mathcal{B}_b(E_n)$, we have $\mathbb{V}_n^\eta(\varphi_n) < \mathbb{V}_n^{\eta,\infty}(\varphi_n)$ while for some other choices of a test function the inequality is reversed. Nevertheless, the next result shows that there are important classes of algorithms where the asymptotic variances are in fact equal, $\mathbb{V}_n^\eta(\varphi_n) = \mathbb{V}_n^{\eta,\infty}(\varphi_n)$ and/or $\mathbb{V}_n^\gamma(\varphi_n) = \mathbb{V}_n^{\gamma,\infty}(\varphi_n)$. Results of the same flavor have been obtained in [5] for adaptive multilevel splitting algorithms.

THEOREM 2.3 (Stability). *Assume* (A1)–(A2).

(1) *Suppose that for any index* $n \geq 1$ *there exists a constant* $C_n > 0$ *such that for any parameter* $\xi \in \mathrm{Range}(\xi_n)$, *we have*

$$\eta_{n-1}(G_{n-1,\xi} M_{n,\xi}) = C_n \eta_n. \tag{2.19}$$

Then for any test function $\varphi_n \in \mathcal{B}_b(E_n)$, *we have* $\mathbb{V}_n^{\gamma}(\varphi_n) = \mathbb{V}_n^{\gamma,\infty}(\varphi_n)$ *and* $\mathbb{V}_n^{\eta}(\varphi_n) = \mathbb{V}_n^{\eta,\infty}(\varphi_n)$.

(2) *Suppose that for any index* $n \geq 1$ *and any parameter* $\xi \in \mathrm{Range}(\xi_n)$, *we have*

$$\frac{\eta_{n-1}(G_{n-1,\xi} M_{n,\xi})}{\eta_{n-1}(G_{n-1,\xi})} = \eta_n. \tag{2.20}$$

Then for any $\varphi_n \in \mathcal{B}_b(E_n)$, *we have* $\mathbb{V}_n^{\eta}(\varphi_n) = \mathbb{V}_n^{\eta,\infty}(\varphi_n)$.

PROOF. Let us prove the two results separately.

(1) To prove the first statement, it suffices to show that $\eta_{n-1}[\partial_\xi Q_{n,\xi}\varphi_n] = 0$ for any $\varphi_n \in \mathcal{B}_b(E_n)$ once (2.19) is satisfied. By differentiation under the integral sign, we have

$$\eta_{n-1}(\partial_\xi Q_{n,\xi}\varphi_n) = \partial_\xi\{\eta_{n-1}(Q_{n,\xi}\varphi_n)\} = \partial_\xi\{C_n\eta_n(\varphi_n)\} = 0,$$

hence the conclusion.

(2) Here, it suffices to show that $\eta_{n-1}(\partial_\xi Q_{n,\xi}[\varphi_n - \eta_n(\varphi_n)]) = 0$ for any test function $\varphi_n \in \mathcal{B}_b(E_n)$ once (2.20) is satisfied. By differentiation under the integral sign, we have

$$\begin{aligned}\eta_{n-1}\big(\partial_\xi Q_{n,\xi}\big[\varphi_n - \eta_n(\varphi_n)\big]\big) &= \partial_\xi\big\{\eta_{n-1}\big(Q_{n,\xi}\big[\varphi_n - \eta_n(\varphi_n)\big]\big)\big\} \\ &= \partial_\xi\big\{\eta_{n-1}(G_{n-1,\xi})\eta_n\big[\varphi_n - \eta_n(\varphi_n)\big]\big\}.\end{aligned}$$

Since $\eta_n\{\varphi_n - \eta_n(\varphi_n)\} = 0$, the conclusion follows. □

Condition (2.19) is strictly stronger than (2.20), as can be seen by considering parametric families of potentials of the type $G_{n,\xi}(x) = f(\xi)g(x)$. Condition (2.19) is very general and applies to the large class of adaptive SMC samplers [13], where for each value of the parameter $\xi \in \mathrm{Range}(\xi_n)$, the Markov kernel $M_{n,\xi}$ lets the distribution η_n be invariant and where the potentials do not actually depend on the parameter ξ; see, for example, Section 2.3.1. This insensitivity to errors in the estimation of the summary statistics means that errors in the estimation of the distributions η_n do not propagate; this stability is the main reason behind the equivalence between the adaptive and nonadaptive variance functionals. Without this stability property, as is typically the case in the setting of McKean–Vlasov diffusions described below, errors can quickly propagate, and the asymptotic variances

are generally different. Note that condition (2.19) also applies to the example of Section 2.3.2 where the weight functions also depend on the summary statistics.

There are important classes of algorithms, though, where the above stability result does not apply. For example, consider the problem of approximating the expectation $\mathrm{E}[\varphi(X_T)]$ where $\varphi : \mathbb{R} \to \mathbb{R}$ is a test function and $\{X_s\}_{s=0}^T$ is a McKean–Vlasov diffusion of the type $dX_t = b(X_t, \eta_t[\xi])\,dt + dW_t$, where $b : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a drift function, $\xi : \mathbb{R} \to \mathbb{R}$ is a given function and $\eta_t(dx) = \mathrm{P}(X_t \in dx)$ denotes the law of the process X at time t. A standard approach consists of using a particle algorithm with a Euler–Maruyama discretization with time step $\Delta \ll 1$; this yields to the class of algorithms described in Section 2 with the underlying Markov chain

$$X_{k+1} \sim X_k + b\big(X_k, \eta_k(\xi)\big)\Delta + \mathrm{N}(0, \Delta).$$

In this setting, conditions (2.19) and (2.20) are typically not satisfied, and in general, the asymptotic variance of the adaptive SMC algorithm is indeed different from the limiting algorithm.

2.8. *Verifying the assumptions.* We consider the sequential Bayesian parameter inference framework of Section 2.3.1 where a parameter $x \in E = \mathbb{R}^p$ has to be estimated from noisy observations $y_i \in \mathcal{Y}$; we assume a prior measure with density $\eta_0(x)$ with respect to the Lebesgue measure and suppose that the following assumptions hold.

ASSUMPTION 3. For each $n \geq 1$ the function $G_n(x) = \mathrm{p}(y_{n+1}|y_{1:n}, x)$ is bounded and strictly positive. The statistics $\xi_n : E \to \mathbb{R}^d$ are bounded: there exists an open and bounded convex $\mathcal{K}$ such that $\xi_n(x) \in \mathcal{K}$ for any index $n \geq 1$ and $x \in E$.

For each index $n \geq 1$, the adaptive SMC algorithm makes use of a parametric family $M_{n,\xi}$, for $\xi \in \mathrm{Range}(\xi_n)$, of random walk Metropolis–Hastings kernels with symmetric proposal density $q : \mathbb{R}^m \times \mathrm{Range}(\xi_n) \to \mathbb{R}^+$,

$$q(z, \xi) = q(-z, \xi).$$

The SMC algorithm targets the sequence $\eta_n(x) \propto \eta_0(x)\,\mathrm{p}(x|y_{1:n})$. At step $n \geq 1$, move $x_n^i \mapsto y_n^i$ is accepted with probability $\min(1, \eta_n(y_n^i)/\eta_n(x_n^i))$. We assume the following second-order regularity conditions on the symmetric proposal kernel q:

ASSUMPTION 4. The first and second derivatives of the kernel density $(z, \xi) \mapsto q(z, \xi)$ are uniformly bounded in $L^1(E)$,

$$\sup\big\{\big\|z \mapsto \partial_{\xi_0} q(z, \xi)\big\|_{L^1(E)} + \big\|z \mapsto \partial_{\xi_0}^2 q(z, \xi)\big\|_{L^1(E)} : \xi_0 \in \mathcal{K}\big\} < \infty.$$

(A3) is weak and satisfied by many statistical models. Consider, for example, the problem of Bayesian parameter estimation for state space models; a vector $x \in E \subset \mathbb{R}^p$ parametrizes the underlying dynamics of a hidden Markov process $\{u_k\}_{k\geq 0}$ and noisy observations distributed as $y_n|u_{1:n} \sim g(u_n, y_n)\,dy_n$ are collected. If the conditional density functions $(u, y) \mapsto g(u, y)$ are bounded, as is, for example, the case when the observations are Gaussian additive perturbations of the latent Markov process with known covariance structure, the weight functions G_n are bounded. The situation where the statistics $\xi_n : E \to \mathbb{R}^d$ are bounded is also common. It is, for example, the case when regularized estimates of the mean and covariance matrix are obtained through the summary statistics $\mathbb{R}^p \ni x \mapsto T(x_i)$ and $\mathbb{R}^p \ni x \mapsto T(x_i x_j)$ for $1 \leq i, j \leq p$ and a threshold function $T : \mathbb{R} \to \mathbb{R}$; a standard choice is $T(x) = \alpha(1_{x\geq\alpha} - 1_{x\leq\alpha}) + x1_{|x|<\alpha}$ for a given threshold $\alpha \gg 1$. Similarly, it is straightforward to construct proposals verifying (A4). Adopting a scalar context ($d = 1$), one can, for example, show that for a bounded interval $\mathcal{K}$ and a precision function $\mathscr{S} : \mathcal{K} \to \mathbb{R}_+$ that is bounded away from zero with bounded first and second derivatives, the Gaussian proposal density $q(z, \xi) \propto \mathscr{S}^{1/2}(\xi) \exp\{-\mathscr{S}(\xi)z^2/2\}$ satisfies (A4). Multi-dimensional extensions of this setting are readily constructed.

PROPOSITION 2.1. *Assume* (A3)–(A4). *The parametric family of Markov kernels* $\{M_{n,\xi}\}_{n\geq 1,\xi\in\mathcal{K}}$ *and potential functions* $\{G_n\}_{n\geq 0}$ *satisfy* (A1)–(A2).

PROOF. By assumption, the potentials $\{G_n\}_{n\geq 0}$ are bounded and strictly positive, and the statistics $\xi_n : E \to \mathbb{R}^d$ are bounded. To verify that (A1)–(A2) are satisfied, it suffices to prove that for any test function $\varphi \in \mathcal{B}_b(E)$, the first and second derivatives of $(x, \xi) \mapsto M_{n,\xi}\varphi(x)$ exist and are uniformly bounded. The Metropolis–Hastings accept–reject ratio of proposal $x \mapsto x + z$ is $r(x, z) := \min(1, \eta_n(x + z)/\eta_n(x))$. Consequently, we have $M_{n,\xi}\varphi(x) = \varphi(x) + \int_{\mathbb{R}^m}[\varphi(x + z) - \varphi(x)]r(x, z)q(z, \xi)\,du$, and differentiation under the integral sign yields that

$$\partial_\xi^\alpha M_{n,\xi}\varphi(x) = \int_{\mathbb{R}^m} \big(\varphi(x + z) - \varphi(x)\big)r(x, z)\partial_\xi^\alpha q(z, \xi)\,dz$$

for $\alpha \in \{1, 2\}$. The conclusion follows from (A4). □

3. Adaptive tempering. We now look at the scenario when one uses the information in the evolving particles population to adapt a sequence of distributions by means of a tempering parameter $\beta \in \mathbb{R}$.

