On the Convergence of Adaptive Sequential Monte Carlo Methods

BY ALEXANDROS BESKOS[1], AJAY JASRA[1], NIKOLAS KANTAS[2] & ALEXANDRE H. THIÉRY[1]

[1]Department of Statistics & Applied Probability, National University of Singapore, Singapore, 117546, SG.

E-Mail: staba@nus.edu.sg, staja@nus.edu.sg, a.h.thiery@nus.edu.sg

[2]Department of Mathematics, Imperial College London, London, SW7 2AZ, UK.

E-Mail: n.kantas@ic.ac.uk

Abstract

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 [5, 17, 20]. 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 [17, 20]. We establish that for a general class of adaptive SMC algorithms [5] the asymptotic variance of the estimators from the adaptive SMC method is *identical* to a so-called 'perfect' 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.

Keywords: Adaptive SMC, Central Limit Theorem, Markov chain Monte Carlo.

1 Introduction

SMC methods are amongst the most widely used computational techniques in statistics, engineering, physics, finance and many other disciplines; see [14] for a recent overview. They are designed to approximate a sequence $\{\eta_n\}_{n\geq 0}$ of probability distributions of increasing dimension or complexity. The method uses $N \geq 1$ weighted samples, or particles, generated in parallel and propagated via Markov kernels and resampling methods. The method has accuracy which increases as the number of particles grows and is typically asymptotically exact. Standard SMC methodology is by now very well understood with regards to its convergence properties and several consistency results have been proved [6, 9]. SMC methods have also recently been proved to be stable in certain high-dimensional contexts [2].

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 [5, 12, 17, 20] 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 or even to *determine* the next probability distribution in the sequence; this is often achieved by using the expectation $\eta_n^N(\xi_{n+1})$ of a summary statistic ξ_{n+1} 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; we expand upon this point below. Such approaches are expected to lead to algorithms that are more efficient than their 'non-adaptive' counter-parts. Critically, such ideas also deliver more automated algorithms by reducing the number of user-specified tuning parameters.

Whilst 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 [8, 17]. Proof sketches are given in [12] with some more realistic but limited analysis in [16]. We are not aware of any other asymptotic analysis of these particular class of algorithms in the literature. Contrary to adaptive MCMC algorithms, we show in this article that it is reasonable to expect most adaptive SMC methods to be asymptotically correct.

1.1 Results and Structure

This article explores two distinct directions. In the first part, an asymptotic analysis of a class of SMC methods with adaptive Markov kernels and weights is carried out. The second part of the article looks at the case where an additional layer of randomness is taken into account through an adaptive tempering procedure. A weak law of large numbers (WLLN) and a central limit theorem (CLT) relevant to each situation are proved. In both cases 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 N-particle SMC 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 the first part of the paper, for each $n \geq 1$ we consider parametric families, 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 *practical* 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 $G_{n,\eta_{n-1}^N(\xi_n)}$ and the Markov kernel $M_{n,\eta_{n-1}^N(\xi_n)}$. This is a substitute for the *perfect* algorithm (as also used by [16] and which cannot be implemented) which employs the Markov kernel $M_{n,\eta_{n-1}(\xi_n)}$ and weight function $G_{n,\eta_{n-1}(\xi_n)}$. We prove a WLLN and a CLT for both the approximation of the probability distribution η_n and its normalising constant. This set-up is relevant, for example, in the context of sequential Bayesian parameter inference [5, 18] when $\{\eta_n\}_{n\geq 0}$ is a sequence of posterior distributions that corresponds to increasing amount of data. The Markov kernel $M_{n,\eta_{n-1}^N(\xi_n)}$ is user-specified and its role is to efficiently move the particles within the state space. In many situations the Markov kernel $M_{n,\eta_{n-1}^N(\xi_n)}$ is constructed so that it leaves the distribution η_n invariant; a random walk Metropolis kernel that uses the estimated covariance structure of η_{n-1}^N for scaling its jump proposals is a popular choice. The case when there is also a tuned parameter in the weight function $G_{n,\eta_{n-1}^N(\xi_n)}$ is relevant to particle filters [14], as described in Section 3.2.

The second part of this article investigates an adaptive 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 combination thereof. It is a standard approach to introduce a bridging sequence of distributions $\{\eta_n\}_{n=0}^{n=n_*}$ between a distribution η_0 that is easy to sample from and the distribution of interest $\eta_{n_*} \equiv \pi$. In accordance with the simulated annealing literature, the probability distribution of interest is written as $\pi(dx) = Z^{-1}\, e^{-\beta_*\, V(x)}\, m(dx)$ for a potential V, temperature parameter $\beta_* \in \mathbb{R}$, dominating measure $m(dx)$ and normalisation constant Z; the bridging sequence of distributions is constructed by introducing a ladder of temperature parameters $\beta_0 \leq \beta_1 \leq \cdots \leq \beta_{n_*} =: \beta_*$ and setting $\eta_n(dx) = Z(\beta_n)^{-1}\, e^{-\beta_n\, V(x)}\, m(dx)$ for a normalisation constant $Z(\beta_n)$. The choice of the bridging sequence of distributions is an important and complex problem, see e.g. [15]. 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' [17, 20], thus obtaining a random increasing sequence of temperature parameters $\{\beta_n^N\}_{n\geq 0}$. In this article, we adopt the following strategy: assuming a particle approximation $\eta_{n-1}^N = (1/N)\sum_{i=1}^N \delta_{x_{n-1}^i}$ with temperature parameter β_{n-1}^N, the particles are assigned weights proportional to $e^{-(\beta_n^N - \beta_{n-1}^N)\, V(x_{n-1}^i)}$ to represent the next distribution in the sequence; the choice of β_n^N is determined from the particle collection $\{x_{n-1}^i\}_{i=1}^N$ by ensuring a minimum effective sample size (ESS) (it is described later on, why this might be a sensible choice). This can efficiently be implemented using a bisection method; see e.g. [17]. We prove a WLLN and a CLT for both the approximation of the probability distribution η_n and the estimates of the normalising constants $Z(\beta_n)$.

One of the contributions of the article is the proof that the asymptotic variance in the CLT, for some algorithms in the first part of the paper, is *identical* to the one of the 'perfect' SMC algorithm using the ideal kernels. One consequence of this effect is that if the asymptotic variance associated to the (relative) normalizing constant estimate increases linearly with respect to time (see e.g. [4]), then so does the asymptotic variance for the adaptive algorithm. We present numerical results on a complex high-dimensional posterior distribution associated with the Navier-Stokes model (as in e.g. [18]), where adapting the proposal kernels over hundreds of different directions is critical for the efficiency of the algorithm. Whilst our theoretical result (with regards to the asymptotic variance) only holds for the case where one adapts the proposal kernel, the numerical application will involve much more advanced adaptation procedures. These experiments provide some evidence that our theory could be relevant in more general scenarios.

This article is structured as follows. In Section 2 the adaptive SMC algorithm is introduced and the associated notations are detailed. In Section 3 we provide some motivating examples for the use of adaptive SMC. In Section 4 we study the asymptotic properties of a class of SMC algorithms with adaptive Markov kernels and weights. In Section 5, we extend our analysis to the case where an adaptive tempering scheme is taken into account. In each situation, we prove a WLLN and a CLT. In Section 6, we verify that our assumptions hold when using the adaptive SMC algorithm in a real scenario. In addition, we provide numerical results associated to the Navier-Stokes model and some theoretical insights associated to the effect of the dimension of the statistic which is adapted. The article is concluded in Section 7 with a discussion of future work. The appendix features a proof of one of the results in the main text.

2 Algorithm and Notations

In this section we provide the necessary notations and describe the SMC algorithm with adaptive Markov kernels and weights. The description of the adaptive tempering procedure is postponed to Section 5.

2.1 Notations 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 set $\mathcal{B}_b(E_n)$ denotes the class of bounded $\mathscr{E}_n/\mathbb{B}(\mathbb{R})$-measurable functions on E_n where $\mathbb{B}(\mathbb{R})$ 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 non-negative 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 non-negative 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

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

We will use the following notion of continuity at several places in this article.

Definition 2.1. *Let $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ be three 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*

$$\limsup_{\delta\to 0^+}\ \Big\{d_{\mathcal{Z}}\big(f(x, y), f(x, y_0)\big)\ :\ x \in \mathcal{X},\ d_{\mathcal{Y}}(y, y_0) < \delta\Big\} = 0\ . \tag{1}$$

We write $\to_{\mathbb{P}}$ and $\Rightarrow$ to denote convergence in probability and in distributions. The Kroenecker product $u \otimes v$ of two vectors $u, v \in \mathbb{R}^d$ designates the matrix $u\cdot v^\top \in \mathbb{R}^{d\times d}$; the covariance of a function $\varphi \in \mathcal{B}_b(E)^r$ with respect to a probability measure $\mu \in \mathcal{P}(E)$ is denoted by $\Sigma_\mu(\varphi) = \int_E[\varphi(x) - \mu(\varphi)] \otimes [\varphi(x) - \mu(\varphi)]\,\mu(dx)$.

2.2 SMC Algorithm

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

$$\eta_n(\varphi) := \gamma_n(\varphi_n)/\gamma_n(1) \tag{2}$$

where γ_n is the unnormalised measure on $(E_n, \mathscr{E}_n)$ given by

$$\gamma_n(\varphi) := \mathbb{E}\Big[\prod_{p=0}^{n-1} G_p(X_p)\cdot\varphi(X_n)\Big] \ . \tag{3}$$

The above expectation is under the law of a non-homogeneous Markov chain $\big\{X_n\big\}_{n\geq 0}$ with initial distribution $X_0 \sim \eta_0 \equiv \gamma_0$ and transition $\mathbb{P}\,[\,X_n \in A \mid X_{n-1} = x\,] = M_n(x, A)$ where we have used the notations

$$M_n \equiv M_{n,\eta_{n-1}(\xi_n)} \ ; \quad G_n \equiv G_{n,\eta_n(\xi_{n+1})} \ .$$

In practice, the expectations $\eta_{n-1}(\xi_n)$ of the summary statistics are not analytically tractable and it is thus impossible to simulate from the Markov chain $\{X_n\}_{n\geq 0}$ or compute the weights G_n. Nevertheless, for the purpose of analysis, we introduce the following idealized algorithm, referred to as the *perfect* SMC algorithm in the sequel, that propagates a set of $N \geq 1$ particles by sampling from the distribution

$$\mathbb{P}\big(\,d(x_0^{1:N}, x_1^{1:N}, \ldots, x_n^{1:N})\,\big) = \prod_{i=1}^{N}\eta_0(dx_0^i)\,\prod_{p=1}^{n}\prod_{i=1}^{N}\Phi_p(\eta_{p-1}^N)(dx_p^i) \tag{4}$$

where the N-particle approximation of the distribution (2) is defined as

$$\eta_n^N = \frac{1}{N}\sum_{i=1}^{N}\delta_{x_n^i} \ . \tag{5}$$

In (4), the operator $\Phi_n : \mathcal{P}(E_{n-1}) \to \mathcal{P}(E_n)$ is

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

Expression (4) is a mathematically concise way to describe a standard particle method that begins by sampling N i.i.d. particles from η_0 and, given particles $\{x_{n-1}^i\}_{i=1}^N$, performs multinomial resampling according to the unnormalised weights $G_{n-1}(x_{n-1}^i)$ before propagating the particles via the Markov kernel $M_n(x, dy)$.