3.1. *Algorithmic set-up*. For clarity, we first describe a nonadaptive version of the algorithm we are interested in. In many situations in Bayesian inference, one seeks to sample from a probability distribution π defined on a measurable space E of the form

$$\pi(dx) = \frac{1}{Z}e^{-V(x)}m(dx),$$

where Z is a normalization constant, $m(dx)$ a dominating measure and $V : E \to \mathbb{R}$ a potential function. A frequently invoked algorithm involves forming a sequence of *tempered* probability distributions

$$\eta_n(dx) = \frac{1}{Z(\beta_n)} e^{-\beta_n V(x)} m(dx)$$

for parameters $\beta_0 \leq \beta_1 \leq \cdots \leq \beta_{n_\star} = 1$. For consistency with the notation of Section 2, and without loss of generality, we assume that $Z(\beta_0) = 1$. The associated unnormalized measures are defined as

$$\gamma_n(dx) = e^{-\beta_n V(x)} m(dx) \quad \text{and} \quad \gamma(dx) = e^{-V(x)} m(dx).$$

Particles are propagated from η_{n-1} to η_n through a Markov kernel M_n that preserves η_n. In other words, the algorithm corresponds to the SMC approach discussed in Section 2 with potential functions

$$G_n(x) = e^{-(\beta_{n+1} - \beta_n) V(x)}$$

and Markov kernels M_n. For a test function $\varphi \in \mathcal{B}_b(E)$, the N-particles approximations of the quantities $\eta_n(\varphi_n)$ and $\gamma_n(\varphi_n)$ are given by $\eta_n^N(\varphi) = (1/N) \sum_{i=1}^N \varphi(x_n^i)$ and $\gamma_n^N(\varphi) = \gamma_n^N(1) \eta_n^N(\varphi)$ with

$$\gamma_n^N(1) = \prod_{p=0}^{n-1} \eta_p^N(G_p).$$

Let us now describe the adaptive version of the above algorithm. In most scenarios of practical interest, it can be difficult or even undesirable to decide a priori upon the tempering sequence $\{\beta_n\}_{n=0}^{n_\star}$. For example, if the chosen sequence of tempering parameters features large gaps, the variance of the resulting weights may potentially be very large due to important discrepancies between consecutive elements of the bridging sequence of probability distributions. At the other extreme, computational time is wasted if the gaps between the tempering parameters are too small. Knowing what constitutes "large" or "small" with regards to the temperature gaps can be very problem specific. Thus an automated procedure for determining the tempering sequence is of great practical importance. The adaptive version of the above described algorithm constructs the (random) tempering sequence $\{\beta_n^N\}_{n \geq 0}$ iteratively. The effective sample size (ESS) is used as measure of diversity; recall that for a measure η on a measurable space E and a weight function $\omega : E \to (0, \infty)$, the ESS is defined by

$$\mathrm{ESS}(\eta, \omega) := \eta(\omega)^2 / \eta(\omega^2).$$

When η is a probability measure and $\eta(\omega) < \infty$, the quantity $\mathrm{ESS}(\eta, \omega)$ can be thought of as a measure of discrepancy between η and the probability measure η_ω whose Radon–Nikodym derivative with respect to η is proportional to ω; a small value of $\mathrm{ESS}(\eta, \omega)$ indicates a large discrepancy between η and η_ω. This measure

of discrepancy is widely used in the SMC literature since normalization constants are not required, and it elegantly deals with probability measures described as weighted particles systems. In our setting, it is natural to fix a threshold $\alpha \in (0, 1)$ and determine the next temperature in the sequence by imposing that the effective sample size stays above α; some discussion on setting α is in [21]. More precisely, the adaptive tempering scheme is defined as follows: the initial tempering parameter β_0^N is set to a prescribed value $\beta_0 < 1$, typically chosen so that the distribution $\eta_0(dx) = e^{-\beta_0 V(x)} m(dx)$ is straightforward to sample from. When the tempering parameter $\beta_n^N \in \mathbb{R}$ and the approximation $\eta_n^N = (1/N) \sum_{i=1}^N \delta_{x_n^i}$ to the probability measure $[1/Z(\beta_n^N)] e^{-\beta_n^N V(x)} m(dx)$ have been computed, the next tempering parameter β_{n+1}^N is determined through the equation

$$\beta_{n+1}^N = \inf\{\beta : \beta > \beta_n^N, \mathrm{ESS}(\eta_n^N, e^{-(\beta-\beta_n^N)V}) = \alpha\}, \tag{3.1}$$

and the next distribution is $\eta_{n+1}^N(dx) \propto e^{-\beta_{n+1}^N V(x)} m(dx)$. Indeed, the adaptive tempering algorithm approximates an SMC algorithm that uses a deterministic tempering sequence $\{\beta_n\}_{n \geq 0}$ given by the analogue of (3.1),

$$\beta_{n+1} = \inf\{\beta : \beta > \beta_n, \mathrm{ESS}(\eta_n, e^{-(\beta-\beta_n)V}) = \alpha\}. \tag{3.2}$$

This nonadaptive algorithm for β_n's as in (3.2) will be referred to as the limiting algorithm. The following result guaranties that under mild assumptions, the effective sample size functional $\beta \mapsto \mathrm{ESS}(\eta_n, e^{-(\beta-\beta_n)V})$ is continuous and decreasing. The proof is deferred to Appendix A.

LEMMA 3.1. *Let η be a probability measure on the measurable space E and $V : E \to \mathbb{R}$ be a bounded potential. Then the function $\lambda \mapsto \mathrm{ESS}(\eta, e^{-\lambda V})$ is continuous and decreasing on $[0, \infty)$. Furthermore, if the random variable $V(X)$ is not almost-surely constant for $X \sim \eta$, the function is strictly decreasing.*

Lemma 3.1 shows that the tempering parameter β_{n+1}^N can be efficiently computed by a standard bisection method. Once the tempering parameter β_{n+1}^N is computed according to (3.1), the particle approximation η_{n+1}^N to the law $\exp\{-\beta_{n+1}^N V(x)\} m(dx)$ is constructed as follows. The particle system $\{x_n^i\}_{i=1}^N$ is re-sampled according to a multinomial scheme with weights proportional to

$$G_{n,N}(x) := e^{-(\beta_{n+1}^N - \beta_n^N)V(x)}$$

and then evolves via an adaptive Markov kernel $M_{n+1,N}$ that preserves the probability distribution $\eta_{n+1}^N(dx) \propto \exp\{-\beta_{n+1}^N V(x)\} m(dx)$. We have assumed, similarly to the previous section, that at each index n a member $M_{n,N} = M_{n,\theta_n^N}$ of a parametric family of Markov kernel $\{M_{n,\theta}\}_{\theta \in \Theta_n}$ is used with

$$\theta_n^N = (\beta_{n-1}^N, \beta_n^N, \eta_{n-1}^N(\xi_n)) \in \Theta_n \equiv [\beta_0, \infty) \times [\beta_0, \infty) \times \mathrm{Range}(\xi_n).$$

The function $\xi_n : E \to \mathbb{R}^d$ is a summary statistic that takes values in an open convex set $\mathrm{Range}(\xi_n) \subset \mathbb{R}^d$, and M_{n,θ_n^N} is a Markov kernel that lets $\eta_n^N(dx) \propto \exp\{-\beta_n^N V(x)\} m(dx)$ be invariant. For convenience, we set

$$M_n \equiv M_{n,\overline{\theta}_n}, \qquad G_n = \exp\{-(\beta_{n+1} - \beta_n)V\}, \qquad \overline{\theta}_n = (\beta_{n-1}, \beta_n, \eta_{n-1}(\xi_n)).$$

As in the previous section, we define

$$Q_{n,N}(x, dy) \equiv G_{n-1,N}(x) M_{n,N}(x, dy)$$

and $Q_n(x, dy) \equiv G_{n-1}(x) M_n(x, dy)$. With this notation, equation (2.5) holds. When the context is clear and in accordance with the previous section, for emphasizing the dependence upon the parameter θ, we sometimes use the notation $M_{n,\theta}$, $G_{n,\theta}$ and $Q_{n,\theta}$. For a function Ψ depending on θ, we use the following shorthand notation: $\partial_\theta \Psi|_{\theta=\overline{\theta}_n} = (\partial_{\beta_{n-1}} \Psi, \partial_{\beta_n} \Psi, \partial_{\overline{\xi}_n} \Psi)$.

In practice, even if equation (3.1) defines an infinite sequence of tempering parameter $\{\beta_n^N\}_{n \geq 0}$, the algorithm finishes after a finite number of steps. We have introduced the infinite sequence $\{\beta_n^N\}_{n \geq 0}$ only to facilitate the presentation of the convergence analysis of the adaptive algorithm. The adaptive tempering algorithm terminates after τ^N steps, with

$$\tau^N = \inf\{k : k \geq 1, \beta_k^N \geq 1\}. \tag{3.3}$$

The ideal limiting algorithm stops after a number of steps $\tau = \inf\{k : k \geq 0, \beta_k \geq 1\}$. For convenience, we assume in the remainder of this section that the following condition holds.

ASSUMPTION 5. The deterministic tempering sequence $\{\beta_n\}_{n \geq 0}$ for the limiting algorithm is such that

$$\beta_\tau > 1. \tag{3.4}$$

This condition holds in most realistic scenarios; indeed in [21, 23] this occurs in quite complex examples. As proved in Theorem 3.2, (A5) ensures that the stopping time τ^N converges in probability, as $N \to \infty$, to the deterministic number of steps $\tau \in \mathbb{N}$. The output of the adaptive SMC algorithm is a particle approximation $\pi^N = (1/N) \sum_{i=1}^N \delta_{x_\star^i}$ of the probability distribution π obtained as follows. (We describe the last step of the algorithm; earlier steps are obtained in the standard way.) Particles $\{x^i_{\tau^N - 1}\}_{i=1}^N$ are resampled with weights proportional to $G_\star(x^i_{\tau^N - 1}, \beta^N_{\tau^N - 1})$, where the function $G_\star : E \times [\beta_0, 1] \to \mathbb{R}$ is defined as

$$G_\star(x, \beta) = \exp\{-(1 - \beta)V(x)\}, \tag{3.5}$$

and evolve via a Markov kernel $M_{\star,N} \equiv M_{\star,\theta_\star^N}$ that leaves π invariant to obtain the set of particles $\{x_\star^i\}_{i=1}^N$, where $\theta_\star^N = (\beta^N_{\tau^N - 1}, \eta^N_{\tau^N - 1}(\xi_{\tau^N})) \in \Theta_\star \subset \mathbb{R} \times \mathbb{R}^d$. For

a test function $\varphi : E \to \mathbb{R}$, the quantity

$$\pi^N(\varphi) = (1/N) \sum_{i=1}^{N} \varphi(x_\star^i) \tag{3.6}$$

is an approximation of $\pi(\varphi)$, and the unnormalized quantity

$$\gamma^N(\varphi) = \gamma^N_{\tau^N-1}(1) \times \eta^N_{\tau^N-1}\big[G_\star(\cdot, \beta^N_{\tau^N-1})\big] \times \pi^N(\varphi) \tag{3.7}$$

approximates $\gamma(\varphi)$.

We prove in Sections 3.3 and 3.4 that under mild assumptions, the output of the SMC algorithm with adaptive tempering sequence and adaptive Markov kernels satisfies a weak law of large numbers (Theorem 3.1) and a central limit theorem (Theorem 3.4). For a test function $\varphi \in \mathcal{B}_b(E)$,

$$\pi^N(\varphi) \xrightarrow{\mathrm{P}} \pi(\varphi) \quad \text{and} \quad N^{1/2}\{\pi^N(\varphi) - \pi(\varphi)\} \xrightarrow{\mathcal{D}} \mathrm{N}\big(0, \widetilde{\mathbb{V}}^\pi(\varphi)\big).$$

The variance functional $\widetilde{\mathbb{V}}^\pi : \mathcal{B}_b(E) \to \mathbb{R}^+$ is discussed in Section 3.4.