The SMC algorithm that is actually simulated in practice, referred to as the *practical* SMC algorithm in the sequel, has joint law

$$\mathbb{P}\left(\,d(x_0^{1:N}, x_1^{1:N}, \ldots, x_n^{1:N})\,\right) = \prod_{i=1}^{N}\eta_0(dx_0^i)\,\prod_{p=1}^{n}\prod_{i=1}^{N}\Phi_{p,N}(\eta_{p-1}^N)(dx_p^i) \ . \tag{6}$$

The operator $\Phi_{n,N}$ approximates the ideal one, Φ_n, and is defined as

$$\Phi_{n,N}(\mu)(dy) = \frac{\mu(G_{n-1,N}\, M_{n,N})(dy)}{\mu(G_{n-1,N})} \ .$$

We have used the short-hand notations

$$M_{n,N} \equiv M_{n,\eta_{n-1}^N(\xi_n)} \ ; \qquad G_{n,N} \equiv G_{n,\eta_n^N(\xi_{n+1})} \ .$$

Throughout this article we assume that the potentials are strictly positive, $G_{n,\xi}(x) > 0$ for all $x \in E_n$ and $\xi \in \mathbb{R}^d$ so that there is no possibility that the algorithm collapses. The particle approximation of the unnormalised distribution (3) is defined as

$$\gamma_n^N(\varphi_n) = \Big\{ \prod_{p=0}^{n-1} \eta_p^N(G_{p,N}) \Big\} \, \eta_n^N(\varphi_n) \ . \tag{7}$$

It will bel useful to introduce the non-negative operator

$$Q_{n,N}(x, dy) = G_{n-1,N}(x) M_{n,N}(x, dy) \tag{8}$$

and the idealised version

$$Q_n(x, dy) = G_{n-1}(x) M_n(x, dy) \equiv G_{n-1,\eta_{n-1}(\xi_n)}(x) M_{n,\eta_{n-1}(\xi_n)}(x, dy) \ .$$

Many times we will be interested in the properties of involved operators as functions of ξ, thus we will also write

$$Q_{n,\xi}(x, dy) := G_{n-1,\xi}(x)\, M_{n,\xi}(x, dy)$$

to emphasise the dependency on the parameter $\xi \in \mathbb{R}^d$. Unless otherwise stated, the differentiation operation ∂_ξ at step n is evaluated at the limiting parameter value $\xi = \eta_{n-1}(\xi_n)$. With these definitions, one can verify that the following identities hold

$$\eta_n(\varphi_n) = \Phi_n(\eta_{n-1})(\varphi_n) = \frac{\eta_{n-1}(Q_n \varphi_n)}{\eta_{n-1}(G_{n-1})} \ ; \quad \gamma_n(\varphi_n) = \gamma_{n-1}(Q_n \varphi_n) \ . \tag{9}$$

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

$$\mathbb{E}\left[\eta_n^N(\varphi_n) \mid \mathscr{F}_{n-1}^N\right] = \Phi_{n,N}(\eta_{n-1}^N)(\varphi_n) \ ; \quad \mathbb{E}\left[\gamma_n^N(\varphi_n) \mid \mathscr{F}_{n-1}^N\right] = \gamma_{n-1}^N(Q_{n,N}\varphi_n) \ . \tag{10}$$

In the sequel, we will use the expressions $\mathbb{E}_{n-1}[\,\cdot\,]$ and $\mathrm{Var}_{n-1}[\,\cdot\,]$ to denote the conditional expectation $\mathbb{E}[\,\cdot \mid \mathscr{F}_{n-1}^N\,]$ and conditional variance $\mathrm{Var}[\,\cdot \mid \mathscr{F}_{n-1}^N\,]$ respectively.

Remark 2.1. *Our results concern multinomial resampling at each time. Extension of our analysis to adaptive resampling [11] is possible but would require many additional calculations and technicalities; this is left as a topic for future work.*

3 Motivating Examples

3.1 Sequential Bayesian Parameter Inference

Consider Bayesian inference for the parameter $x \in E$, observations $y_i \in \mathcal{Y}$ and prior measure $\eta_0(dx)$. The posterior distribution η_n after having observed $y_{1:n} \in \mathcal{Y}^{n+1}$ reads

$$\eta_n(dx) = \big(\mathbb{P}\,[\, y_{1:n} \mid x \,] \,/\, \mathbb{P}\,[\, y_{1:n} \,] \big)\, \eta_0(dx)\,.$$

The approach in [5] fits in the framework described in Section 2.2 with state spaces $E_n = E$ and potential functions $G_n(x) = \mathbb{P}\,[\, y_{n+1} \mid y_{1:n}, x \,]$. 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 unnormalised measure γ_n is defined as in (3). A popular choice consists in choosing for $M_{n,\eta_{n-1}(\xi_n)}$ a random walk Metropolis kernel reversible with respect to η_n and jump covariance structure matching the one of the distribution η_{n-1}. Under our assumptions, the analysis of Section 4 applies in this context.

Whilst such an example is quite simple it is indicative of more complex applications in the literature. Article [18] considers a state-space with dimension of about 10^4 and dimension of adapted statistic of about 500. In such a setting, pre-specifying the covariance structure of the random walk Metropolis proposals is impractical; the adaptive SMC strategy of Section 2 provides a principled framework for automatically setting this covariance structure, see also Section 6.2.

3.2 Filtering

This section illustrates the case of having 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 probability $\eta_0(u_0)\, \lambda_{\mathcal{U}}(du_0)$ is the prior distribution for the initial state of the unobserved Markov chain, $g_p(u_p, y_p)\, \lambda_{\mathcal{Y}}(dy_p)$ is the conditional observation probability at time p and $f_p(u_{p-1}, u_p)\, \lambda_{\mathcal{U}}(du_p)$ describes the dynamics of the unobserved Markov process.

A standard particle filter with proposal at time p corresponding to the Markov kernel $\mathbb{P}[U_p \in du_p \mid U_{p-1} = u_{p-1}] = m_p(u_{p-1}, u_p)\, \lambda_{\mathcal{U}}(du_p)$ has 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 $M_p(x_{p-1}, dx_p) = \delta_{x_{p-1}^{(2)}}(dx_p^{(1)})\, m_p(x_{p-1}^{(2)}, x_p^{(2)})\, \lambda_{\mathcal{U}}(dx_p^{(2)})$. The marginals of the sequence of probability distributions η_n described in Equation (2) are the standard predictors.

In practice, the choice of the proposal kernel m_n is critical to the efficiency of the SMC algorithm. In such settings, 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 adapted in Section 2. See [14] and the references therein for ideas associated to such approaches.

4 Asymptotic Results for Adaptive SMC via Summary Statistics

In this section we develop an asymptotic analysis of the class of adaptive SMC algorithm described in section 2. After first stating our assumptions in Section 4.1, we give a WLLN in Section 4.2 and a CLT in Section 4.3.

4.1 Assumptions

Our results will make use of conditions (A1-2) below. By $\mathrm{Dom}(\xi_n) \subset \mathbb{R}^d$ we denote a convex set that contains the range of the statistic $\xi_n : E_{n-1} \to \mathbb{R}^d$.

(**A1**) For each $n \geq 0$, function $(x, \xi) \mapsto G_{n,\xi}(x)$ is bounded and continuous at $\xi = \eta_n(\xi_{n+1})$ uniformly over $x \in E_n$. Statistics $\xi_{n+1} : E_n \to \mathbb{R}^d$ are bounded. For any test function $\varphi_{n+1} \in \mathcal{B}_b(E_{n+1})$ the function $(x, \xi) \mapsto Q_{n+1,\xi}\varphi_{n+1}(x)$ is bounded, continuous at $\xi = \eta_n(\xi_{n+1})$ uniformly over $x \in E_n$.

(**A2**) For each $n \geq 0$ and test function $\varphi_{n+1} \in \mathcal{B}_b(E_{n+1})$, function $(x, \xi) \mapsto \partial_\xi Q_{n+1,\xi}\varphi_{n+1}(x)$ is well defined on $E_n \times \mathrm{Dom}(\xi_{n+1})$, bounded and continuous at $\xi = \eta_n(\xi_{n+1})$ uniformly over $x \in E_n$.

Assumptions (A1-2) are reasonably weak in comparison to some assumptions used in the SMC literature, such as in [9], but are certainly not the weakest adopted for WLLN and CLTs (see e.g. [6]). The continuity assumptions in (A2) are associated to the use of a first order-Taylor expansion. We have defined $\mathrm{Dom}(\xi_p)$ as a convex set because we need to compute integrals along segments between points of $\mathrm{Dom}(\xi_p)$. In general, we expect that the assumptions can be relaxed for unbounded functions at the cost of increased length and complexity of the proofs.

4.2 Weak Law of Large Numbers

In this section we establish a weak law of large numbers (WLLN). To do so, we state first a slightly stronger result that will be repeatedly used in the fluctuation analysis presented in Section 4.3.

Theorem 4.1. *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* $n \geq 0$*,* $r \geq 1$ *and* $\varphi_n : E_n \times \mathsf{V} \to \mathbb{R}^r$ *be a bounded function continuous at* $\mathsf{v} \in \mathsf{V}$ *uniformly on* E_n*. The following limit holds in probability*

$$\lim_{N\to\infty} \eta_n^N \left[\varphi_n(\cdot, V_N) \right] = \eta_n \left[\varphi_n(\cdot, \mathsf{v}) \right] .$$

Corollary 4.1 (WLLN)**.** *Assume (A1). Let* $n \geq 0$*,* $r \geq 1$ *and* $\varphi_n : E_n \to \mathbb{R}^r$ *a bounded measurable function. The following limit holds in probability,* $\lim_{N\to\infty} \eta_n^N(\varphi_n) = \eta_n(\varphi_n)$*.*

Proof of Theorem 4.1. It suffices to concentrate on the scalar case $r = 1$. The proof is by induction on n. The initial case $n = 0$ is a direct consequence of WLLN for i.i.d. random variables and Definition 2.1. For notational convenience, in the rest of the proof we write $\bar{\varphi}_n(\cdot)$ instead of $\varphi_n(\cdot, \mathsf{v})$. We assume the result at rank $n - 1$ and proceed to the induction step. Since V_N converges in probability to $\mathsf{v} \in \mathsf{V}$, Definition 2.1 shows that it suffices to prove that $[\eta_n^N - \eta_n]\big(\bar{\varphi}_n\big)$ converges in probability to zero. We use the decomposition

$$\begin{aligned}
[\eta_n^N - \eta_n](\bar{\varphi}_n) &= \big(\eta_n^N(\bar{\varphi}_n) - \mathbb{E}_{n-1}[\eta_n^N(\bar{\varphi}_n)]\big) + \big(\mathbb{E}_{n-1}[\eta_n^N(\bar{\varphi}_n)] - \eta_n(\bar{\varphi}_n)\big) \\
&= [\eta_n^N - \Phi_{n,N}(\eta_{n-1}^N)](\bar{\varphi}_n) + [\Phi_{n,N}(\eta_{n-1}^N) - \eta_n](\bar{\varphi}_n) =: A(N) + B(N) .
\end{aligned}$$

To conclude the proof, we now prove that each of these terms converges to zero in probability.

- 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. To this end, it suffices to notice that

$$\mathbb{E}_{n-1}\big[A(N)^2\big] = \tfrac{1}{N}\, \mathbb{E}_{n-1}\Big[\big(\bar{\varphi}(x_n^i) - \mathbb{E}_{n-1}[\bar{\varphi}(x_n^i)]\big)^2\Big] \leq \frac{\|\bar{\varphi}\|_\infty^2}{N} .$$

- To treat the quantity $B(N)$, we use the definition of $\Phi_{n,N}(\eta_{n-1}^N)$ in (9) and decompose it as the sum of three terms $B(N) = B_1(N) + B_2(N) + B_3(N)$ with

$$\begin{aligned}
B_1(N) &= \eta^N_{n-1}\big\{[Q_{n,N} - Q_n](\bar\varphi_n)\big\} \,/\, \eta^N_{n-1}(G_{n-1,N}) \;; \\
B_2(N) &= [\eta^N_{n-1} - \eta_{n-1}]\big(Q_n(\bar\varphi_n)\big) \,/\, \eta^N_{n-1}(G_{n-1,N}) \;; \\
B_3(N) &= \eta^N_{n-1}[Q_n(\bar\varphi_n)] \times \big\{1/\eta^N_{n-1}(G_{n-1,N}) - 1/\eta_{n-1}(G_{n-1})\big\} \;.
\end{aligned}$$

We prove that $B_i(N)$ converges in probability to zero for $i = 1, 2, 3$. The induction hypothesis shows that $\eta^N_{n-1}(\xi_n)$ converges to $\eta_{n-1}(\xi_n)$ in probability. By Assumption 1, the bounded function $(x, \xi) \mapsto G_{n-1,\xi}(x)$ is continuous at $\xi = \eta_{n-1}(\xi_n)$ uniformly on E_{n-1}; the induction hypothesis applies and $\eta^N_{n-1}(G_{n-1,N})$ converges in probability to $\eta_{n-1}(G_{n-1})$. Similarly, since $Q_n(\bar\varphi) \in \mathcal{B}_b(E_{n-1})$ is bounded by boundedness of $\bar\varphi_n$, it follows that $\eta^N_{n-1}[Q_n(\bar\varphi_n)]$ converges in probability to $\eta_{n-1}[Q_n(\bar\varphi_n)]$. Slutsky's Lemma thus yields that $B_2(N)$ and $B_3(N)$ converge to zero in probability. Finally, note that by Assumption 1 the bounded function $(x, \xi) \mapsto Q_{n,\xi}(x, \bar\varphi_n)$ is continuous at $\xi = \eta_{n-1}(\xi_n)$ uniformly on E_{n-1}; the induction yields

$$\begin{aligned}
\lim_{N\to\infty} \eta^N_{n-1}\big\{[Q_{n,N} - Q_n](\bar\varphi_n)\big\} &= \lim_{N\to\infty} \big\{\, \eta^N_{n-1}[Q_{n,N}(\bar\varphi_n)] - \eta_{n-1}[Q_n(\bar\varphi_n)] \,\big\} \\
&\quad - \lim_{N\to\infty} \big\{\, \eta^N_{n-1}[Q_n(\bar\varphi_n)] - \eta_{n-1}[Q_n(\bar\varphi_n)] \,\big\} = 0 \;,
\end{aligned}$$

which is enough for concluding that $B_1(N)$ converges to zero in probability.

□

As a corollary, one can establish a similar consistency result for the sequence of particle approximations $\gamma^N_n(\varphi_n)$, defined in Equation (7), of the unnormalised quantity $\gamma_n(\varphi_n)$.

Corollary 4.2. *Assume (A1). Let $n \geq 0$, $r \geq 1$ and $\varphi_n : E_n \to \mathbb{R}^r$ be a bounded measurable function. The following limit holds in probability, $\lim_{N\to\infty} \gamma^N_n(\varphi_n) = \gamma_n(\varphi_n)$.*

Proof. Since $\gamma^N_n(\varphi_n) = \gamma^N_n(1)\, \eta^N_n(1)$ and $\gamma_n(\varphi_n) = \gamma_n(1)\, \eta^N_n(1)$, by Corollary 4.1 it suffices to prove that $\gamma^N_n(1) = \eta^N_0(G_0) \times \ldots \times \eta^N_{n-1}(G_{n-1})$ converges in probability to the value $\gamma_n(1) = \eta_0(G_0) \times \ldots \times \eta_{n-1}(G_{n-1})$. By Assumption 1, the potentials $\{G_p\}_{p\geq 0}$ are bounded so that Corollary 4.1 applies and the quantity $\eta^N_p(G_p)$ converges in probability to $\eta_p(G_p)$ for any index $p \geq 0$. The conclusion directly follows. □

4.3 Central Limit Theorems

In this section, for a test function $\varphi_n : E_n \to \mathbb{R}^r$, we carry out a fluctuation analysis of the particle approximations $\gamma^N_n(\varphi_n)$ and $\eta^N_n(\varphi_n)$ around their limiting value. As expected, we prove that there is convergence at standard Monte-Carlo rate $N^{-1/2}$; in some situations, comparison with the perfect and non-adaptive algorithm is discussed in Section 4.4.

Theorem 4.2. *Assume (A1-2). Let $n \geq 0$, $r \geq 1$ and $\varphi_n : E_n \to \mathbb{R}^r$ be a bounded measurable function. The sequence $\sqrt{N}\,[\gamma_n^N - \gamma_n](\varphi_n)$ converges weakly to a centered Gaussian distribution with covariance*

$$\sum_{p=0}^{n} \gamma_p(1)^2\, \Sigma_{\eta_p}(\mathscr{L}_{p,n}\varphi_n) \tag{11}$$

where the linear operator $\mathscr{L}_p : \mathcal{B}_b(E_p)^r \to \mathcal{B}_b(E_{p-1})^r$ is defined by

$$\mathscr{L}_p\varphi_p = \eta_{p-1}[\partial_\xi Q_p \varphi_p]\,\big(\,\xi_p - \eta_{p-1}(\xi_p)\,\big) + Q_p(\varphi_p) \tag{12}$$

with $\mathscr{L}_{p,n} := \mathscr{L}_{p+1} \circ \ldots \circ \mathscr{L}_n$ and $\mathscr{L}_{n,n} = \mathrm{Id}$.

Proof. For notational convenience, we concentrate on the scalar case $r = 1$. The proof of the multi-dimensional case is identical, with covariance matrices replacing scalar variances. We proceed by induction on the parameter $n \geq 0$. The case $n = 0$ follows from the usual CLT for i.i.d. random variables. To prove the induction step it suffices to show that for any $t \in \mathbb{R}$ the following identity holds

$$\lim_{N\to\infty} \mathbb{E}\,\big[\, e^{it\sqrt{N}\,[\gamma_n^N - \gamma_n](\varphi_n)}\,\big] = e^{-\frac{1}{2}t^2\,\gamma_n(1)^2\,\Sigma_{\eta_n}(\varphi_n)} \lim_{N\to\infty} \mathbb{E}\,\big[\, e^{it\sqrt{N}\,[\gamma_{n-1}^N - \gamma_{n-1}](\mathscr{L}_n\varphi_n)}\,\big]\,. \tag{13}$$

Indeed, assuming that the induction hypothesis holds at time $n-1$, we have that

$$\lim_{N\to\infty} \mathbb{E}\,\big[\, e^{it\sqrt{N}\,[\gamma_{n-1}^N - \gamma_{n-1}](\mathscr{L}_n\varphi_n)}\,\big] = \exp\big\{-\tfrac{1}{2}t^2 \sum_{p=0}^{n-1} \gamma_p(1)^2\,\Sigma_{\eta_p}(\mathscr{L}_{p,n}\varphi_n)\big\}$$

and the proof of the induction step then follows from Levy's continuity theorem and (13). To prove (13) we use the following decomposition

$$\begin{aligned}[\gamma_n^N - \gamma_n](\varphi_n) &= \big\{\gamma_n^N(\varphi_n) - \mathbb{E}_{n-1}[\gamma_n^N(\varphi_n)]\big\} + \big\{\mathbb{E}_{n-1}[\gamma_n^N(\varphi_n)] - \gamma_n(\varphi_n)\big\} \\ &=: \widetilde{A}(N) + \widetilde{B}(N)\ .\end{aligned}$$

Since $\widetilde{B}(N) \in \mathscr{F}_{n-1}^N$ the expectation $\mathbb{E}[e^{it\sqrt{N}\,[\gamma_n^N - \gamma_n](\varphi_n)}]$ can be decomposed as

$$\begin{aligned}\mathbb{E}\Big[\Big(\mathbb{E}_{n-1}\big[e^{it\sqrt{N}\widetilde{A}(N)}\big] - e^{-\frac{1}{2}t^2\,\gamma_n(1)^2\,\Sigma_{\eta_n}(\varphi_n)}\Big) \times e^{it\sqrt{N}\,\widetilde{B}(N)}\Big] \\ + e^{-\frac{1}{2}t^2\,\gamma_n(1)^2\,\Sigma_{\eta_n}(\varphi_n)} \times \mathbb{E}\big[e^{it\sqrt{N}\,\widetilde{B}(N)}\big]\ .\end{aligned}$$

As a consequence, (13) follows once it is established that the limit

$$\lim_{N\to\infty} \mathbb{E}_{n-1}\Big[e^{it\sqrt{N}\widetilde{A}(N)}\Big] = \exp\big\{-\tfrac{1}{2}t^2\,\gamma_n(1)^2\,\Sigma_{\eta_n}(\varphi_n)\big\} \tag{14}$$

holds in probability and that $\sqrt{N}\,\widetilde{B}(N) = \sqrt{N}\,[\gamma_{n-1}^N - \gamma_{n-1}](\mathscr{L}_n(\varphi_n)) + o_{\mathbb{P}}(1)$. We finish the proof of Theorem 4.2 by establishing these two results.

- Quantity $\widetilde{A}(N)$ also reads as $\gamma_n^N(1)\,A(N)$ with $A(N) := \big[\eta_n^N - \Phi_{n,N}(\eta_{n-1}^N)\big](\varphi_n)$. By Corollary 4.2, $\gamma_n^N(1)$ converges in probability to $\gamma_n(1)$; to prove that $\mathbb{E}_{n-1}\big[e^{it\sqrt{N}\widetilde{A}(N)}\big]$ converges in probability to $\exp\big\{-\frac{1}{2}t^2\,\gamma_n(1)^2\,\Sigma_{\eta_n}(\varphi_n)\big\}$ it thus suffices to show that $\mathbb{E}_{n-1}\big[e^{it\sqrt{N}A(N)}\big]$ converges in probability to $\exp\big\{-\frac{1}{2}t^2\,\Sigma_{\eta_n}(\varphi_n)\big\}$. We will exploit the following identity

$$\mathbb{E}_{n-1}\left[e^{i\,t\,\sqrt{N}\,A(N)}\right] = \mathbb{E}_{n-1}\left[e^{i\,t\,\{\varphi_n(X_N)-\mathbb{E}_{n-1}[\varphi_n(X_N)]\}/\sqrt{N}}\right]^N$$

with X_N is distributed according to $\sum_{i=1}^N \frac{G_{n-1,N}(x_{n-1}^i)}{\sum_{j=1}^N G_{n-1,N}(x_{n-1}^i)}\,M_{n,N}(x_{n-1}^i, dx)$. Since the test function φ_n is bounded, a Taylor expansion yields that

$$\mathbb{E}_{n-1}\left[e^{i\,t\,\{\varphi_n(X_N)-\mathbb{E}_{n-1}[\varphi_n(X_N)]\}/\sqrt{N}}\right] = 1 - \tfrac{t^2}{N}\,\mathrm{Var}_{n-1}[\varphi_n(X_N)] + N^{-3/2}\times\mathcal{O}_{\mathbb{P}}(1)\ .$$

Consequently, $\mathbb{E}_{n-1}[e^{it\sqrt{N}\,A(N)}] = \exp\big\{-t^2\,\mathrm{Var}_{n-1}[\varphi_n(X_N)]/2\big\} + o_{\mathbb{P}}(1)$ and the proof is complete once it is shown that

$$\begin{aligned}\mathrm{Var}_{n-1}[\varphi_n(X_N)] &= \sum_{i=1}^N G_{n-1,N}(x_{n-1}^i)M_{n,N}(\varphi_n^2)(x_{n-1}^i)\ /\ \sum_{i=1}^N G_{n-1,N}(x_{n-1}^i)\\ &\qquad - \Big\{\sum_{i=1}^N G_{n-1,N}(x_{n-1}^i)M_{n,N}(\varphi_n)(x_{n-1}^i)\ /\ \sum_{i=1}^N G_{n-1,N}(x_{n-1}^i)\Big\}^2\\ &= \eta_{n-1}^N\big[Q_{n-1,\eta_{n-1}^N(\xi_n)}\varphi_n^2\big]\ /\ \eta_{n-1}^N\big[G_{n-1,\eta_{n-1}^N(\xi_n)}\big]\\ &\qquad - \Big\{\eta_{n-1}^N\big[Q_{n-1,\eta_{n-1}^N(\xi_n)}\varphi_n\big]\ /\ \eta_{n-1}^N\big[G_{n-1,\eta_{n-1}^N(\xi_n)}\big]\Big\}^2\end{aligned}$$

converges in probability to $\Sigma_{\eta_n}(\varphi_n)$. By Assumption 1, functions $(x,\xi)\mapsto G_{n-1,\xi}(x)$, $(x,\xi)\mapsto Q_{n,\xi}\varphi_n(x)$, $(x,\xi)\mapsto Q_{n,\xi}\varphi_n^2(x)$ are bounded and continuous at $\xi = \eta_{n-1}(\xi_n)$ uniformly on E_{n-1}. By Corollary 4.1, $\eta_{n-1}^N(\xi_n)$ converges in probability to $\eta_{n-1}(\xi_n)$; by Theorem 4.1 and Slutsky's Lemma we get that $\mathrm{Var}_{n-1}[\varphi_n(X_N)]$ converges in probability to

$$\eta_{n-1}[Q_n(\varphi_n^2)]/\eta_{n-1}(G_n) - \big(\eta_{n-1}[Q_n(\varphi_n)]/\eta_{n-1}(G_n)\big)^2\ ,$$

which is another formula for $\eta_n(\varphi_n^2) - \eta_n(\varphi_n)^2 = \Sigma_{\eta_n}(\varphi_n)$, as required.