3.2. *Assumptions.* The results to be presented in the next section make use of the following hypotheses:

ASSUMPTION 6. The potential $V : E \to \mathbb{R}$ and the summary statistics $\xi_n : E \to \mathbb{R}^d$ are bounded on E. For each $n \geq 1$ and test function $\varphi \in \mathcal{B}_b(E)$, the functions $(x, \theta) \mapsto G_{n-1,\theta}(x)$ and $(x, \theta) \mapsto Q_{n,\theta}\varphi(x)$ are continuous at $\overline{\theta}_n$ uniformly on E.

ASSUMPTION 7. For each $n \geq 1$ and test function $\varphi : E \to \mathbb{R}$, the function $(x, \theta) \mapsto \partial_\theta Q_{n,\theta}\varphi(x)$ is well defined, bounded and continuous at $\overline{\theta}_n$ uniformly on E.

These assumptions are analogous to (A1)–(A2). As previously mentioned, we believe that these conditions could be relaxed to accommodate unbounded test functions at the cost of considerable technical complications in the proofs.

3.3. *Weak law of large numbers.* In this section we prove that the adaptive tempering procedure is consistent. To do so, we first establish that the empirical tempering parameters are such that $\beta_n^N \xrightarrow{\mathrm{P}} \beta_n$ and that for any test function $\varphi \in \mathcal{B}_b(E)$, we have $\eta_n^N(\varphi) \xrightarrow{\mathrm{P}} \eta_n(\varphi)$. These results are then used to establish that $\tau^N \xrightarrow{\mathrm{P}} \tau$ and $\pi^N(\varphi) \xrightarrow{\mathrm{P}} \pi(\varphi)$.

THEOREM 3.1 (WLLN for SMC with adaptive scaling and adaptive tempering). *Assume* (A6). *For any $n \geq 0$ we have*

$$\beta_n^N \xrightarrow{\mathrm{P}} \beta_n. \tag{3.8}$$

For a Polish space V, *sequence* $\{V_N\}_{N\geq 0}$ *of* V*-valued random variables such that* $V_N \xrightarrow{\mathrm{P}} \mathsf{v} \in \mathsf{V}$ *and bounded measurable function* $\varphi_n : E \times \mathsf{V} \to \mathbb{R}$ *continuous at* $\mathsf{v} \in \mathsf{V}$ *uniformly on* E, *we have*

$$\eta_n^N\big[\varphi_n(\cdot, V_N)\big] \xrightarrow{\mathrm{P}} \eta_n\big[\varphi_n(\cdot, \mathsf{v})\big]. \tag{3.9}$$

COROLLARY 3. *Assume* (A6). *For a test function* $\varphi_n \in \mathcal{B}_b(E)$ *we have* $\eta_n^N(\varphi_n) \xrightarrow{\mathrm{P}} \eta_n(\varphi_n)$ *and* $\gamma_n^N(\varphi_n) \xrightarrow{\mathrm{P}} \gamma_n(\varphi_n)$.

PROOF OF THEOREM 3.1. The proof is by induction on the rank $n \geq 0$ that $\beta_n^N \xrightarrow{\mathrm{P}} \beta_n$ and that equation (3.9) holds for any test function $\varphi : E \times \mathsf{V} \to \mathbb{R}$ bounded and continuous at $\mathsf{v} \in \mathsf{V}$ uniformly on E. The initial case $n = 0$ is a direct consequence of WLLN for i.i.d. random variables and Definition 2.1. We assume the result at rank $n-1$ and proceed to the induction step:

- We first focus on proving that $\beta_n^N \xrightarrow{\mathrm{P}} \beta_n$. Note that β_n^N can also be expressed as

$$\beta_n^N = \inf\left\{\beta : \beta > \beta_0, \frac{\zeta_{1,n-1}^N(\beta)}{\zeta_{2,n-1}^N(\beta)} = \alpha\right\}$$

for the functionals $\zeta_{1,n-1}^N(\beta) = \eta_{n-1}^N[e^{-\max(0,\beta-\beta_{n-1}^N)V}]^2$ and $\zeta_{2,n-1}^N(\beta) = \eta_{n-1}^N[e^{-2\max(0,\beta-\beta_{n-1}^N)V}]$. Indeed, the limiting temperature β_n can also be expressed as

$$\beta_n = \inf\left\{\beta : \beta > \beta_0, \frac{\zeta_{1,n-1}(\beta)}{\zeta_{2,n-1}(\beta)} = \alpha\right\},$$

where $\zeta_{1,n-1}(\beta)$ and $\zeta_{2,n-1}(\beta)$ are the limiting values of $\zeta_{1,n-1}^N(\beta)$ and $\zeta_{2,n-1}^N(\beta)$. The dominated convergence theorem now implies that the mappings $\beta \mapsto \zeta_{1,n-1}^N(\beta)/\zeta_{2,n-1}^N(\beta)$ and $\beta \mapsto \zeta_{1,n-1}(\beta)/\zeta_{2,n-1}(\beta)$ are continuous; it is thus sufficient to prove that on any compact set $[\beta_0, \beta_+]$, where β_+ is any real number, we have

$$\sup\left\{\frac{\zeta_{1,n-1}^N(\beta)}{\zeta_{2,n-1}^N(\beta)} - \frac{\zeta_{1,n-1}(\beta)}{\zeta_{2,n-1}(\beta)} : \beta \in [\beta_0, \beta_+]\right\} \xrightarrow{\mathrm{P}} 0. \tag{3.10}$$

Lemma 3.1 shows that the function $\beta \mapsto \zeta_{1,n-1}^N(\beta)/\zeta_{1,n-1}^N(\beta)$ is decreasing on $[\beta_0, \beta_+]$ for any $n, N \geq 1$. By Dini's theorem, on a compact interval, if a sequence of decreasing functions converges pointwise toward a continuous function, then the convergence is also uniform; it follows that for proving equation (3.10), it suffices to show that for any fixed parameter $\beta \in [\beta_0, \beta_+]$, the difference $\zeta_{1,n-1}^N(\beta)/\zeta_{2,n-1}^N(\beta) - \zeta_{1,n-1}(\beta)/\zeta_{2,n-1}(\beta)$ converges to zero in

probability. Indeed, one can focus on proving that $\zeta_{i,n-1}^N(\beta)$ converges in probability to $\zeta_{i,n-1}(\beta)$ for $i \in \{1,2\}$. We present the proof for $i=2$, the case $i=1$ being similar. The difference $\zeta_{2,n-1}^N(\beta) - \zeta_{2,n-1}(\beta)$ decomposes as

$$\begin{aligned} &\eta_{n-1}^N\big[e^{-2\max(0,\beta-\beta_{n-1}^N)V} - e^{-2\max(0,\beta-\beta_{n-1})V}\big] \\ &\quad + \big(\eta_{n-1}^N - \eta_{n-1}\big)\big[e^{-2\max(0,\beta-\beta_{n-1})V}\big]. \end{aligned} \tag{3.11}$$

For any fixed β, since $\beta_{n-1}^N \xrightarrow{\mathrm{P}} \beta_{n-1}$ and the potential function V is bounded, the first term in (3.11) goes to zero in probability. The induction hypothesis yields that the second term also goes to zero in probability.

- Since $\beta_n^N \xrightarrow{\mathrm{P}} \beta_n$, $\eta_{n-1}^N(\xi_n) \xrightarrow{\mathrm{P}} \eta_{n-1}(\xi_n)$ and $V_n \xrightarrow{\mathrm{P}} \mathsf{v}$, the same approach as the one used in the proof of Theorem 2.1 shows that $\eta_n^N[\varphi_n(\cdot, V_N)] \to \eta_n[\varphi_n(\cdot,\mathsf{v})]$ in probability. □

We conclude this section by proving that the output (3.6) of the adaptive SMC algorithm with adaptive tempering sequence is consistent.

THEOREM 3.2. *Assume* (A5)–(A6). *Consider a test function* $\varphi \in \mathcal{B}_b(E)$. *We have*

$$\tau^N \xrightarrow{\mathrm{P}} \tau \quad \textit{and} \quad \pi^N(\varphi) \xrightarrow{\mathrm{P}} \pi(\varphi).$$

PROOF. By Theorem 3.1 and Assumption 5, we have $\beta_{\tau-1}^N \xrightarrow{\mathrm{P}} \beta_{\tau-1} < 1$ and $\beta_\tau^N \xrightarrow{\mathrm{P}} \beta_\tau > 1$, from which it directly follows that $\tau^N \xrightarrow{\mathrm{P}} \tau$. For proving that $\pi^N(\varphi) \xrightarrow{\mathrm{P}} \pi(\varphi)$, it is enough to notice that the convergence $\tau^N \xrightarrow{\mathrm{P}} \tau$ yields that for any test function $h \in \mathcal{B}_b(E)$, we have $\eta_{\tau^N-1}(h) \xrightarrow{\mathrm{P}} \eta_{\tau-1}(h)$, from which the conclusion readily follows by the definition of π^N and the same arguments used in the proof of Theorem 2.1. □

3.4. *Central limit theorem.* Here we extend the fluctuation analysis of Section 2.6 to the adaptive tempering setting. We show that the differences $N^{1/2}(\beta_n^N - \beta_n)$ and $N^{1/2}[\gamma_n^N - \gamma_n](\varphi_n)$ satisfy a joint CLT and then use these results to obtain a CLT for the outputs $\pi^N(\varphi)$ and $\gamma^N(\varphi)$ of the adaptive SMC algorithm. The proofs are more intricate than the proof of Theorem 2.2 since the correlation between the fluctuations of $(\beta_n^N - \beta_n)$ and $[\gamma_n^N - \gamma_n](\varphi_n)$ have to be taken into account. Theorem 3.3 states that under mild assumptions, for any index $n \geq 0$ there exist variance functionals $\widetilde{\mathbb{V}}_n^\gamma, \widetilde{\mathbb{V}}_n^\eta : \mathbb{R} \times \mathcal{B}_b(E) \to \mathbb{R}^+$ such that for any test function $\varphi_n \in \mathcal{B}_b(E)$ and scalar $\alpha \in \mathbb{R}$,

$$\begin{cases} N^{1/2}\{\alpha(\beta_n^N - \beta_n) + [\gamma_n^N - \gamma_n](\varphi_n)\} \xrightarrow{\mathcal{D}} \mathrm{N}\big(0, \widetilde{\mathbb{V}}_n^\gamma(\alpha,\varphi_n)\big), \\ N^{1/2}\{\alpha(\beta_n^N - \beta_n) + [\eta_n^N - \eta_n](\varphi_n)\} \xrightarrow{\mathcal{D}} \mathrm{N}\big(0, \widetilde{\mathbb{V}}_n^\eta(\alpha,\varphi_n)\big). \end{cases} \tag{3.12}$$