- To prove that $\sqrt{N}\,\widetilde{B}(N) = \sqrt{N}\,[\gamma_{n-1}^N - \gamma_{n-1}](\mathscr{L}_n(\varphi_n)) + o_{\mathbb{P}}(1)$ we write $\widetilde{B}(N)$ as

$$\gamma_{n-1}^N(1)\times\eta_{n-1}^N[Q_{n,N} - Q_n](\varphi_n)\ +\ [\gamma_{n-1}^N - \gamma_{n-1}](Q_n\varphi_n)\ . \tag{15}$$

Furthermore, we have

$$\eta_{n-1}^N[Q_{n,N} - Q_n](\varphi_n) = \eta_{n-1}^N\left[\omega(\cdot,\eta_{n-1}^N(\xi_n))\right]\times[\eta_{n-1}^N - \eta_{n-1}](\xi_n) \tag{16}$$

with $\omega(x,z) := \int_0^1 \partial_\xi Q_{n,\xi}\varphi_n(x)|_{\xi=\eta_{n-1}(\xi_n)+\lambda(z-\eta_{n-1}(\xi_n))}\, d\lambda$. Under Assumption 2, function ω is bounded and continuous at $z = \eta_{n-1}(\xi_n)$ uniformly over $x \in E_{n-1}$. Theorem 4.1 applies so that $\eta_{n-1}^N\left[\,\omega(\cdot, \eta_{n-1}^N(\xi_n))\,\right] \to \eta_{n-1}\left[\,\omega(\cdot, \eta_{n-1}(\xi_n))\,\right] = \eta_{n-1}[\partial_\xi Q_n(\varphi)]$, in probability. The induction hypothesis, Slutky's Lemma and standard manipulations yield that $\sqrt{N} \times \gamma_n^N[Q_{n,N} - Q_n](\varphi_n)$ equals

$$\sqrt{N} \times \eta_{n-1}\big[\partial_\xi Q_n(\varphi)\big] \times [\gamma_{n-1}^N - \gamma_{n-1}](\xi_n - \eta_{n-1}(\xi_n)) + o_{\mathbb{P}}(1)\ .$$

It then follows from (15) that $\sqrt{N}\,\widetilde{B}(N) = \sqrt{N}\,[\gamma_{n-1}^N - \gamma_{n-1}](\mathscr{L}_n\varphi_n)\ +\ o_{\mathbb{P}}(1)$.

This concludes the proof of the induction steps and finishes the proof of Theorem 4.2. □

In the case where the summary statistics are constant, i.e. $\xi_p \equiv C \in \mathbb{R}$ for $p \geq 0$, expression (11) reduces to the usual non-adaptive asymptotic variance as presented, for example, in [9]. In the special case $\varphi_n \equiv 1$, one obtains the following expression for the asymptotic variance of the relative normalisation constant $\gamma_n^N(1)/\gamma_n(1)$.

Corollary 4.3. *Assume (A1-2) and let $n \geq 0$ be a non-negative integer. Then the quantity $\sqrt{N}\left\{\gamma_n^N(1)/\gamma_n(1) - 1\right\}$ converges, as $N \to \infty$, to a centered Gaussian distribution with variance*

$$\sum_{p=0}^{n} \frac{\mathrm{Var}_{\eta_p}(\mathscr{L}_{p,n}\,1)}{\prod_{k=p}^{n-1}\eta_k(G_k)^2}\ .$$

Similarly, one can obtain a CLT for the empirical normalised measures $\eta_n^N(\varphi_n)$:

Theorem 4.3. *Assume (A1-2). Let $n \geq 0$, $r \geq 1$ and $\varphi_n : E_n \to \mathbb{R}^r$ be a bounded measurable function. The sequence $\sqrt{N}\,[\eta_n^N - \eta_n](\varphi_n)$ converges weakly to a centered Gaussian distribution with covariance*

$$\Sigma_n(\varphi_n) := \sum_{p=0}^{n} \frac{\gamma_p(1)^2}{\gamma_n(1)^2}\,\Sigma_{\eta_p}\big[\mathscr{L}_{p,n}\big(\varphi_n - \eta_n(\varphi_n)\big)\big] \tag{17}$$

with the linear operators $\mathscr{L}_p$ for $p \geq 0$ as defined in (12). The asymptotic variances satisfy

$$\Sigma_n(\varphi_n) := \Sigma_{\eta_n}(\varphi_n) + \frac{\Sigma_{n-1}\big[\mathscr{L}_n\big(\varphi_n - \eta_n(\varphi_n)\big)\big]}{\eta_{n-1}(G_{n-1})^2}\ . \tag{18}$$

Proof. One can verify that the normalised measure η_n^N is related to the unnormalised measure γ_n^N through the identity ([9, pp. 301])

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

By Corollary 4.2, $\gamma_n(1)/\gamma_n^N(1)$ converges in probability to 1. Slutsky's Lemma and Theorem 4.2 yield that $\sqrt{N}\,[\eta_n^N - \eta_n](\varphi_n)$ converges weakly to a centered Gaussian variable with variance $\sum_{p=0}^n \gamma_p(1)^2\,\Sigma_{\eta_p}[\mathscr{L}_{p,n}\big(\gamma_n(1)^{-1}(\varphi_n - \eta_n(\varphi_n)\big)]$, which is just another way of writing (17). Equation (18) follows from the identities $\gamma_p(1) = \prod_{k=0}^{p-1}\eta_k(G_k)$, $\eta_{n-1}(\mathscr{L}_n\varphi_n) = \eta_n(\varphi_n)$. □

4.4 Stability

We now show that in the majority of applications of interest, the asymptotic variance of the adaptive SMC algorithm is identical to the asymptotic variance of the *perfect* algorithm.

Theorem 4.4 (Stability)**.** *Assume (A1-2). Suppose further that for any index $n \geq 1$ the identity*

$$\eta_{n-1}(G_{n-1,\xi} M_{n,\xi}) / \eta_{n-1}(G_{n-1,\xi}) = \eta_n \tag{19}$$

holds for any parameter $\xi \in Dom(\xi_n)$. For any test function $\varphi_n \in \mathcal{B}_b(E_n)$, the asymptotic variance of the adaptive SMC algorithm identified in Theorem 4.2 equals the asymptotic variance of the perfect SMC algorithm.

Proof. Formula (12) shows that it suffices to prove that the term $\eta_{n-1}(\partial_\eta Q_n \varphi_n)$ vanishes. By differentiation under the integral sign, it is enough to prove that the mapping $\xi \mapsto \eta_{n-1}(Q_{n,\xi}\varphi_n)$ is constant on $\mathrm{Dom}(\xi_n)$. Indeed, it follows from (19) that $\eta_{n-1}(Q_{n,\xi}\varphi_n) = \eta_n(\varphi_n)$ for any $\xi \in \mathrm{Dom}(\xi_n)$, concluding the proof of Theorem 4.4. □

Theorem 4.4 applies for instance to the sequential Bayesian parameter inference context discussed in Section 3.1 and to the filtering setting of Section 3.2. A consequence of Theorem 4.4 is that standard behaviours for the asymptotic variance of the *perfect* SMC algorithm, such as linear growth of the asymptotic variance of $\sqrt{N}\left(\gamma_n^N(1)/\gamma_n(1) - 1\right)$, are inherited by the adaptive SMC algorithm.

5 Adaptive Tempering

We now look at the scenario when one uses the information in the evolving particle population to adapt a sequence of distributions by means of a tempering parameter $\beta \in (0,1)$.

5.1 Algorithmic Set-Up

In many situations in Bayesian inference one seeks to sample from a distribution π on a set E of the form

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

where Z is a normalisation constant, $m(dx)$ a dominating measure on the set E and $V : E \to \mathbb{R}$ a potential. Coefficient $\beta_* \in \mathbb{R}$ can be thought of as an inverse temperature parameter. 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 inverse temperatures $\beta_0 \leq \ldots \leq \beta_{n-1} \leq \beta_n \leq \cdots \leq \beta_{n_*} = \beta_*$; in many applications $\beta_* = 1$. The associated unnormalised measures are

$$\gamma_n(dx) = e^{-\beta_n V(x)}\, m(dx) \ .$$

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 potentials

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

and Markov kernels M_n satisfying $\eta_n M_n = \eta_n$. For test function $\varphi_n \in \mathcal{B}_b(E)$, the N-particle approximation of the normalised and unnormalised distribution are given in (5), (7). To be consistent with the notations introduced in Section 4.3, note that the normalisation constants also read as $Z(\beta_n) = \gamma_n(1)$ and $Z = Z(\beta_*) = \gamma_{n^*}(1)$. In most scenarios of practical interest, it can be difficult or even undesirable to decide *a-priori* upon the annealing sequence $\{\beta_n\}_{n=0}^{n_*}$. Indeed, if the chosen sequence features big gaps, one may reach the terminal temperature rapidly, the variance of the weights being potentially very large due to large discrepancies between consecutive elements of the bridging sequence of probability distributions. Alternatively, if the gaps between the annealing parameters are too small, the variance of the final weights can be very small; this comes at the price of needlessly wasting a lot of computation time. Knowing what constitutes 'big' or 'small' with regards to the temperature gaps can be very-problem specific. Thus, an automated procedure for determining the annealing sequence is of great practical importance. In this section we investigate the asymptotic properties of an algorithm where the temperatures, as well as statistics of the MCMC kernel, are determined empirically by the evolving population of particles.

A partial analysis of the algorithm to be described can be found in [16]. However, the way in which the annealing sequence is determined in that work does not correspond to one typically used in the literature. In addition, the authors assume that the perfect MCMC kernels are used at each time step, whereas we do not assume so. It should also be noted, however, that the analysis in [16] is non-asymptotic.

The adaptive version of the above described algorithm constructs the (random) temperatures sequence $\{\beta_p^N\}_{p\geq 0}$ 'on the fly' as follows. Once a proportion $\alpha \in (0,1)$ has been specified, the random tempering sequence is determined through the recursive equation

$$\beta_{n+1}^N = \inf \left\{ \beta_n^N < \beta \leq \beta_* \ : \ \mathrm{ESS}(\eta_n^N, e^{-(\beta - \beta_n^N)\, V}) = \alpha \right\} \tag{20}$$

initialized at a prescribed value β_0 typically chosen so that the distribution η_0 is easy to sample from. For completeness, we use the convention that $\inf \varnothing = \beta_*$. In the above displayed equation, we have used the ESS functional defined for a measure η on the set E and a weight function $\omega : E \to (0, \infty)$ by

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

The following lemma guaranties that under mild assumptions the effective sample size functional $\beta \mapsto \mathrm{ESS}(\eta_p, e^{-(\beta - \beta_n)\, V})$ is continuous and decreasing so that (20) is well-defined and the inverse temperature β_{n+1} can be efficiently computed by a standard bisection method.

Lemma 5.1. *Let η be a finite measure on the set 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 $\mathbb{P}\,[\, V(X) \neq V(Y)\,] > 0$ for X, Y independent and distributed according to η, the function is strictly decreasing.*

Proof. We treat the case where $\mathbb{P}\,[\, V(X) \neq V(Y)\,] > 0$, the case $\mathbb{P}\,[\, V(X) \neq V(Y)\,] = 0$ being trivial. Let X and Y be two independent random variables distributed according to η. The dominated convergence theorem shows 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 $\eta(V\, e^{-\lambda V})\, \eta(e^{-2\lambda V}) > \eta(e^{-\lambda V})\, \eta(V\, e^{-2\lambda V})$, which is equivalent to the condition

$$\mathbb{E}\left[\, e^{-\lambda\{V(X)+V(Y)\}} \times \Big\{ V(X) - V(Y) \Big\} \times \Big\{ e^{-\lambda V(X)} - e^{-\lambda V(Y)} \Big\} \,\right] < 0 \ .$$

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 $x \neq y$. □

We will assume that the sequence of temperatures $\{\beta_n\}_{n \geq 0}$ and $\{\beta_n^N\}_{n \geq 0}$ are defined for *any* index $n \geq 0$, using the convention that the first time that the parameter β_n^N reaches the level β_*, which is random for the practical algorithm, the algorithm still goes on with fixed inverse temperatures equal to β_*. Under this convention, we can carry out an asymptotic analysis using an induction argument. Ideally one would like to prove asymptotic consistency (and a CLT) for the empirical measure at the random termination time of the practical algorithm; we do not do this, due to the additional technical challenge that it poses. We believe that the result to be proven still provides a very satisfying theoretical justification for the practical adaptive algorithm. We assume from now on that for the perfect algorithm the sequence of inverse temperatures is given by the limiting analogue of (20),

$$\beta_{n+1} = \inf \Big\{ \beta_n < \beta \le \beta_* \; : \; \mathrm{ESS}(\eta_n, e^{-(\beta - \beta_n)\, V}) = \alpha \Big\} \; . \tag{21}$$

We will show in the next section that under mild assumptions the adaptive version β_n^N converges in probability towards β_n. For statistics $\xi_{n+1} : E \to \mathbb{R}^d$ we set

$$\theta_n^N = \big(\beta_n^N , \beta_{n+1}^N , \eta_n^N(\xi_{n+1})^\top \big)^\top$$

and denote by θ_n its limiting value. At time n, for a particle system $\{x_n^i\}_{i=1}^N$ and associated empirical distribution η_n^N targeting the distribution η_n, the next inverse temperature β_{n+1}^N is computed according to (20); the particle system is re-sampled according to a multinomial scheme with weights

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

and then evolves via a Markov kernel $M_{n+1,N} \equiv M_{n+1, \eta_n^N(\xi_{n+1}), \beta_{n+1}^N}$ that preserves the preserves $Z(\beta_{n+1}^N)^{-1}\, e^{-\beta_{n+1}^N \, V}\, m(dx)$. Similarly to Section 2.2, we will make use of the operator

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

and its limiting analogue Q_n. With these notations, note that Equation (6) holds. To emphasise the dependencies upon the parameter $\theta = (\beta_1, \beta_2, \eta)$, we will sometimes use the expression $Q_{n,\theta} = G_{n,\theta}(x)\, M_{n,\eta,\beta_2}(x,dy)$ with $G_{n,\theta} = e^{-(\beta_2 - \beta_1)\, V} = e^{-\Delta\, V}$ and $\Delta = \beta_2 - \beta_1$. For notational convenience, we sometimes write ∂_Δ when the meaning is clear. For example, by differentiation under the integral sign, the quantity $\partial_\Delta \eta_n(G_n)$ also equals $-\eta_n(V\, G_n)$. Unless otherwise stated, the derivative ∂_θ is evaluated at the limiting parameter $\theta_n = (\beta_n, \beta_{n+1}, \eta_n(\xi_{n+1}))$.

5.2 Assumptions

We define $\mathrm{Dom}(\beta) = \{(\beta_1, \beta_2) \in [\beta_0, \beta_*]^2 \; ; \; \beta_1 \le \beta_2\}$. By $\mathrm{Dom}(\xi_p) \subset \mathbb{R}^d$ we denote a convex set that contains the range of the statistic $\xi_p : E_{p-1} \to \mathbb{R}^d$. The results to be presented in the next section make use of the following hypotheses.

(**A3**) The potential V is bounded on the set E. For each $n \ge 0$ the function $(x, \theta) \mapsto G_{n,\theta}(x)$ is bounded and continuous at $\theta_n = \big(\beta_n, \beta_{n+1}, \eta_n(\xi_{n+1})\big)$ uniformly on E. The statistic

$\xi_n : E \to \mathbb{R}^d$ is bounded. For any bounded Borel test function $\varphi_n : E \to \mathbb{R}^r$, the function $(x, \theta) \mapsto Q_{n,\theta}\varphi_n(x)$ is bounded and continuous at $\theta = \theta_{n-1}$ uniformly on E.

(**A4**) For each $n \geq 1$, $r \geq 1$ and bounded Borel test function $\varphi_n : E \to \mathbb{R}^r$ the function $(x, \theta) \mapsto \partial_\theta Q_{n,\theta}\, \varphi_n(x)$ is well defined, bounded and continuous at $\theta = \theta_{n-1}$ uniformly on E.

These conditions could be relaxed at the cost of considerable technical complications in the proofs.

5.3 Weak Law of Large Numbers

In this section we prove that the consistency results of Section 4.2 also hold in the adaptive annealing setting. To do so, we prove that for any index $n \geq 0$ the empirical inverse temperature parameter β_n^N converges in probability towards β_n.

Theorem 5.1 (WLLN)**.** *Assume (A3). For any $n \geq 0$, the empirical inverse temperature β_n^N converges in probability to β_n as $N \to \infty$. Also, 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 $r \geq 1$ and $\varphi_n : E \times \mathsf{V} \to \mathbb{R}^r$ a bounded function continuous at $\mathsf{v} \in \mathsf{V}$ uniformly on E. Then, the following limit holds in probability*

$$\lim_{N \to \infty} \eta_n^N[\varphi_n(\cdot, V_N)] = \eta_n[\varphi_n(\cdot, \mathsf{v})] \ .$$

Corollary 5.1 (WLLN)**.** *Assume (A3). Let $n \geq 0$, $r \geq 1$ and $\varphi_n : E \to \mathbb{R}^r$ be a bounded measurable function. The following limit holds in probability, $\lim_{N\to\infty} \eta_n^N(\varphi_n) = \eta_n(\varphi_n)$.*

Corollary 5.2. *Assume (A3). Let $n \geq 0$, $r \geq 1$ and $\varphi_n : E \to \mathbb{R}^d$ a bounded measurable function. The following limit holds in probability, $\lim_{N\to\infty} \gamma_n^N(\varphi_n) = \gamma_n(\varphi_n)$.*

Proof of Theorem 5.1. Clearly, it suffices tp concentrate on the case $r = 1$. We prove by induction on the rank $n \geq 0$ that β_n^N converges in probability to β_n and for any test function $\varphi : E \times \mathsf{V} \to \mathbb{R}$ bounded and continuous at $\mathsf{v} \in \mathsf{V}$ uniformly on E that $[\eta_n^N - \eta_n](\varphi) \to_{\mathbb{P}} 0$. 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 β_n^N converges in probability to β_n. Note that β_n^N can also be expressed as

$$\beta_n^N := \inf \left\{ \beta \in [\beta_0, \beta_*] \ : \ \frac{\zeta_{1,n-1}^N(\beta)}{\zeta_{2,n-1}^N(\beta)} \leq \alpha \right\}$$

with $\zeta^N_{1,n-1}(\beta) = \eta^N_{n-1}[e^{-\max(0,\beta-\beta^N_{n-1})\,V}]^2$ and $\zeta^N_{2,n-1}(\beta) = \eta^N_{n-1}[e^{-2\max(0,\beta-\beta^N_{n-1})\,V}]$. Indeed, the limiting temperature β_n can also be expressed as

$$\beta_n := \inf\left\{\beta \in [\beta_0, \beta_*] \;:\; \frac{\zeta_{1,n-1}(\beta)}{\zeta_{2,n-1}(\beta)} \leq \alpha\right\}$$

where $\zeta_{1,n-1}(\beta)$ and $\zeta_{2,n-1}(\beta)$ are the limiting values of $\zeta^N_{1,n-1}(\beta)$ and $\zeta^N_{2,n-1}(\beta)$. The dominated convergence theorem shows that the paths $\beta \mapsto \zeta^N_{1,n-1}(\beta)/\zeta^N_{2,n-1}(\beta)$ and $\beta \mapsto \zeta_{1,n-1}(\beta)/\zeta_{2,n-1}(\beta)$ are continuous; it thus suffices to prove that the limit

$$\lim_{N\to\infty} \left\|\zeta^N_{1,n-1}(\beta)/\zeta^N_{2,n-1}(\beta) - \zeta_{1,n-1}(\beta)/\zeta_{2,n-1}(\beta)\right\|_{\infty,[\beta_0,\beta_*]} = 0 \tag{22}$$

holds in probability. Lemma 5.1 shows that the function $\beta \mapsto \zeta^N_{i,n-1}(\beta)$ is decreasing on $[\beta_0, \beta_*]$ for any $1 \leq i \leq 2$ and $n, N \geq 1$; by standard arguments, for proving (22) it suffices to show that for any fixed inverse temperature $\beta \in [\beta_0, \beta_*]$ the difference $\zeta^N_{1,n-1}(\beta)/\zeta^N_{2,n-1}(\beta) - \zeta_{1,n-1}(\beta)/\zeta_{2,n-1}(\beta)$ converges to zero in probability. Indeed, one can focus on proving that $\zeta^N_{i,n-1}(\beta)$ 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 entirely similar.

  – For the case $\beta < \beta_{n-1}$, the induction hypothesis shows that β^N_{n-1} converges in probability to β_{n-1}. Since $\zeta^N_{2,n-1}(\beta) = 1 = \zeta_{2,n-1}(\beta)$ for $\beta \leq \min(\beta^N_{n-1}, \beta_{n-1})$, the conclusion follows.

  – The case $\beta \geq \beta_{n-1}$ follows from the convergence in probability of β^N_{n-1} to β_{n-1} and $\eta^N_{n-1}(e^{-(\beta-\beta_{n-1})\,V})$ to $\eta_{n-1}(e^{-(\beta-\beta_{n-1})\,V})$.

- To prove that $\eta^N_n[\varphi_n(\cdot, V_N)]$ converges in probability towards $\eta_n[\varphi_n(\cdot, \mathsf{v})]$, because of the convergence in probability of β^N_n to β_n, of $\eta^N_{n-1}(\xi_n)$ to $\eta_{n-1}(\xi_n)$ and of V_n to v, one can use exactly the same approach as the one in the proof of Theorem 4.1.

□

5.4 Central Limit Theorem

In this section we extend the fluctuation analysis of Section 4.3 to the adaptive annealing setting. We prove that for a test function φ_n the empirical quantity $\gamma^N_n(\varphi_n)$ converges at $N^{-1/2}$-rate towards its limiting value $\gamma_n(\varphi_n)$; we give explicit recursive expressions for the asymptotic variances. It is noted that results for $\eta^N_n(\varphi_n)$ may also be proved as in Section 4.3, but are omitted for brevity. Before stating the main result of this section, several notations need to be introduced. For any $n \geq 0$ and test function $\varphi_n : E \to \mathbb{R}^r$

we consider the extension operator Ext_n that maps the test function φ_n to the function $\mathrm{Ext}_n(\varphi_n) : E \to \mathbb{R}^{r+2}$ defined by

$$\mathrm{Ext}_n(\varphi) := \Big(G_n - \eta_n(G_n),\, G_n^2 - \eta_n(G_n^2),\, \varphi_n\Big)^\top.$$

The linear operator $\mathcal{A}_n$ maps the bounded Borel function $\varphi_n : E \to \mathbb{R}^r$ to the rectangular $(r+1) \times (r+3)$ matrix $\mathcal{A}_n(\varphi_n)$ defined by $[\mathcal{A}_n(\varphi_n)]_{1,1} = 1$, $[\mathcal{A}_n(\varphi)]_{1,[4:r+3]} = 0_{1\times r}$, $[\mathcal{A}_n(\varphi_n)]_{[2:r+1],[4:r+3]} = I_{r\times r}$ and

$$\begin{aligned}
&[\mathcal{A}_n(\varphi_n)]_{1,2} = -2\gamma_{n-1}^{-1}(1)\, \frac{\eta_{n-1}(G_{n-1})}{\eta_{n-1}(G_{n-1}^2)} \cdot \Big\{\partial_\Delta\Big[\frac{\eta_{n-1}(G_{n-1})^2}{\eta_{n-1}(G_{n-1}^2)}\Big]\Big\}^{-1} \;;\\
&[\mathcal{A}_n(\varphi_n)]_{1,3} = \gamma_{n-1}^{-1}(1)\, \frac{\eta_{n-1}(G_{n-1})^2}{\eta_{n-1}(G_{n-1}^2)^2} \cdot \Big\{\partial_\Delta\Big[\frac{\eta_{n-1}(G_{n-1})^2}{\eta_{n-1}(G_{n-1}^2)}\Big]\Big\}^{-1} \;;\\
&[\mathcal{A}_n(\varphi_n)]_{2:r+1,1} = \big(\partial_{\beta_{n-1}} + \partial_{\beta_n}\big)\, \eta_{n-1}(Q_n\varphi_n) \;;\\
&[\mathcal{A}_n(\varphi_n)]_{2:r+1,2} = \gamma_{n-1}(1)\, \eta_{n-1}[\partial_{\beta_n} Q_n\varphi_n] \times [\mathcal{A}_n(\varphi_n)]_{1,2} \;;\\
&[\mathcal{A}_n(\varphi_n)]_{2:r+1,3} = \gamma_{n-1}(1)\, \eta_{n-1}[\partial_{\beta_n} Q_n\varphi_n] \times [\mathcal{A}_n(\varphi_n)]_{1,3} \;.
\end{aligned}$$

Theorem 5.2 (CLT)**.** *Assume (A3)-(A4). Let $n \geq 0$, $r \geq 1$ and $\varphi_n : E_n \to \mathbb{R}^r$ be a bounded measurable function. The sequence $\sqrt{N}\,\big(\beta_n^N - \beta_n, [\gamma_n^N - \gamma_n](\varphi_n)\big)^\top$ converges weakly to a centred Gaussian distribution with covariance*

$$\Sigma_n(\varphi_n) \;=\; \mathcal{A}_n(\varphi_n) \cdot \Sigma_{n-1}\big(\mathrm{Ext}_{n-1}(Q_n\varphi_n)\big) \cdot \mathcal{A}_n(\varphi_n)^\top + \gamma_n^2(1)\, \widetilde{\Sigma}_{\eta_n}(\varphi_n) \tag{23}$$

where $\widetilde{\Sigma}_{\eta_n}(\varphi_n)$ is the covariance matrix of the function $\big(0, \varphi_n\big)^\top$ under η_n.