The variance functionals are such that $\widetilde{\mathbb{V}}_0^{\eta}(\alpha,\varphi)=\widetilde{\mathbb{V}}_0^{\gamma}(\alpha,\varphi)=\mathrm{Var}_{\eta_0}(\varphi)$ and for any $n\geq 1$,

$$\begin{cases}\widetilde{\mathbb{V}}_n^{\eta}(\alpha,\varphi_n)=\widetilde{\mathbb{V}}_{n-1}^{\eta}\big(\alpha+\Gamma_n(\varphi_n),\alpha\overline{H}_{n-1}+\widetilde{\mathscr{L}}_n\varphi_n\big)+\gamma_n^2(1)\,\mathrm{Var}_{\eta_n}(\varphi_n),\\ \widetilde{\mathbb{V}}_n^{\gamma}(\alpha,\varphi_n)=\widetilde{\mathbb{V}}_n^{\gamma}\big(\alpha,[\varphi_n-\eta_n(\varphi_n)]/\gamma_n(1)\big).\end{cases}\tag{3.13}$$

The operators $\widetilde{\mathscr{L}}_n:\mathcal{B}_b(E_n)\to\mathcal{B}_b(E_{n-1})$ and $\Gamma_n:\mathcal{B}_b(E_n)\to\mathbb{R}$ are defined as

$$\begin{cases}\widetilde{\mathscr{L}}_n\varphi_n=\gamma_{n-1}(\partial_{\beta_n}Q_{n,\theta}\varphi_n)(\overline{H}_{n-1})+Q_n\varphi_n,\\ \Gamma_n(\varphi_n)=\gamma_{n-1}\big[(\partial_{\beta_n}+\partial_{\beta_{n-1}})Q_{n,\theta}\varphi_n\big],\end{cases}\tag{3.14}$$

where we have defined, for convenience, the quantity

$$\begin{aligned}H_n&=\frac{1}{\gamma_{n-1}(1)}\frac{\eta_n(e^{-2\Delta_nV})e^{-\Delta_nV}-\eta_n(e^{-\Delta_nV})e^{-2\Delta_nV}/2}{\eta_n(e^{-2\Delta_nV})\eta_n(Ve^{-\Delta_nV})-\eta_n(e^{-\Delta_nV})\eta_n(Ve^{-2\Delta_nV})}\\&\in\mathcal{B}_b(E)\end{aligned}\tag{3.15}$$

and its centered version $\overline{H}_n=H_n-\eta_n(H_n)$; also $\Delta_n\equiv\beta_{n+1}-\beta_n$. Equation (A.1) readily implies that H_n is well defined as soon as the potential V is not π-almost surely constant. Indeed, equation (3.12) yields that for any test function $\varphi_n\in\mathcal{B}_b(E)$, we have

$$\begin{cases}N^{1/2}[\eta_n^N-\eta_n](\varphi_n)\xrightarrow{\mathcal{D}}\mathrm{N}\big(0,\widetilde{\mathbb{V}}_n^{\eta}(\varphi_n)\big),\\ N^{1/2}[\gamma_n^N-\gamma_n](\varphi_n)\xrightarrow{\mathcal{D}}\mathrm{N}\big(0,\widetilde{\mathbb{V}}_n^{\gamma}(\varphi_n)\big),\end{cases}\tag{3.16}$$

with the shorthand notation $\widetilde{\mathbb{V}}_n^{\gamma}(\varphi_n)\equiv\widetilde{\mathbb{V}}_n^{\gamma}(0,\varphi_n)$ and $\widetilde{\mathbb{V}}_n^{\eta}(\varphi_n)\equiv\widetilde{\mathbb{V}}_n^{\eta}(0,\varphi_n)$. The induction formula (3.14) shows that, in the spirit of Section 2.7, since no derivative with respect to ξ is present, the adaptivity of the Markov kernel does not influence the asymptotic variance functionals; only the adaptivity of the tempering scheme matters. As explained in Section 2.7, the main reason behind this phenomenon is that for any value ξ of the summary statistics, the Markov kernels $M_{n,(\beta_{n-1}^N,\beta_n^N,\xi)}$ leave the probability with the density proportional to $\exp\{-\beta_n^NV\}$ invariant.

THEOREM 3.3 (CLTs for SMC with adaptive scaling and adaptive tempering). *Assume* (A6)–(A7). *For* $n\geq 0$ *and* $\varphi_n\in\mathcal{B}_b(E_n)$, *the equations in* (3.12) *hold*, *and the asymptotic variance functionals* $\widetilde{\mathbb{V}}_n^{\gamma}$ *and* $\widetilde{\mathbb{V}}_n^{\eta}$ *satisfy* (3.13).

PROOF. Let us first prove the result for the unnormalized measures γ_n^N. We proceed by induction on the index $n\geq 0$. The case $n=0$ follows from the usual CLT for i.i.d. random variables. To prove the induction step we decompose the random vector $(\beta_n^N-\beta_n,[\gamma_n^N-\gamma_n](\varphi_n))^{\top}$ as $P(N)+R(N)$ with

$$\begin{aligned}P(N)&=\mathrm{E}\left[\left\{\begin{array}{c}\beta_n^N-\beta_n\\ [\gamma_n^N-\gamma_n](\varphi_n)\end{array}\right\}\Big|\mathscr{F}_{n-1}^N\right],\\ R(N)&=\left\{\begin{array}{c}\beta_n^N-\beta_n\\ [\gamma_n^N-\gamma_n](\varphi_n)\end{array}\right\}-P(N).\end{aligned}$$

To establish (3.13) one can verify that for any vector $t = (u, v)^\top \in \mathbb{R}^2$,

$$\mathrm{E}\big[\exp\{iN^{1/2}\langle t, R(N)\rangle\}|\mathscr{F}_{n-1}^N\big] \xrightarrow{\mathrm{P}} \exp\left\{-\frac{v^2}{2}\gamma_n(1)^2\,\mathrm{Var}_{\eta_n}(\varphi_n)\right\}, \tag{3.17}$$

and the random vector $P(N)$ is such that

$$\begin{aligned} N^{1/2}P(N) = N^{1/2}&\begin{Bmatrix} (\beta_{n-1}^N - \beta_{n-1}) + [\gamma_{n-1}^N - \gamma_{n-1}](\overline{H}_{n-1}) \\ \Gamma_n(\varphi_n) \times (\beta_{n-1}^N - \beta_{n-1}) + [\gamma_n^N - \gamma_n](\widetilde{\mathscr{L}}_n\varphi_n) \end{Bmatrix} \\ &+ o_{\mathrm{P}}(1). \end{aligned} \tag{3.18}$$

Equation (3.13) immediately follows from equations (3.17) and (3.18) and Levy's characterization of convergence in distribution. We now prove that equations (3.17) and (3.18) hold:

- *Proof of equation* (3.18).
  We first deal with the first coordinate of $P(N)$ and start by noting that $\mathrm{E}_{n-1}[\beta_n^N - \beta_n] = \beta_n^N - \beta_n$. For convenience, we set $\Delta_{n-1} \equiv \beta_n - \beta_{n-1}$ and $\Delta_{n-1}^N \equiv \beta_n^N - \beta_{n-1}^N$. The equation $\mathrm{ESS}(\eta_{n-1}^N, e^{-\Delta_{n-1}^N V}) = \alpha = \mathrm{ESS}(\eta_{n-1}, e^{-\Delta_{n-1}V})$ can then be re-arranged to obtain that

$$\begin{aligned} &\eta_{n-1}(e^{-\Delta_{n-1}V})^2\{\eta_{n-1}^N(e^{-2\Delta_{n-1}^N V}) - \eta_{n-1}(e^{-2\Delta_{n-1}V})\} \\ &\quad = \eta_{n-1}(e^{-2\Delta_{n-1}V})\{\eta_{n-1}^N(e^{-\Delta_{n-1}^N V})^2 - \eta_{n-1}(e^{-\Delta_{n-1}V})^2\}. \end{aligned} \tag{3.19}$$

  The term $\eta_{n-1}^N(e^{-2\Delta_{n-1}^N V}) - \eta_{n-1}(e^{-2\Delta_{n-1}V})$ decomposes as the sum of $\eta_{n-1}^N(e^{-2\Delta_{n-1}^N V} - e^{-2\Delta_{n-1}V})$ and $[\eta_{n-1}^N - \eta_{n-1}](e^{-2\Delta_{n-1}V})$. The boundedness of potential V and Theorem 3.1 yield that [we use the notation $\eta_{n-1}^N(x \mapsto \varphi(x))$ to represent $\eta_{n-1}^N(\varphi)$, as this clarifies some calculations]

$$\begin{aligned} &\eta_{n-1}^N(e^{-2\Delta_{n-1}^N V} - e^{-2\Delta_{n-1}V})/(\Delta_{n-1}^N - \Delta_{n-1}) \\ &\quad = \eta_{n-1}^N\left(x \mapsto \int_{\lambda=0}^{\lambda=1} \frac{d}{ds}e^{-2sV(x)}\Big|_{s=\Delta_{n-1}+\lambda(\Delta_{n-1}^N - \Delta_{n-1})} d\lambda\right) \\ &\quad = -2\eta_{n-1}(Ve^{-2\Delta_{n-1}V}) + o_{\mathrm{P}}(1). \end{aligned} \tag{3.20}$$

  It follows that the difference $\eta_{n-1}^N(e^{-2\Delta_{n-1}^N V}) - \eta_{n-1}(e^{-2\Delta_{n-1}V})$ equals

$$\begin{aligned} &\{-2\eta_{n-1}(Ve^{-2\Delta_{n-1}V}) + o_{\mathrm{P}}(1)\}(\Delta_{n-1}^N - \Delta_{n-1}) \\ &\quad + [\eta_{n-1}^N - \eta_{n-1}](e^{-2\Delta_{n-1}V}). \end{aligned} \tag{3.21}$$

  Similarly, the difference $\eta_{n-1}^N(e^{-\Delta_{n-1}^N V})^2 - \eta_{n-1}(e^{-\Delta_{n-1}V})^2$ equals

$$\begin{aligned} &\{-2\eta_{n-1}(e^{-\Delta_{n-1}V})\eta_{n-1}(Ve^{-\Delta_{n-1}V}) + o_{\mathrm{P}}(1)\} \times (\Delta_{n-1}^N - \Delta_{n-1}) \\ &\quad + \{2\eta_{n-1}(e^{-\Delta_{n-1}V}) + o_{\mathrm{P}}(1)\}[\eta_{n-1}^N - \eta_{n-1}](e^{-\Delta_{n-1}V}). \end{aligned} \tag{3.22}$$

Since $(\Delta_{n-1}^N - \Delta_{n-1})$ equals $(\beta_n^N - \beta_n) - (\beta_{n-1}^N - \beta_{n-1})$, standard algebraic manipulations, Slutsky's lemma, equations (3.19), (3.21), (3.22) and the induction hypothesis yield that

$$\begin{aligned} N^{1/2}(\beta_n^N - \beta_n) &= N^{1/2}(\beta_{n-1}^N - \beta_{n-1}) \\ &\quad + N^{1/2}[\eta_{n-1}^N - \eta_{n-1}](\gamma_{n-1}(1) \times H_{n-1}) + o_{\mathrm{P}}(1) \\ &= N^{1/2}(\beta_{n-1}^N - \beta_{n-1}) + N^{1/2}[\gamma_{n-1}^N - \gamma_{n-1}](\overline{H}_{n-1}) + o_{\mathrm{P}}(1), \end{aligned} \tag{3.23}$$

where the functions $\{H_n\}_{n\geq 0}$ are defined in equation (3.15). This completes the proof of the first coordinate of equation (3.18). For dealing with the second coordinate of $P(N)$, we use

$$\begin{aligned} &\mathrm{E}_{n-1}[(\gamma_n^N - \gamma_n)(\varphi_n)] \\ &\quad = \gamma_{n-1}^N(1)\eta_{n-1}^N[Q_{n,N} - Q_n](\varphi_n) + [\gamma_{n-1}^N - \gamma_{n-1}](Q_n\varphi_n). \end{aligned} \tag{3.24}$$