Proof. The proof follows closely the one of Theorem 4.2. For the reader's convenience, we only highlight the differences. The proof proceeds by induction, the case $n = 0$ directly following from the CLT for i.i.d random variables. For proving the induction step, assuming that the result holds at rank $n-1$, it suffices to prove that

$$\mathbb{E}_{n-1}\left(\begin{array}{c} \beta_n^N - \beta_n \\ \big[\gamma_n^N - \gamma_n\big](\varphi_n) \end{array}\right) \;=\; \mathcal{A}_{n,N}(\varphi_n)\left(\begin{array}{c} \beta_{n-1}^N - \beta_{n-1} \\ \big[\gamma_n^N - \gamma_n\big]\big(\mathrm{Ext}[Q_n\varphi_n]\big) \end{array}\right), \tag{24}$$

with $\mathcal{A}_{n,N}(\varphi_n) \in \mathsf{M}_{r+1,r+3}(\mathbb{R})$ converging in probability to $\mathcal{A}_n(\varphi_n)$, and that for any vector $t \in \mathbb{R}^r$ the following limit holds in probability

$$\lim_{N\to\infty} \mathbb{E}\Big[\exp\Big\{i\, t\, \sqrt{N}\, C(N)\Big] \;=\; \exp\big\{-\gamma_n^2(1)\, \langle t, \Sigma_{\eta_n}(\varphi_n)\, t\rangle/2\big\}$$

with $C(N) = (\gamma_n^N - \gamma_n)(\varphi_n) - \mathbb{E}_{n-1}\big[(\gamma_n^N - \gamma_n)(\varphi_n)\big]$. The proof of the above displayed equation is identical to the proof of (14) and is thus omitted. We now prove (24).

- We first treat the term $\mathbb{E}_{n-1}[\beta_n^N - \beta_n] = \beta_n^N - \beta_n$. The relation $\mathrm{ESS}(\eta_{n-1}^N, e^{-\Delta_{n-1}^N V}) = \alpha = \mathrm{ESS}(\eta_{n-1}, e^{-\Delta_{n-1} V})$ can be rearranged as

$$\begin{aligned} \eta_{n-1}(G_{n-1})^2 \Big\{ \eta_{n-1}^N(e^{-2\Delta_{n-1}^N V}) - \eta_{n-1}(e^{-2\Delta_{n-1} V}) \Big\} &= \\ \eta_{n-1}(G_{n-1}^2) \Big\{ \eta_{n-1}^N(e^{-\Delta_{n-1}^N V})^2 - \eta_{n-1}(e^{-\Delta_{n-1} V})^2 \Big\} & . \end{aligned} \tag{25}$$

Decomposing $\eta_{n-1}^N(e^{-2\Delta_{n-1}^N V}) - \eta_{n-1}(e^{-2\Delta_{n-1} V})$ as the sum of $\eta_{n-1}^N(e^{-2\Delta_{n-1}^N V}) - e^{-2\Delta_{n-1} V})$ and $[\eta_{n-1}^N - \eta_{n-1}](G_{n-1}^2)$, and using a similar decomposition for the difference $\eta_{n-1}^N(e^{-\Delta_{n-1}^N V})^2 - \eta_{n-1}(e^{-\Delta_{n-1} V})^2$, one can exploit the boundedness of the potential V, Theorem 5.1 and the same approach as the one used for proving (16) to obtain that $\eta_{n-1}^N(e^{-2\Delta_{n-1}^N V}) - \eta_{n-1}(e^{-2\Delta_{n-1} V})$ equals

$$\Big\{ \partial_\Delta \eta_{n-1}(G_{n-1}^2) + o_{\mathbb{P}}(1) \Big\} \times (\Delta_{n-1}^N - \Delta_{n-1}) + [\eta_{n-1}^N - \eta_{n-1}](G_{n-1}^2) \tag{26}$$

and $[\eta_{n-1}^N(\kappa^{\Delta_{n-2}^N})^2 - \eta_{n-2}(\kappa^{\Delta_{n-2}})^2]$ equals

$$\begin{aligned} &\Big\{ 2\eta_{n-1}(G_{n-1}) \partial_\Delta \eta_{n-1}(G_{n-1}) + o_{\mathbb{P}}(1) \Big\} \times (\Delta_{n-2}^N - \Delta_{n-2}) \\ &\qquad + \Big\{ 2\eta_{n-1}(G_{n-1}) + o_{\mathbb{P}}(1) \Big\} \times [\eta_{n-1}^N - \eta_{n-1}](G_{n-1}) \ . \end{aligned} \tag{27}$$

Since $(\Delta_{n-1}^N - \Delta_{n-1})$ equals $(\beta_n^N - \beta_n) + (\beta_{n-1}^N - \beta_{n-1})$, Slutsky's Lemma, Equations (25), (26), (27) and standard algebraic manipulations yield

$$\begin{aligned} (\beta_n^N - \beta_n) = [\mathcal{A}_{n,N}(\varphi)]_{1,1} \, &(\beta_{n-1}^N - \beta_{n-1}) \\ &+ [\mathcal{A}_{n,N}(\varphi)]_{1,2} \, [\gamma_{n-1}^N - \gamma_{n-1}]\big(G_{n-1} - \eta_{n-1}(G_{n-1})\big) \\ &+ [\mathcal{A}_{n,N}(\varphi)]_{1,3} \, [\gamma_{n-1}^N - \gamma_{n-1}]\big(G_{n-1}^2 - \eta_{n-1}(G_{n-1}^2)\big) \end{aligned} \tag{28}$$

where $[\mathcal{A}_{n,N}(\varphi)]_{1,i}$ converges in probability to $[\mathcal{A}_{n,N}(\varphi)]_{1,i}$ for $1 \le i \le 3$.

- To deal with the term $\mathbb{E}_{n-1}\big[(\gamma_n^N - \gamma_n)(\varphi_n)\big]$ we make use of the decomposition

$$\mathbb{E}_{n-1}\big[(\gamma_n^N - \gamma_n)(\varphi_n)\big] = \gamma_{n-1}^N(1) \times \eta_{n-1}^N[Q_{n,N} - Q_n](\varphi_n) + [\gamma_{n-1}^N - \gamma_{n-1}](Q_n \varphi_n) \ . \tag{29}$$

Assumptions (A3)-(A4), Theorem 5.1 and the same approach as the one used for proving (16) show that the term $\eta_{n-1}^N[Q_{n,N} - Q_n](\varphi_n)$ equals

$$\Big\{ \eta_{n-1}[\partial_{\beta_{n-1}} Q_n \varphi_n] + o_{\mathbb{P}}(1) \Big\} (\beta_{n-1}^N - \beta_{n-1}) + \Big\{ \eta_{n-1}[\partial_{\beta_n} Q_n \varphi_n] + o_{\mathbb{P}}(1) \Big\} (\beta_n^N - \beta_n) \ .$$

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 $\xi \in \mathbb{R}^r$ the Markov kernel $M_{n,\xi}$ preserves η_n so that one can readily check that $\eta_{n-1}[\partial_\xi Q_{n,\xi} \varphi_n] = 0$. One can then use (28) to express $(\beta_n^N - \beta_n)$ in terms of the three quantities $(\beta_{n-1}^N - \beta_{n-1})$,

$[\gamma_{n-1}^N - \gamma_{n-1}]\big(G_{n-1} - \eta_{n-1}(G_{n-1})\big)$ and $[\gamma_{n-1}^N - \gamma_{n-1}]\big(G_{n-1}^2 - \eta_{n-1}(G_{n-1}^2)\big)$ and obtain, via Slutsky's Lemma and (29), that for any coordinate $1 \le i \le r$,

$$\begin{aligned}\mathbb{E}_{n-1}\big[(\gamma_n^N - \gamma_n)(\varphi_n)\big]_i = [\mathcal{A}_{n,N}(\varphi)]_{i+1,1}\,(\beta_{n-1}^N - \beta_{n-1}) \\ + [\mathcal{A}_{n,N}(\varphi)]_{i+1,2}\,[\gamma_{n-1}^N - \gamma_{n-1}]\big(G_{n-1} - \eta_{n-1}(G_{n-1})\big) \\ + [\mathcal{A}_{n,N}(\varphi)]_{i+1,3}\,[\gamma_{n-1}^N - \gamma_{n-1}]\big(G_{n-1}^2 - \eta_{n-1}(G_{n-1}^2)\big) \\ + [\gamma_{n-1}^N - \gamma_{n-1}](Q_n\varphi_n)_i\end{aligned} \tag{30}$$

where $[\mathcal{A}_{n,N}(\varphi)]_{i+1,j}$ converges in probability to $[\mathcal{A}_n(\varphi)]_{i+1,j}$ for $1 \le j \le 3$.

Equation (24) is a simple rewriting of (28) and (30). This concludes the proof. □

6 Applications

6.1 Verifying the Assumptions

We consider the sequential Bayesian parameter inference framework of Section 3.1. That is, for a parameter $x \in E = \mathbb{R}^m$, observations $y_i \in \mathcal{Y}$ and prior measure with density $\eta_0(x)$ with respect to the Lebesgue measure in $\mathbb{R}^m$. We assume the following.

(**B1**) For each $n \ge 1$ the function $G_n(x) := \mathbb{P}[y_{n+1} \mid y_{1:n}, x]$ is bounded and strictly positive. The statistics $\xi_n : E \to \mathbb{R}^d$ is bounded.

(**B2**) For each $n \ge 1$, the parametric family of Markov kernel $M_{n,\xi}$ is given by a Random-Walk-Metropolis kernel. The proposal density $q(\cdot;\xi)$ is symmetric; for a current position $x \in E$ the proposal y is such that $\mathbb{P}(y - x \in du) = q(u;\xi)\,du$. We suppose that the first and second derivatives

$$\xi \mapsto \nabla_\xi q(u;\xi)\ ; \quad \xi \mapsto \nabla_\xi^2 q(u;\xi)\ ,$$

are bounded on the range $\mathrm{Dom}(\xi_n)$ of the adaptive statistics $\xi_n : E \to \mathrm{Dom}(\xi_n) \subset \mathbb{R}^d$.

Assumption (B1) is reasonable and satisfied by many real statistical models. Similarly, it is straightforward to construct proposals verifying Assumption (B2); one can for example show that for a function $\sigma : \mathrm{Dom}(\xi_n) \to \mathbb{R}_+$, bounded away from zero with bounded first and second derivatives, the Gaussian proposal density $q(u;\xi) := \exp\big\{-u^2/[2\sigma^2(\xi)]\big\}/\sqrt{2\pi\sigma^2(\xi)}$ satisfies Assumption (B2); multi-dimensional extensions of this settings are readily constructed.

Proposition 6.1. *Assume (B1-2). The kernels* $(M_{n,\cdot})_{n\geq 1}$ *and potentials* $(G_n)_{n\geq 0}$ *satisfy Assumptions (A1-2).*

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 Assumptions (A1-2) 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 the proposal $x \mapsto x+u$ is $r(x,u) := \min\big\{1, \big(\mathbb{P}[y_{1:n} \mid x+u]\,\eta_0(x+u)\big) \,/\, \big(\mathbb{P}[y_{1:n} \mid x]\,\eta_0(x)\big)\big\}$ and we have $M_{n,\xi}(\varphi)(x) = \varphi(x) + \int_{\mathbb{R}^m} \big[\varphi(x+u) - \varphi(x)\big]\, r(x,u)\, q(u;\xi)\, du$. Differentiation under the integral sign yields

$$\begin{aligned}
\nabla_\xi M_{n,\xi}(\varphi)(x) &= \int \big[\varphi(x+u) - \varphi(x)\big]\, r(x,u)\, \nabla_\xi q(u;\xi)\, du \ , \\
\nabla^2_\xi M_{n,\xi}(\varphi)(x) &= \int \big[\varphi(x+u) - \varphi(x)\big]\, r(x,u) \nabla^2_\xi q(u;\xi)\, du \ ,
\end{aligned}$$

and the conclusion follows by boundedness of the first and second derivative of $q(u;\xi)$ with respect to the parameter $\xi \in \mathrm{Dom}(\xi_n)$. □

6.2 Numerical Example

We now provide a numerical study of a high-dimensional sequential Bayesian parameter inference, as described in Section 3.1, applied to the Navier-Stokes model. In this section, we briefly describe the Navier-Stokes model, the associated SMC algorithm and focus on the analysis of the behavior of the method when estimating the normalising constant. The SMC method to be presented is described in detail in [18]. In the subsequent discussion, we highlight the algorithmic challenges and the usefulness of the adaptive SMC methodology when applied to such high-dimensional scenarios. This motivates theoretical results presented in Section 6.2.3 where the stability properties of the SMC estimates are investigated in the regime where the dimension d of the adaptive statistics is large.

6.2.1 Model Description

We work with the Navier-Stokes dynamics describing the incompressible flow of a fluid in a two dimensional torus $\mathbb{T} = [0, 2\pi) \times [0, 2\pi)$. The time-space varying velocity field is denoted by $v(t,x) : [0,\infty) \times \mathbb{T} \to \mathbb{R}^2$. The Newton's laws of motion yield the Navier-Stokes system of partial differential equations [13]

$$\partial_t v - \nu \Delta v + (v \cdot \nabla)\, v = f - \nabla \mathfrak{p} \ , \qquad \nabla \cdot v = 0 \ , \qquad \int_{\mathbb{T}} v(x,\cdot)\, dx = 0 \ , \tag{31}$$

with initial condition $v(x,0) = u(x)$. The quantity $\nu > 0$ is a viscosity parameter, $\mathfrak{p} : \mathbb{T} \times [0,\infty) \to \mathbb{R}$ is the pressure field and $f : \mathbb{T} \to \mathbb{R}^2$ is an exogenous time-homogeneous forcing. For simplicity, we assume periodic boundary conditions. We adopt a Bayesian approach for inferring the unknown initial condition $u = u(x)$ from noisy measurements of the evolving velocity field $v(\cdot,t)$ on a fixed grid of points $\big(x_1,\ldots,x_M\big) \in \mathbb{T}$. Performing inference with this type of data is referred to as Eulerian data assimilation. Measurements are available at time $t_j := j \times \delta$ for time increment $\delta > 0$ and index $1 \le j \le T$ at each fixed location $x_m \in \mathbb{T}$. We assume i.i.d Gaussian measurements error with standard deviation $\varepsilon > 0$ so that the noisy observations $y := \big\{y_{j,m}\big\}_{j,m}$ for $1 \le j \le T$ and $1 \le m \le M$ can be modelled as

$$y_{j,m} = v\,(x_m, t_j) + \varepsilon\,\zeta_{j,m}$$

for an i.i.d sequence $\zeta_{j,m} \overset{iid}{\sim} \mathcal{N}(0, I_2)$. We follow the notations of [18] and set

$$\mathbb{U} = \Big\{\ 2\pi\text{-periodic trigonometric polynomials } u :\ \mathbb{T} \to \mathbb{R}^2 \Big|\ \nabla \cdot u = 0\ ,\ \int_{\mathbb{T}} u(x) dx = 0 \Big\}\ .$$

We use a Gaussian random field prior for the unknown initial condition; as will become apparent from the discussion to follow, it is appropriate in this setting to assume that the initial condition $u = u(x)$ belongs the closure U of $\mathbb{U}$ with respect to the $\big(L^2(\mathbb{T})\big)^2$ norm. The semigroup operator for the Navier-Stokes PDE is denoted by $\Psi : U \times [0,\infty) \to U$ so that the likelihood for the noisy observation y reads

$$\ell(y;u) = \exp\Big\{ -\frac{1}{2\varepsilon^2} \sum_{j=1}^{T} \sum_{m=1}^{M} \big\| y_{j,m} - [\Psi(u,t_j)](x_m)\,\big\|^2 \Big\} / (2\pi\varepsilon^2)^{MT}\ . \tag{32}$$

Under periodic boundary conditions, an appropriate orthonormal basis for U is comprised of the functions $\psi_k(x) := \big(k^\perp/(2\pi\,|k|\big)\,e^{ik\cdot x}$ for $k \in \mathbb{Z}^2_* := \mathbb{Z}^2 \setminus \{(0,0)\}$ and $k^\perp := (-k_2, k_1)^\top$, $|k| := \sqrt{k_1^2 + k_2^2}$. The index k corresponds to a bivariate frequency and the Fourier series decomposition of an element $u \in U$ reads

$$u(x) = \sum_{k \in \mathbb{Z}^2_*} u_k\,\psi_k(x) \tag{33}$$

with Fourier coefficients $u_k = \langle u, \psi_k \rangle = \int_{\mathbb{T}} u(x) \cdot \overline{\psi}_k(x)\,dx$. Since the initial condition $u \in U$ is real-valued we have $\overline{u_k} = -u_{-k}$ and one can focus on reconstructing the frequencies in the subset

$$\mathbb{Z}^2_{\uparrow} = \Big\{ k = (k_1,k_2) \in \mathbb{Z}^2_* \,:\, [k_1 + k_2 > 0] \text{ or } [k_1 = -k_2 > 0] \Big\}.$$

We adopt a Bayesian framework and assume a centred Gaussian random field prior η_0 on the unknown initial condition

$$\eta_0 = \mathcal{N}(0, \beta^2 A^{-\alpha}) \tag{34}$$

with hyper-parameters α, β affecting the roughness and magnitude of the initial vector field. In (34), $A = -P\Delta$ denotes the Stokes operator where $\Delta = \big(\partial^2_{x_1} + \partial^2_{x_2}, \, \partial^2_{x_1} + \partial^2_{x_2}\big)$ is the usual Laplacian and $P : \big(L^2(\mathbb{T})\big)^2 \to U$ is the Leray-Helmholtz orthogonal projector that maps a field to its divergence-free and zero-mean part. A simple understanding of the prior distribution η_0 can be obtained through the Karhunen-Loéve representation; a draw from the prior distribution η_0 can be realised as the infinite sum

$$\mathfrak{z} = \beta \sum_{k \in \mathbb{Z}^2_*} |k|^{-\alpha} \, \xi_k \, \psi_k \;\sim\; \eta_0 \tag{35}$$

where variables $\{\xi_k\}_{k \in \mathbb{Z}^2_*}$ correspond standard complex centred Gaussian random variables with $\big(\mathrm{Re}(\xi_k), \mathrm{Im}(\xi_k)\big) \overset{iid}{\sim} \mathcal{N}\big(0, \tfrac{1}{2} I_2\big)$ for $k \in \mathbb{Z}^2_\uparrow$ and $\xi_k = -\overline{\xi_{-k}}$ for $k \in \mathbb{Z}^2_* \setminus \mathbb{Z}^2_\uparrow$. In other words, *a-priori*, the Fourier coefficients u_k with $k \in \mathbb{Z}^2_\uparrow$ are assumed independent, normally distributed, with a particular rate of decay for their variances as $|k|$ increases. Statistical inference is carried out by sampling from the posterior probability measure η on U defined as the Gaussian change of measure

$$\frac{d\eta}{d\eta_0}(u) = \frac{1}{Z(y)} \, \ell(y; u) \tag{36}$$

for a normalisation constant $Z(y) > 0$.

6.2.2 Algorithmic Challenges and Adaptive SMC

With a slight abuse of notation we will henceforth use a single subscript to count the observations and set $y_{(j-1)M+m} \equiv y_{j,m}$. We will apply an SMC sampler on the sequence of distributions $\{\eta_n\}_{n=0}^{M \times T}$ defined by

$$\frac{d\eta_n}{d\eta_0}(u) = \frac{1}{Z(y_{1:n})} \, \ell(y_{1:n}; u) \tag{37}$$

for a normalisation constant $Z(y_{1:n})$ and likelihood $\ell(y_{1:n}; u)$. Note that the state space U is infinite-dimensional even though in practice, as described in [18], our solver truncates the Fourier expansion (33) on a pre-specified window of frequencies $-k_{\max} + 1 \le k_1, k_2 \le k_{\max}$ for $k_{\max} = 32$.

We now describe the MCMC mutation steps used for propagating the N-particle system. For a tuning parameter $\rho \in (0, 1)$, a simple Markov kernel suggested in several articles (see e.g. [7] and the references therein) for target distributions that are Gaussian changes of measure of the form (37) is the following. Given the current position $u \in U$, the proposal $\widetilde{u}$ is defined as

$$\widetilde{u} = \rho \, u + (1 - \rho^2)^{1/2} \, \mathfrak{z} \tag{38}$$

with $\mathfrak{z} \sim \eta_0$; the proposal is accepted with probability $\min\big(1, \ell(y_{1:n};\widetilde{u})/\ell(y_{1:n};u)\big)$. Proposal (38) preserves the prior Gaussian distribution (34) for any $\rho \in (0,1)$ and the above Markov transition is well-defined on the infinite-dimensional space U. It follows that the method is robust upon mesh-refinement in the sense that ρ does not need to be adjusted as $k_{\max}$ increases [19]. In contrast, for standard Random-Walk Metropolis proposals, one would have to pick a smaller step-size upon mesh-refinement; for the optimal step-size, the mixing time will typically deteriorate as $\mathcal{O}(k_{\max}^2)$, see e.g. [3]. Still, proposal (38) can be inefficient when targeting the posterior distribution η when it differs significantly from the prior distribution η_0. Indeed, *a-priori* the Fourier coefficients u_k have known scales appropriately taken under consideration in (38); *a-posteriori*, information from the data spreads non-uniformly on the Fourier coefficients, with more information being available for low frequencies than for high ones. Taking a glimpse into results from the execution of the adaptive SMC algorithm yet to be defined, in Figure 1 we plot the fractions, as estimated by the SMC method, between posterior and prior standard deviations for the Fourier coefficient $\mathrm{Re}(u_k)$ (left panel) and $\mathrm{Im}(u_k)$ (right panel) over all pairs of frequencies $k=(k_1,k_2)$ with $-20 \le k_1, k_2 \le 20$. In this case it is apparent that most of the information in the data concentrates on a window of frequencies around the origin; still there is a large number of variables (around $2\cdot 10^2$ in this example) which have diverse posterior standard deviations under the posterior distribution. The standard deviations of these Fourier coefficients can potentially be very different from their prior standard deviations.

The approach followed in [18] for constructing better-mixing Markov kernels involves selecting a 'window' of frequencies $\mathbf{K} = \big\{k \in \mathbb{Z}_*^2 \,:\, \max(k_1,k_2) \le K\big\}$, for a user pre-specified threshold $K \ge 1$, and using the following Markov mutation steps within an SMC algorithm.

- Use the currently available particles approximation $\{u^i\}_{i=1}^N$ of η_n to estimate the current marginal mean and covariance $\mathfrak{m}_k^N$ and Σ_k^N of the two-dimensional variable $u_k = \big(\mathrm{Re}(u_k), \mathrm{Im}(u_k)\big)$ over the window $k=(k_1,k_2) \in \mathbf{K} \cap \mathbb{Z}_\uparrow^2$,

  $$\mathfrak{m}_k^N = \tfrac{1}{N}\sum_{i=1}^N u_k^i\ ; \quad \Sigma_k^N = \tfrac{1}{N-1}\sum_{i=1}^N (u_k^i - \mathfrak{m}_k^N)\otimes(u_k^i - \mathfrak{m}_k^N)\ .$$

  For high-frequencies $k=(k_1,k_2) \in \mathbf{K}^c \cap \mathbb{Z}_\uparrow^2$, only the information contained in the prior distribution is used and we thus set $\mathfrak{m}_k^N = 0$ and $\Sigma_k^N = \frac{1}{2}\,|k|^{-2\alpha}\, I_2$.