(A6)–(A7), Theorem 3.1 and the same approach as that used to prove (3.20) show that the term $\eta_{n-1}^N[Q_{n,N} - Q_n](\varphi_n)$ equals

$$\begin{aligned} &\{\eta_{n-1}[\partial_{\beta_{n-1}} Q_n\varphi_n] + o_{\mathrm{P}}(1)\}(\beta_{n-1}^N - \beta_{n-1}) \\ &\quad + \{\eta_{n-1}[\partial_{\beta_n} Q_n\varphi_n] + o_{\mathrm{P}}(1)\}(\beta_n^N - \beta_n). \end{aligned}$$

Note that there is no term involving the derivative with respect to the value of the summary statistics; indeed, this is because for any value of $\xi \in \mathrm{Range}(\xi_n)$, the Markov kernel $M_{n,\theta}$, with $\theta = (\beta_{n-1}, \beta_n, \xi)$, preserves η_n so that one can readily check that $\eta_{n-1}[\partial_{\overline{\xi}_n} Q_{n,\theta}\varphi_n] = 0$. Equation (3.23), Slutsky's lemma and equation (3.24) yield that

$$\begin{aligned} N^{1/2}\mathrm{E}_{n-1}[(\gamma_n^N - \gamma_n)(\varphi_n)] &= N^{1/2}\Gamma_n(\varphi_n) \times (\beta_{n-1}^N - \beta_{n-1}) \\ &\quad + N^{1/2}[\gamma_{n-1}^N - \gamma_{n-1}](\widetilde{\mathscr{L}}_n\varphi_n) + o_{\mathrm{P}}(1). \end{aligned}$$

The operators $\widetilde{\mathscr{L}}_n$ and Γ_n are defined in equation (3.14). This completes the proof of (3.18).

- *Proof of equation* (3.17).
  Since $\beta_n^N \in \mathscr{F}_{n-1}^N$, the first coordinate of $R(N)$ is zero, and proving equation (3.17) is equivalent to showing that for any $v \in \mathbb{R}$, we have

$$\mathrm{E}[\exp\{iN^{1/2}vR_1(N)\}|\mathscr{F}_{n-1}^N] \xrightarrow{\mathrm{P}} \exp\{-\tfrac{1}{2}v^2\gamma_n(1)^2\,\mathrm{Var}_{\eta_n}(\varphi_n)\}$$

  with $R_1(N) = \gamma^N(\varphi_n) - \mathrm{E}[\gamma^N(\varphi_n)|\mathscr{F}_{n-1}^N]$; this quantity can also be expressed as $R_1(N) = \gamma_n^N(1)\{\sum_{i=1}^N U_{N,i} - \mathrm{E}[U_{N,i}|\mathscr{F}_{n-1}^N]\}/N$ where the random variables $\{U_{N,i}\}_{i=1}^N$ are, conditionally on $\mathscr{F}_{n-1}^N$, independent and identically distributed as $\varphi_n(X_{N,n})$ with $X_{N,n}$ distributed as described in (2.14). The proof of equation (3.17) is thus identical to the proof of equation (2.12), and thus omitted.

The proof of the CLT for the normalized measures γ_n^N is complete. The proof of the CLT for the normalized measures η_n^N then directly follows from equation (2.18). $\square$

We conclude this section by proving that the output of the adaptive SMC algorithm with adaptive tempering sequence also satisfies a CLT. Because the last step of the adaptive tempering algorithm is slightly different from the other steps, the final parameter β being known (and equal to one) and not empirically determined through equation (3.1), in order to state the next result, we need to introduce some notation. Let $\gamma_\star^N$ be the random unnormalized measure $\gamma_\star^N = \gamma_\star^N(1)\pi_\star^N$ with

$$\gamma_\star^N(1) = \gamma_{\tau-1}^N(1) \times \eta_{\tau-1}^N\big[G_\star(\cdot, \beta_{\tau-1}^N)\big];$$

the random probability measure $\pi_\star^N = (1/N)\sum_{i=1}^N \delta_{x_\star^i}$ is obtained from the measure $\eta_{\tau-1}^N = (1/N)\sum_{i=1}^N \delta_{x_{\tau-1}^i}$ after a multinomial re-sampling step with weights proportional to $G_\star(x_{\tau-1}^i, \beta_{\tau-1}^N)$ and a Markovian transport step with kernel $M_{\star,\beta_{\tau-1}^N,\eta_{\tau-1}^N(\xi_\tau)}$. Note the subtle difference between $(\pi_\star^N, \gamma_\star^N)$ and (π^N, γ^N): the former are obtained after τ steps of the adaptive SMC algorithm while the later are computed after τ^N steps. We have $\gamma_\star^N(\varphi) = \gamma_{\tau-1}^N(Q_{\star,N}\varphi)$ with

$$Q_{\star,N}(x, dy) = G_\star\big(x, \beta_{\tau-1}^N\big)M_{\star,\beta_{\tau-1}^N,\eta_{\tau-1}^N(\xi_\tau)}(x, dy).$$

Corollary 3 yields that $\gamma_\star^N(1) \to \gamma(1) = Z(1)/Z(\beta_0) = Z$. The next result shows that under mild assumptions, for any test function $\varphi \in \mathcal{B}_b(E)$, we have

$$\begin{cases} N^{1/2}\big[\gamma^N - \gamma\big](\varphi) \xrightarrow{\mathcal{D}} \mathrm{N}\big(0, \widetilde{\mathbb{V}}^\gamma(\varphi)\big), \\ N^{1/2}\big[\pi^N - \pi\big](\varphi) \xrightarrow{\mathcal{D}} \mathrm{N}\big(0, \widetilde{\mathbb{V}}^\pi(\varphi)\big), \end{cases} \tag{3.25}$$

where the variance functionals $\widetilde{\mathbb{V}}^\gamma, \widetilde{\mathbb{V}}^\pi : \mathcal{B}_b(E) \to \mathbb{R}$ are defined as

$$\begin{cases} \widetilde{\mathbb{V}}^\gamma(\varphi) = \widetilde{\mathbb{V}}_{\tau-1}^\gamma\big(\gamma_{\tau-1}(\partial_{\beta_{\tau-1}} Q_\star\varphi), Q_\star\varphi\big) + Z^2 \operatorname{Var}_\pi(\varphi), \\ \widetilde{\mathbb{V}}^\pi(\varphi) = \widetilde{\mathbb{V}}^\gamma\big(\big[\varphi - \pi(\varphi)\big]/Z\big), \end{cases} \tag{3.26}$$

where $Q_\star(x, dy) = G_\star(x, \beta_{\tau-1})M_{\star,\beta_{\tau-1},\eta_{\tau-1}(\xi_\tau)}(x, dy)$.

THEOREM 3.4. *Assume* (A5)–(A7). *For any test function $\varphi_n \in \mathcal{B}_b(E)$ the convergence in distribution* (3.25) *holds. The variance functional $\widetilde{\mathbb{V}}^\pi$ satisfies equation* (3.26).

PROOF. Since $\tau^N \xrightarrow{\mathrm{P}} \tau$ and the rescaled output of the algorithm can be expressed as

$$N^{1/2}\big\{\gamma^N(\varphi) - \gamma(\varphi)\big\} = N^{1/2}\big\{\gamma^N(\varphi) - \gamma(\varphi)\big\}\mathbb{I}\big(\tau^N = \tau\big) + o_{\mathrm{P}}(1),$$

to prove Theorem 3.2, it is enough to prove that $N^{1/2}[\gamma_\star^N - \gamma](\varphi)$ and $N^{1/2}[\pi_\star^N - \pi](\varphi)$ converge in distribution toward centered Gaussian distribution with variance $\widetilde{\mathbb{V}}^\gamma(\varphi)$ and $\widetilde{\mathbb{V}}^\pi(\varphi)$. The same computations as those presented in the proof of Theorem 3.3 yield that

$$N^{1/2} P_\star(N) = N^{1/2}\{\gamma_{\tau-1}(\partial_{\beta_{\tau-1}} Q_\star\varphi)(\beta_{\tau-1}^N - \beta_{\tau-1}) + [\gamma_{\tau-1}^N - \gamma_{\tau-1}](Q_\star\varphi)\} + o_{\mathrm{P}}(1)$$

and

$$\mathrm{E}[\exp\{iN^{1/2}tR_\star(N)\}|\mathscr{F}_{\tau-1}^N] \xrightarrow{\mathrm{P}} \exp\{-\tfrac{1}{2}t^2 Z^2(1)\,\mathrm{Var}_\pi(\varphi)\}$$

with $P_\star(N) = \mathrm{E}[(\gamma_\star^N - \gamma)(\varphi)|\mathscr{F}_{\tau-1}^N]$ and $R_\star(N) = [\gamma_\star^N - \gamma](\varphi) - P_\star(N)$, which readily implies that $N^{1/2}[\gamma_\star^N - \gamma](\varphi) \to \mathrm{N}(0, \widetilde{\mathbb{V}}^\gamma(\varphi))$, in distribution. The weak convergence for the sequence $N^{1/2}[\pi_\star^N - \gamma](\varphi)$ follows from the identity $[\pi_\star^N - \pi](\varphi) = \{\gamma_\star^N(1)\}^{-1}[\gamma_\star^N - \gamma](\varphi - \pi(\varphi))$. □

4. Applications.

4.1. *A simple numerical example.* To investigate the impact of our CLT, we consider an SMC algorithm which adapts the scale of the proposal in an MCMC kernel, by using the sample covariance of the previous target. Let $x \in \mathbb{R}^p$, and consider the Gaussian target distribution η_n with density with respect to the Lebesgue measure in $\mathbb{R}^p$ given by

$$\eta_n(x) \propto \exp\{-\tfrac{1}{2}\langle x, \Gamma_n^{-1}x\rangle\}.$$

The covariance matrices are given by $\Gamma_n = L_n L_n^\top$ for $L_n = (10(1 - n/99) + 0.1n/99)I + 0.5nJ/99$, where I is an identity matrix, and J is a lower triangular matrix with $J_{i,j} = 1$, for $i \le j - 1$. This particular sequence starts at $n = 0$ with a Gaussian distribution whose components are independent with variance 10, and as n grows, so do the correlations among the p coordinates. We implement SMC with adaptive scaling and the ideal (limiting) SMC algorithm that uses perfect scaling for two cases: $p \in \{5, 10\}$. Here, for $n \ge 1$ we have $G_n(x) = \exp\{-\frac{1}{2}\langle x, (\Gamma_{n+1}^{-1} - \Gamma_n^{-1})x\rangle\}$; for this example, (A1)–(A2) can be verified by using (A3)–(A4) and Proposition 2.1.

For $n = 0, \ldots, 50$ we run both the adaptive SMC and the limiting algorithms and compute an estimate of $N \times \mathrm{E}([\eta_n^N(\varphi) - \eta_n(\varphi)]^2)$ for $\varphi(x) = x_1$ and $\varphi(x) = x_1^2$ as well as an estimate of the variance of the normalizing constant, $N \times \mathrm{E}([\gamma_n^N(1) - \gamma_n(1)]^2)$. The estimates are computed in each case by averaging over 500 independent replications of the SMC algorithm. For large N this quantity is an approximation of the asymptotic variance in the CLT described in Theorem 2.2. In Figure 1 we present the results using $N = 10^4$ and $N = 10^6$. In

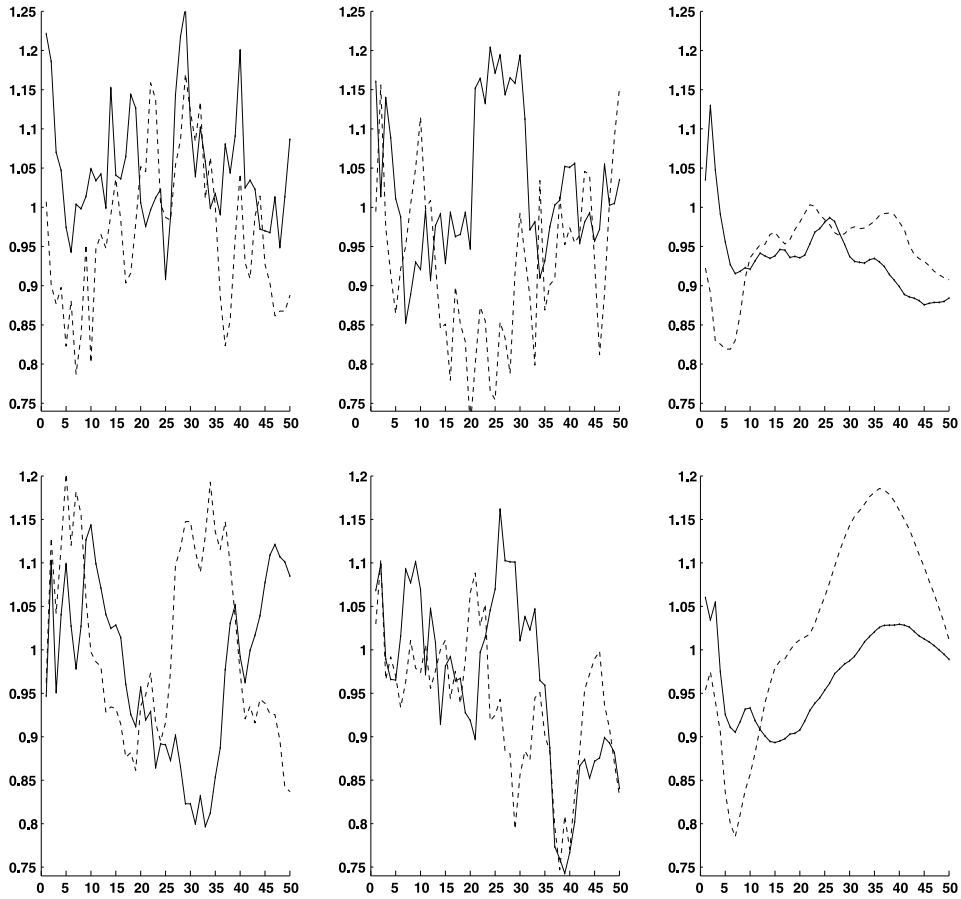

FIG. 1. *Ratio of estimated asymptotic variances of adaptive to limiting algorithms against n. We are estimating the asymptotic variance for the estimate of the first moment* (*left panels*), *second moment* (*middle panels*) *both in the first coordinate of x and the normalising constant* (*right panels*). *Top panels are for* $p = 5$ *and bottom for* $p = 10$. *In the simulations we used* $N = 10^4$ (*dotted line*), $N = 10^6$ (*solid lines*) *and* 500 *independent runs*.

each panel of Figure 1 we show the ratio of the estimated asymptotic variance of the adaptive algorithm over the one of the limiting algorithm against n. The top panels are for $p = 5$ and the lower ones for $p = 10$. The results show that the adaptive and limiting algorithms behave very similarly in terms of the Monte Carlo variance. This is in agreement with the theoretical result in Theorem 2.3. In addition, it appears that the increase in dimension does not lead to a loss of efficiency in estimation. Note that when $p = 5$ the adaptive algorithm has to estimate $d = 15$ quantities per time step, but this grows to $d = 55$ for $p = 10$. We believe this is due to using a very high number of particles compared to d and will be investigated later on in the article.

4.2. *Adaptive SMC in high dimensions.* In a large number of statistical problems, inference involves calibrating a posterior defined as a change of measure from a Gaussian prior $\eta_0 = \mathrm{N}(0, \mathcal{C})$ on infinite-dimensional function spaces; see, for example, [3, 9]. It is convenient to work with the Karhunen–Loéve representation of η_0. That is, given the eigen-decomposition $\{\lambda_k^2, \psi_k\}_{k\in\mathcal{K}}$ of $\mathcal{C}$, for some relevant set $\mathcal{K}$, we expand the function-valued parameter of interest u as $u = \sum_k u_k \psi_k$, with $u_k = \langle u, \psi_k \rangle$, so that a priori under η_0 we have $u_k \sim \mathrm{N}(0, \lambda_k^2)$ independently over k. The arriving observations can be informative for a high number of u_k's, and the task is to infer their posterior distribution. Within this high-dimensional setting, we look at a Bayesian parameter inference problem (as in Section 2.3.1) involving the 2D Navier–Stokes partial differential equation (PDE). The PDE describes the evolution in time/space of a velocity field, say $v(x,t) \in \mathbb{R}^2$, with $t \geq 0$ denoting time and x denoting space with $x \in \mathbb{T} = [0, 2\pi) \times [0, 2\pi)$, a 2D-torus. The parameter of interest $u \in E$ is the initial condition $u = v(\cdot, 0)$ of the PDE. The field is observed at discrete times and locations. We apply an SMC sampler that uses adaptive tempering and scaling. The SMC method and model are described in detail in [22]. Here we focus on the estimation of the normalizing constant, not considered in [22], and highlight the algorithmic challenges and usefulness of the adaptive SMC methodology when applied in high dimensions. This motivates some theoretical results presented in Section 4.3 where the stability properties of SMC are investigated in a particular scenario when the dimension d of the adaptive statistic is large.

In more detail, following [22], we choose the prior covariance operator as $\beta^2 \Delta^{-\alpha}$, for hyper-parameters $\beta > 0$, $\alpha > 1$ and the Laplacian Δ. This gives the representation

$$\eta_0 = \bigotimes_{k\in\mathbb{Z}^2\setminus\{0,0\}} \mathrm{N}(0, \beta^2 |k|^{-2\alpha}). \tag{4.1}$$

Each $k = (k_1, k_2)$ can be thought of as a bivariate frequency in a Fourier expansion. Data correspond to measurements $y_{j,m} = v(x_m, t_j) + \varepsilon \zeta_{j,m}$ at times $t_j = j\delta$ for $\delta > 0$ and locations on a fixed grid $(x_1, \ldots, x_M) \in \mathbb{T}$, for an i.i.d. sequence $\zeta_{j,m} \sim \mathrm{N}(0, I_2)$ and some known $\varepsilon > 0$. Assuming that $\Psi : E \times [0, \infty) \to E$ denotes the semigroup solution operator for the Navier–Stokes PDE, so that $v = v(x,t) = \Psi(u,t)$, the likelihood of each observation is

$$p(y_{j,m}|u) = \frac{1}{2\pi\varepsilon^2} \exp\left\{ -\frac{1}{2\varepsilon^2} \left\| y_{j,m} - \left[\Psi(u, t_j)\right](x_m) \right\|^2 \right\}. \tag{4.2}$$

Let $n = (j-1)M + m$. We will use SMC to sample from the sequence of laws $\{\eta_n\}_{n=0}^{M\times T}$ defined as

$$\frac{d\eta_n}{d\eta_0}(u) = \frac{1}{Z_n} \prod_{l=0}^{j-1} \prod_{j=1}^{m} p(y_{j,m}|u), \tag{4.3}$$

where Z_n denotes the unknown normalization constant. In this context $G_n = d\eta_n/d\eta_{n-1}$ for $n \geq 1$, and $M_{n,\cdot}$ is the η_n-invariant MCMC mutation steps used for propagating the N-particles system.

A priori, the Fourier coefficients u_k have known scales, but a posteriori, they can have widely varying unknown scales with the posterior being more informative for the low frequencies than the high ones. This information is contained in the particle population and can be used to construct effective proposals. Indeed, given the current particle approximation $\{u^i\}_{i=1}^N$ of η_n, we estimate the marginal mean and covariance μ_k^N, Σ_k^N for frequency k

$$\mu_k^N = \frac{1}{N}\sum_{i=1}^N u_k^i; \qquad \Sigma_k^N = \frac{1}{N-1}\sum_{i=1}^N (u_k^i - \mu_k^N) \otimes (u_k^i - \mu_k^N).$$

Then, for a current position $u = \sum u_k \psi_k$, an MCMC proposal $\widetilde{u} = \sum \widetilde{u}_k \psi_k$ can be defined as

$$\widetilde{u}_k = \mu_k^N + \rho(u_k - \mu_k^N) + (1-\rho^2)^{1/2}\mathrm{N}(0, \Sigma_k^N),$$

for a global scaling parameter $\rho > 0$ and Gaussian noise $\mathrm{N}(0, \Sigma_k^N)$ independent over k. This proposal is accepted with the relevant Metropolis–Hastings ratio, given in [22]. The challenge here is that the number of important frequencies k to be tuned can be large; for example, the actual total number of frequencies when truncating the expansion at $k_{\max} = 32$ is 4096. Kantas, Beskos and Jasra [22] suggest using adaptation only on a "window" $\{k \in \mathbb{Z}^2 \setminus \{0,0\} : \max(k_1, k_2) \leq K\}$, for a user-specified threshold $K \geq 1$. For higher frequencies, only the information contained in the prior distribution is used; thus we set $\mu_k = 0$ and $\Sigma_k^N = |k|^{-2\alpha} I_2$. The full algorithm also uses adaptive tempering as in Section 3, between every pair η_{n-1}, η_n. We found this to be important for avoiding weight degeneracy and getting a stable algorithm. As in Section 3, the temperatures are determined on the fly, according to a minimum requirement for ESS (we choose $\alpha = \frac{1}{3}$).

We present some results on estimating the normalization constant with the adaptive SMC algorithm when $N = 500$ and $K = 7$. Figure 2 shows a plot of an estimate of the variance of Z_n^N/Z_n against n, where Z_n^N is the N-particle's approximation of Z_n. In this complex setting, the numerical results seem to confirm the theoretical asymptotic results of Theorem 2.3, and the estimated asymptotic variance seems to grow linearly with n, as one would expect for the ideal SMC algorithm that uses the correct (constant) variances (see [4]). Notice that the asymptotic behavior predicted in Theorem 2.3 is likely to hold in far more general contexts, as in this application, for instance, the regularity conditions of the theorem are not satisfied.

4.3. *Algorithmic stability in large-scale adaptation.* In the Navier–Stokes example implementation we use $K = 7$, so $d \approx 2 \times (2K+1)^2 = 450$. In this case, and potentially in other scenarios, it is of interest to investigate the effect of a large

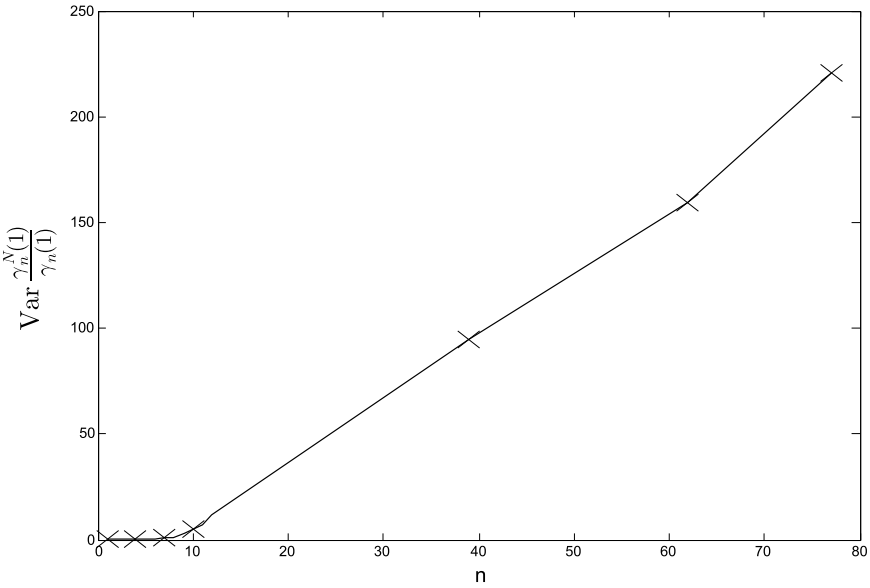

FIG. 2. *Estimated variance for the estimate of the normalizing constant of adaptive SMC. The "true" normalizing constant is estimated from* 1000 *independent runs with* $N = 500$ *and the relative variance is estimated when* $N = 500$ *over* 500 *independent runs. The crosses are the estimated values of the relative variance.*

d on the performance of the SMC method. We make a first modest attempt to shed some light on this issue via a very simple modeling structure motivated by the Navier–Stokes example and allow for some simple calculations.

For each $n \geq 1$ we assume a product form Gaussian target on $E_n = \mathbb{R}^\infty$,

$$\eta_n = \bigotimes_{j=1}^{\infty} \mathrm{N}(0, \sigma_j^2),$$

for variances $\{\sigma_j^2\}_{j=1}^\infty$ that do not depend on $n \geq 1$. Thus the weights $G_n(x)$ are assumed small enough to be irrelevant for the study of the influence of the dimension d; we have $G_n(x) \equiv 1$. It is assumed that the SMC has worked well up to time $(n-1)$, producing i.i.d. samples $\{x_{n-1}^i\}_{i=1}^N$ from η_{n-1}. For the mutation step, we consider an adaptive kernel $M_{n,\xi}$ preserving η_n that proposes, when the current position is $x \in \mathbb{R}^\infty$, a new position $\widetilde{x} \in \mathbb{R}^\infty$ distributed as

$$\begin{aligned} \widetilde{x}_j &= \rho x_j + \bigl(1-\rho^2\bigr)^{1/2} \mathrm{N}\bigl(0, \widehat{\sigma}_j^2\bigr) \qquad \text{for } 1 \leq j \leq d, \\ \widetilde{x}_j &= \rho x_j + \bigl(1-\rho^2\bigr)^{1/2} \mathrm{N}\bigl(0, \sigma_j^2\bigr) \qquad \text{for } j \geq d+1, \end{aligned} \tag{4.4}$$

for $\widehat{\sigma}_j^2 := (1/N)\sum_{i=1}^N \{x_{n-1,j}^i\}^2$; thus we adapt $\xi_n(x) = (x_1^2, \ldots, x_d^2)$. The d first coordinates are adapted to the estimated marginal variance while the ideal variance

is used for the rest. We will investigate the effect of the amount of adaptation on the accuracy of $\eta_n^N(\varphi)$ for a bounded φ that only depends on the $(d+1)$th coordinate, $\varphi(x) = \varphi(x_{d+1})$. In this simplistic scenario the proposal for the ideal kernel $M_{n,\eta_{n-1}(\xi_n)}$ preserves η_n and is always accepted. Thus any deviation from a $\mathcal{O}(N^{-1/2})$ rate of convergence for the estimator $\eta_n^N(\varphi)$ will be due to the effect of the adaptation.

Following the proof of Theorem 2.1 we use the decomposition

$$[\eta_n^N - \eta_n](\varphi) = A(N) + B_1(N) + B_2(N),$$

where $A(N) = [\eta_n^N - \Phi_{n,N}(\eta_{n-1}^N)](\varphi)$, $B_1(N) = \eta_{n-1}^N[Q_{n,N} - Q_n](\varphi)$, $B_2(N) = [\eta_{n-1}^N - \eta_{n-1}](Q_n\varphi)$. Denoting by $\|\cdot\|_2$ the L_2-norm of random variables and conditioning upon $\mathcal{F}_{n-1}^N$, we have that $\|A(N)\|_2^2 = \frac{1}{N}\mathrm{E}[\mathrm{Var}[\varphi(x_n^1)|\mathcal{F}_{n-1}^N]] = \mathcal{O}(\frac{1}{N})$. For $B_2(N)$ one can notice that $Q_n(\varphi)$ is a bounded mapping from $\mathbb{R}^\infty$ to $\mathbb{R}$, thus $\|B_2(N)\|_2^2 = \frac{1}{N}\mathrm{Var}_{\eta_{n-1}}[Q_n(\varphi)] = \mathcal{O}(\frac{1}{N})$. The critical term with regards to the effect of the dimension d on the magnitude of the difference $[\eta_n^N - \eta_n](\varphi)$ is $B_1(N)$. An approach similar to equation (2.17) in the proof of Theorem 2.2 yields

$$\begin{aligned} B_1(N) &= \eta_{n-1}^N[Q_{n,N} - Q_n](\varphi) = \eta_{n-1}^N\big([M_{n,N} - M_n](\varphi)\big) \\ &= \eta_{n-1}^N[\partial_\xi M_n\varphi] \cdot \big[\eta_{n-1}^N - \eta_{n-1}\big](\xi_n) + R =: \widetilde{B}_1(N) + R, \end{aligned}$$

for a random variable term R containing second (or higher) order derivatives in the Taylor expansion. Controlling the R term poses enormous technical challenges, and we restrict our analysis to the first term in the Taylor expansion $\widetilde{B}_1(N)$, as we feel this will give an impression for the choice of N as a function of d. The proof of the following result is provided in the Appendix.

PROPOSITION 4.1. *The term $\widetilde{B}_1(N)$ satisfies*

$$\|\widetilde{B}_1(N)\|_2 = \mathcal{O}\Big(\frac{\sqrt{d}}{N}\Big) + \mathcal{O}\Big(\frac{d}{N^{3/2}}\Big).$$

Proposition 4.1, combined with the earlier results, suggests that in a high-dimensional setting with $d \gg 1$, it is reasonable to choose N of order $\mathcal{O}(d)$, yielding a mean squared error of order $\mathcal{O}(1/d)$. Even if this choice of N should be thought of as a minimum requirement for the complete sequential method, it could potentially explain the fairly accurate SMC estimates of the marginal expectation obtained in the Navier–Stokes example when $N = 500$ and $d \approx 500$.

5. Summary. This article studies the asymptotic properties of a class of adaptive SMC algorithms; weak law of large numbers and a central limit theorems are established in several practical settings. There are several potential directions for extension to the work in this article. One could relax the boundedness assumptions used in the paper; our proof technique, also used in [8], is particularly amenable to this. Also, one can extend the analysis to the context of adaptive resampling.

APPENDIX A: PROOF OF LEMMA 3.1

The case where the random variable $V(X)$ is almost surely constant is trivial. Let us thus consider the case where $V(X)$ is not almost surely constant and prove that $\lambda \mapsto \mathrm{ESS}(\eta, e^{-\lambda V})$ is continuous and strictly decreasing. Let X and Y be two independent random variables distributed according to η. The dominated convergence theorem implies that the function $\lambda \mapsto \mathrm{ESS}(\eta, e^{-\lambda V})$ is continuous, with a continuous derivative. Standard manipulations show that the derivative is strictly negative if $\eta(Ve^{-\lambda V})\eta(e^{-2\lambda V}) > \eta(e^{-\lambda V})\eta(Ve^{-2\lambda V})$, which is equivalent to the condition

$$\mathrm{E}\big[e^{-\lambda\{V(X)+V(Y)\}} \times \big\{V(X) - V(Y)\big\} \times \big\{e^{-\lambda V(X)} - e^{-\lambda V(Y)}\big\}\big] < 0. \tag{A.1}$$

This last condition is satisfied since for any $x, y \in \mathbb{R}$ and any $\lambda > 0$, we have the inequality $\{V(x) - V(y)\}\{e^{-\lambda V(x)} - e^{-\lambda V(y)}\} < 0$, with strict inequality for $V(x) \neq V(y)$, and we assume that the random variable $V(X)$ is not almost surely constant.

APPENDIX B: PROOF OF PROPOSITION 4.1

First of all, notice that without loss of generality we can assume that σ_j^2 is a constant. We have that

$$\begin{aligned}\widetilde{B}_1(N) &= \frac{\sqrt{d}}{N} \times \sum_{j=1}^{d} \Bigg\{ \frac{\sum_{i=1}^{N} \partial_{\xi_j} M_{n,\xi}(\varphi)(x_{n-1}^i)|_{\xi=\eta_{n-1}(\xi_n)}}{\sqrt{N}} \\ &\qquad \times \sqrt{N}\big(\eta_{n-1}^N - \eta_{n-1}\big)(\xi_{n,j})\Bigg\} \Big/ \sqrt{d} \\ &\equiv \frac{\sqrt{d}}{N} \times \sum_{j=1}^{d} \big[\sqrt{N}\eta_{n-1}^N(\overline{\Xi}_{n,j}) \cdot \sqrt{N}\eta_{n-1}^N(\overline{\xi}_{n,j})\big]/\sqrt{d},\end{aligned} \tag{B.1}$$

where we have set $\overline{\Xi}_{n,j}(x) = \partial_{\xi_j} M_{n,\xi}(\varphi)(x)|_{\xi=\eta_{n-1}(\xi_n)}$ and $\overline{\xi}_{n,j}(x) = \xi_{n,j}(x) - \eta_{n-1}(\xi_{n,j})$. Clearly, the expectation of the latter variable over η_{n-1} is zero, but the same is also true for the former. Initially, we will focus on the term $\overline{\Xi}_{n,j}(x)$ as it has some structure which will be exploited in subsequent calculations. Indeed, considering $M_{n,\xi_j}(\varphi)(x)$, for an arbitrary ξ_j and the rest ξ_k, $k \neq j$, at their limiting "correct" values, we have that

$$\begin{aligned}M_{n,\xi_j}(\varphi)(x) &= \mathrm{E}\big[\varphi\big(x'_{d+1}\big)|x\big] \\ &= \varphi(x_{d+1}) + \mathrm{E}\big[a(x_j, \xi_j, Z_j)|x_j\big]\Delta\varphi(x_{d+1}),\end{aligned} \tag{B.2}$$

where we have set $\Delta\varphi(x_{d+1}) = \mathrm{E}[\varphi(x'_{d+1}) - \varphi(x_{d+1})|x_{d+1}]$; x'_{d+1} denotes the Metropolis–Hastings proposal for the $(d+1)$th co-ordinate as specified in (4.4); $a(x_j, \xi_j, Z_j)$ denotes the Metropolis–Hastings acceptance probability, which depends only on the current position x_j, the (arbitrary) scaling choice ξ_j and the

noise $Z_j \sim \mathrm{N}(0,1)$ for simulating the proposal for the jth coordinate, assuming a scaling ξ_j (i.e., $x'_j = \rho x_j + \sqrt{1-\rho^2}\xi_j^{1/2} Z_j$). We will give the explicit formula for $a(\cdot)$ below. Notice that due to the proposal for x_{d+1} preserving the target marginally at the $(d+1)$th coordinate, we have that $\mathrm{E}_{\eta_{n-1}}[\Delta\varphi(x_{d+1})] = 0$. Recall $\overline{\Xi}_{n,j}(x) = \partial_{\xi_j} M_{n,\xi_j}(\varphi)(x)|_{\xi_j=\eta_{n-1}(\xi_{n,j})}$. Thus to check for the differentiability of $\xi_j \mapsto \mathrm{E}[a(x_j,\xi_j,Z_j)|x_j]$ we can only resort to analytical calculations, starting with the fact that (after some algebraic manipulations)

$$\begin{aligned} &a(x_j,\xi_j,Z_j) \\ &\quad = \min\bigl(1, \exp\bigl\{-\tfrac{1}{2}\bigl(\xi_j^{-1}-\sigma_j^{-2}\bigr)\bigl(x_j^2 - \bigl\{\rho x_j + \sqrt{1-\rho^2}\xi_j^{1/2} Z_j\bigr\}^2\bigr)\bigr\}\bigr). \end{aligned}$$

After several analytic calculations (which are omitted for brevity) we can integrate out variable Z_j and find that: (i) the derivative $D(x_j,\eta_{n-1}(\xi_{n,j})) = \partial_{\xi_j}\mathrm{E}[a(x_j,\xi_j,Z_j)|x_j]|_{\xi_j=\eta_{n-1}(\xi_{n,j})}$ exists; (ii) $D(x_j,\eta_{n-1}(\xi_{n,j}))$, with $x_j \sim \mathrm{N}(0,\sigma_j^2)$, has a finite second moment. Thus, continuing from (B.2), we have

$$\overline{\Xi}_{n,j}(x) = \partial_{\xi_j} M_{n,\xi_j}(\varphi)(x)|_{\xi_j=\eta_{n-1}(\xi_{n,j})} = D\bigl(x_j,\eta_{n-1}(\xi_{n,j})\bigr)\Delta\varphi(x_{d+1}). \tag{B.3}$$

The factorization in (B.3) will be exploited in the remaining calculations.

Continuing from (B.1), we now have that

$$\begin{aligned} &\left\| \frac{N}{\sqrt{d}} \widetilde{B}_1(N) \right\|_2^2 \\ &\quad = \frac{1}{d}\sum_{j=1}^{d} N^2 \mathrm{E}\bigl[\bigl\{\eta_{n-1}^N(\overline{\Xi}_{n,j})\bigr\}^2\bigl\{\eta^N(\overline{\xi}_{n,j})\bigr\}^2\bigr] \\ &\qquad + \frac{1}{d}\sum_{\substack{j,k=1,2,\ldots,d \\ j\neq k}} N^2 \mathrm{E}\bigl[\eta_{n-1}^N(\overline{\Xi}_{n,j})\eta_{n-1}^N(\overline{\xi}_{n,j})\eta_{n-1}^N(\overline{\Xi}_{n,k})\eta_{n-1}^N(\overline{\xi}_{n,k})\bigr] \\ &\quad =: T_1 + T_2. \end{aligned} \tag{B.4}$$

The following zero-expectations obtained for terms involved in T_1, T_2 are a direct consequence of the fact that $\overline{\xi}_{n,j}(x)$ only depends on x_j and has zero expectation under η_{n-1}, and that $\overline{\Xi}_{n,j}(x)$ only depends on x_j, x_{d+1} through the product form in (B.3) with the x_{d+1}-term having zero-expectation; critically, recall that particles $x_{n-1,j}^i$ are independent over both i, j. Focusing on the T_1-term and the expectation $\mathrm{E}[\{\eta_{n-1}^N(\overline{\Xi}_{n,j})\}^2\{\eta_{n-1}^N(\overline{\xi}_{n,j})\}^2]$, we note that all 4-way product terms arising after replacing η_{n-1}^N with its sum-expression will have expectation 0, except for those that involve cross-products of the form $\{\overline{\Xi}_{n,j}(x_{n-1}^i)\}^2 \times \{\overline{\xi}_{n,j}(x_{n-1}^{i'})\}^2$, thus

$$T_1 = \frac{1}{d}\sum_{j=1}^{d} N^2 \cdot \frac{1}{N^4} \cdot \mathcal{O}\bigl(N^2\bigr) = \mathcal{O}(1). \tag{B.5}$$

Moving on to T_2, notice that all 4-way products in $\mathrm{E}[\eta_{n-1}^N(\overline{\Xi}_{n,j})\eta_{n-1}^N(\overline{\xi}_{n,j}) \times \eta_{n-1}^N(\overline{\Xi}_{n,k})\eta_{n-1}^N(\overline{\xi}_{n,k})]$ have expectation 0, except for the products involving the same particles, that is, except for the terms $\overline{\Xi}_{n,j}(x_{n-1}^i)\overline{\xi}_{n,j}(x_{n-1}^i)\overline{\Xi}_{n,k}(x_{n-1}^i) \times \overline{\xi}_{n,k}(x_{n-1}^i)$. Thus we have that

$$T_2 = \frac{1}{d} \sum_{j,k=1,\, j\neq k}^{d} N^2 \cdot \frac{1}{N^4} \cdot \mathcal{O}(N) = \mathcal{O}\left(\frac{d}{N}\right).$$

Thus overall we have that

$$\|\widetilde{B}_1(N)\|_2 = \mathcal{O}\left(\frac{\sqrt{d}}{N}\right) + \mathcal{O}\left(\frac{d}{N^{3/2}}\right). \tag{B.6}$$

Results (B.5), (B.6), used within (B.4) complete the proof.

Acknowledgments. We thank the referees and Associate Editor for their comments, which lead to substantial improvements in the article. We also thank Arnaud Doucet for conversations on this work.

REFERENCES

[1] ANDRIEU, C. and MOULINES, É. (2006). On the ergodicity properties of some adaptive MCMC algorithms. *Ann. Appl. Probab.* **16** 1462–1505. MR2260070

[2] BESKOS, A., CRISAN, D. and JASRA, A. (2014). On the stability of sequential Monte Carlo methods in high dimensions. *Ann. Appl. Probab.* **24** 1396–1445. MR3211000

[3] BESKOS, A., ROBERTS, G. and STUART, A. (2009). Optimal scalings for local Metropolis–Hastings chains on nonproduct targets in high dimensions. *Ann. Appl. Probab.* **19** 863–898. MR2537193

[4] 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

[5] CÉROU, F. and GUYADER, A. (2014). Fluctuation analysis of adaptive multilevel splitting. Technical report, INRIA.

[6] CHAN, H. P. and LAI, T. L. (2013). A general theory of particle filters in hidden Markov models and some applications. *Ann. Statist.* **41** 2877–2904. MR3161451

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

[8] CHOPIN, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.* **32** 2385–2411. MR2153989

[9] COTTER, S. L., ROBERTS, G. O., STUART, A. M. and WHITE, D. (2013). MCMC methods for functions: Modifying old algorithms to make them faster. *Statist. Sci.* **28** 424–446. MR3135540

[10] CRISAN, D. and DOUCET, A. (2000). Convergence of sequential Monte Carlo methods. Technical report, CUED/F-INFENG/, Cambridge Univ.

[11] DEL MORAL, P. (2004). *Feynman–Kac Formulae*: *Genealogical and Interacting Particle Systems with Applications*. Springer, New York. MR2044973

[12] DEL MORAL, P. (2013). *Mean Field Simulation for Monte Carlo Integration*. Chapman & Hall, London.

[13] DEL MORAL, P., DOUCET, A. and JASRA, A. (2006). Sequential Monte Carlo samplers. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 411–436. MR2278333

[14] DEL MORAL, P., DOUCET, A. and JASRA, A. (2012). On adaptive resampling strategies for sequential Monte Carlo methods. *Bernoulli* **18** 252–278. MR2888706

[15] DEL MORAL, P., DOUCET, A. and JASRA, A. (2012). An adaptive sequential Monte Carlo method for approximate Bayesian computation. *Stat. Comput.* **22** 1009–1020. MR2950081

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

[17] DOUC, R., MOULINES, E. and OLSSON, J. (2014). Long-term stability of sequential Monte Carlo methods under verifiable conditions. *Ann. Appl. Probab.* **24** 1767–1802. MR3226163

[18] DOUCET, A. and JOHANSEN, A. (2011). A tutorial on particle filtering and smoothing: Fifteen years later. In *Handbook of Nonlinear Filtering* (D. Crisan and B. Rozovsky, eds.). Oxford Univ. Press, Oxford.

[19] GELMAN, A. and MENG, X.-L. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statist. Sci.* **13** 163–185. MR1647507

[20] GIRAUD, F. and DEL MORAL, P. (2015). Non-asymptotic analysis of adaptive and annealed Feynman–Kac particle models. *Bernoulli*. To appear.

[21] JASRA, A., STEPHENS, D. A., DOUCET, A. and TSAGARIS, T. (2011). Inference for Lévy-driven stochastic volatility models via adaptive sequential Monte Carlo. *Scand. J. Stat.* **38** 1–22. MR2760137

[22] KANTAS, N., BESKOS, A. and JASRA, A. (2014). Sequential Monte Carlo methods for high-dimensional inverse problems: A case study for the Navier–Stokes equations. *SIAM/ASA J. Uncertain. Quantificat.* **2** 464–489. MR3283917

[23] SCHÄFER, C. and CHOPIN, N. (2013). Sequential Monte Carlo on large binary sampling spaces. *Stat. Comput.* **23** 163–184. MR3016936

A. BESKOS
DEPARTMENT OF STATISTICAL SCIENCE
UNIVERSITY COLLEGE LONDON
1-19 TORRINGTON PLACE
LONDON, WC1E 7HB
UNITED KINGDOM
E-MAIL: a.beskos@ucl.ac.uk
URL: http://www.homepages.ucl.ac.uk/~ucakabe/

A. JASRA
A. THIERY
DEPARTMENT OF STATISTICS
AND APPLIED PROBABILITY
NATIONAL UNIVERSITY OF SINGAPORE
6 SCIENCE DRIVE 2
SINGAPORE, 117546
SINGAPORE
E-MAIL: staja@nus.edu.sg
a.h.thiery@nus.edu.sg
URL: https://sites.google.com/site/ajayjasra0/home
http://www.normalesup.org/~athiery/

N. KANTAS
DEPARTMENT OF MATHEMATICS
IMPERIAL COLLEGE LONDON
180 QUEENS GATE
LONDON, SW7 2AZ
UNITED KINGDOM
E-MAIL: n.kantas@ic.ac.uk
URL: http://www.imperial.ac.uk/people/n.kantas