- For a current position $u = \sum u_k\,\psi_k$, the proposal $\widetilde{u} = \sum \widetilde{u}_k\,\psi_k$ is defined as

  $$\widetilde{u}_k = \mathfrak{m}_k^N + \rho\,(u_k - \mathfrak{m}_k^N) + (1-\rho^2)^{1/2}\,\mathfrak{z}_k$$

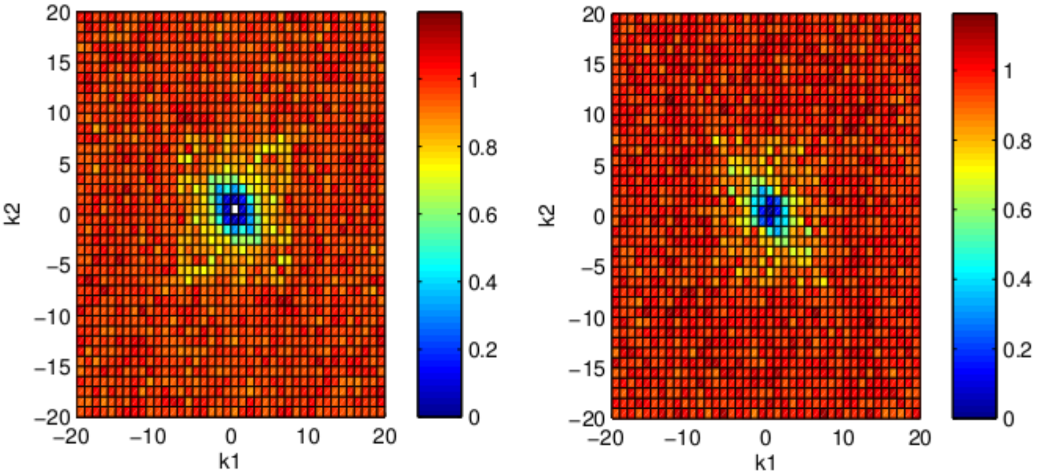

Figure 1: Ratio of (estimated) posterior vs prior standard deviations for $\mathrm{Re}(u_k)$ (left panel) and $\mathrm{Im}(u_k)$ (right panel) over all pairs $k = (k_1, k_2)$ with $-20 \le k_1, k_2 \le 20$. The model here corresponds to: $\delta = 0.2$, $m = 4$, $T = 20$, $\alpha = 2$, $\beta^2 = 5$, $\varepsilon^2 = 0.2$, $f(x) = \nabla^{\perp} \cos((5,5)' \cdot x)$. The $m = 4$ observation locations were at $(0, \pi)$, $(\pi, 0)$, $(0, 0)$, (π, π). Samples from the posterior were generated by applying a version of the adaptive SMC algorithm described in Section 6.2.2 for $K = 7$, see [18] for full details. The 'true' initial condition was sampled from the prior; data were then simulated accordingly.

for $k \in \mathbb{Z}^2_{\uparrow}$ and $\mathfrak{z}_k \sim \mathcal{N}(0, \Sigma^N_k)$ and $\widetilde{u}_{-k} = -\overline{\widetilde{u}_k}$ for $\mathbb{Z}^2_* \setminus \mathbb{Z}^2_{\uparrow}$; this proposal is accepted with the relevant Metropolis-Hastings ratio.

- In addition to the above adaptation at the Markov kernel, the analytical algorithm also involved an annealing step as described in Section 5, whereby additional intermediate distributions were introduced, if needed, in between any pairs η_{n-1}, η_n. We found this to be important for avoiding weight degeneracy and getting a stable algorithm. As explained in Section 5, the choice of temperatures was determined on the fly, according to a minimum requirement of the effective sample size (we choose $\alpha = \frac{1}{3}$).

It is important to note that in this Navier-Stokes setting, the regularity assumptions adopted in the theoretical parts of this article for the derivation of the asymptotic results do not apply anymore. As illustrated by this numerical analysis, the asymptotic behaviour predicted in Theorem 4.4 is likely to hold in far more general contexts. Figure 2 shows a plot of an estimate of the variance of $Z^N(y_{1:n})/Z(y_{1:n})$, where $Z^N(y_{1:n})$ is the N-particle

particle approximation of normalisation constant $Z^N(y_{1:n})$, as a function of the amount of data n for an adaptive SMC algorithm using $N = 500$ particles. In this complex setting, the numerical results seem to confirm the theoretical asymptotic results of Theorem 4.4: the estimated asymptotic variance seems to grow linearly with n, as one would have expected to be true for the perfect SMC algorithm that does not use adaptation. This is an indication that Theorem 4.4 is likely to hold under weaker assumptions than adopted in this article.

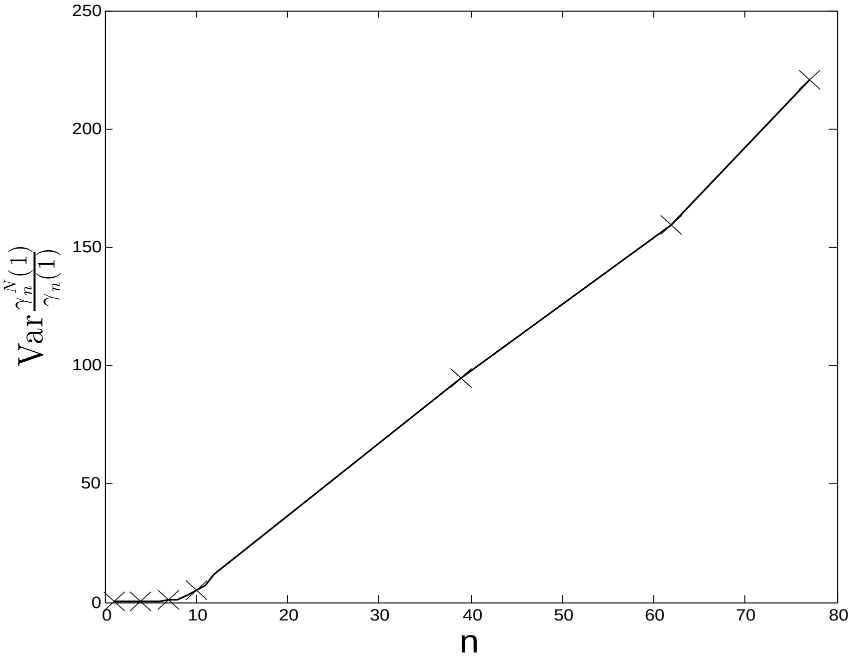

Figure 2: Estimated variance for the estimate of the normalizing constant of adaptive SMC. The 'true' normalizing constant was estimated from 1000 independent runs with $N = 500$ and the relative variance is estimated when $N = 500$ over 500 independent runs. The crosses are the estimated values of the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

6.2.3 Algorithmic Stability in Large Scale Adaptation

When the dimension d of the adapted statistics is large, as in the Navier-Stokes case (in our simulation study $d = \mathbf{Card}(\mathbf{K} \cap \mathbb{Z}^2_{\uparrow}) \times 5 \approx [(2K)^2/2] \times 5 \approx 500$) and potentially in other scenarios, it is certainly of interest to quantify the effect of the dimensionality d of the adaptive statistics on the overall accuracy of the SMC estimators. We will make a first modest attempt to shed some light on this issue via the consideration of a very simple modelling structure motivated by the Navier-Stokes example and allowing for some simple calculations.

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

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

for a given sequence of variances $\{\sigma_j^2\}_{j=1}^\infty$ that does not depend on the index $n \geq 1$. This represents an optimistic case where the incremental weights $G_n(x)$ are small enough to be irrelevant for the study of the influence of the dimension d; we set $G_n(x) \equiv 1$. It is assumed that the SMC method has worked well up-to time $(n-1)$ and has produced a collection of i.i.d. samples $\{x_{n-1}^i\}_{i=1}^N$ from η_{n-1}. For the mutation step, we consider an adaptive Metropolis-Hastings Markov 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

$$\begin{aligned} \widetilde{x}_j &= \rho\, x_j + (1-\rho^2)^{1/2}\, \mathcal{N}(0, \widehat{\sigma}_j^2) \ , \quad \text{for} \quad 1 \leq j \leq d \ , \\ \widetilde{x}_j &= \rho\, x_j + (1-\rho^2)^{1/2}\, \mathcal{N}(0, \sigma_j^2) \ , \quad \text{for} \quad j \geq d+1 \ , \end{aligned} \tag{39}$$

where we have set $\widehat{\sigma}_j^2 := (1/N)\sum_{i=1}^N \{x_{n-1,j}^i\}^2$. This corresponds to the adaptive SMC approach described in Section 2 with a d-dimensional adaptive statistics $\xi_n(x) = (x_1^2, \ldots, x_d^2)$. Thus, the d first coordinates of the proposal are adapted to the estimated marginal variance while the ideal variance is used for the remaining coordinates. We want to investigate the effect of the amount of adaptation on the accuracy of the estimator $\eta_n^N(\varphi)$ for a bounded function φ that only depends on the $(d+1)$-th coordinate,

$$\varphi(x) = \varphi(x_{d+1}) \ .$$

Notice that in this simple scenario the Metropolis-Hastings proposal corresponding to the ideal kernel $M_{n,\eta_{n-1}(\xi_n)}$ preserves η_n and is thus always accepted; under the ideal kernel, the particles at time n would still be a set of N i.i.d. samples from η_n. Consequently, any deviation from the $\mathcal{O}(N^{-1/2})$ rate of convergence for the estimator $\eta_n^N(\varphi)$ will be solely due to the effect of the adaptation.

We now investigate in this context the behavior of the difference $\eta_n^N(\varphi) - \eta_n(\varphi)$. Following the proof of Theorem 4.1 we use the decomposition

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

where, using the notations of Section 2, we have set $A(N) = [\eta_n^N - \Phi_{n,N}(\eta_{n-1}^N)](\varphi)$, $B_1(N) = \eta_{n-1}^N[Q_{n,N} - Q_n](\varphi)$ and $B_2(N) = [\eta_{n-1}^N - \eta_{n-1}](Q_n\varphi)$. Denoting by $\|\cdot\|_2$ the L_2-norm of random variables and conditioning upon $\mathcal{F}_{n-1}^N$, we have that

$$\|A(N)\|_2^2 = \tfrac{1}{N}\, \mathbb{E}\left[\, \mathrm{Var}\left[\, \varphi(x_n^1) \,|\, \mathcal{F}_{n-1}^N \,\right] \,\right] = \mathcal{O}(\tfrac{1}{N}) \ . \tag{40}$$

For $B_2(N)$ one can notice that $Q_n(\varphi)$ is a bounded mapping from $\mathbb{R}^\infty$ to $\mathbb{R}$, thus

$$\|B_2(N)\|_2^2 = \tfrac{1}{N}\,\mathrm{Var}_{\eta_{n-1}}\,[\,Q_n(\varphi)\,] = \mathcal{O}(\tfrac{1}{N})\ . \tag{41}$$

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 (16) in the proof of Theorem 4.2 yields

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

for a residual random variable R. Controlling the residual term in the above expansion poses enormous technical challenges and we restrict our analysis to the main order term $\widetilde{B}_1(N)$.

Proposition 6.2. *The term $\widetilde{B}_1(N)$ satisfies*

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

Proof. See the Appendix. □

Proposition 6.2 combined with (40)-(41) 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 maybe explain the fairly accurate SMC estimates of the marginal expectation obtained in the Navier-Stokes example when $N = 500$ and $d \approx 500$; we refer the reader to [18] for further simulation studies.

7 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 settings. There are several extensions to the work in this article. First, one could relax the boundedness assumptions used in the paper; our proof technique, also used in [6], is particularly amenable to this. Second, an approach to deal with the random stopping of some adaptive SMC algorithms (see Section 5) also needs to be developed. Lastly, one can extend the analysis to the context of adaptive resampling.

Acknowledgements

AB and AT were supported by a Singapore MOE grant. AJ was supported by Singapore MOE grant R-155-000-119-133 and is also affiliated with the risk management institute at the National University of Singapore.

A Proof of Proposition 6.2

First of all, notice that without loss of generality we can assume that $\sigma_j^2 = const.$. We have that:

$$\widetilde{B}_1(N) = \frac{\sqrt{d}}{N} \times \sum_{j=1}^{d} \Big\{ \tfrac{\sum_{i=1}^{N} \partial_{\xi_j} M_{n,\xi}(\varphi)(x_{n-1}^i)|_{\xi=\eta_{n-1}(\xi_n)}}{\sqrt{N}} \cdot \sqrt{N}\,(\eta_{n-1}^N - \eta_{n-1})(\xi_{n,j}) \Big\} / \sqrt{d}$$

$$\equiv \frac{\sqrt{d}}{N} \times \sum_{j=1}^{d} \big[\sqrt{N}\,\eta_{n-1}^N(\bar{\Xi}_{n,j}) \cdot \sqrt{N} \eta_{n-1}^N(\bar{\xi}_{n,j}) \big] / \sqrt{d} \qquad (42)$$

where we have set $\bar{\Xi}_{n,j}(x) = \partial_{\xi_j} M_{n,\xi}(\varphi)(x)|_{\xi=\eta_{n-1}(\xi_n)}$ and $\bar{\xi}_{n,j}(x) = \xi_{n,j}(x) - \eta_{n-1}(\xi_{n,j})$. Clearly, the expectation of the latter variable over η_{n-1} is zero, but the same is also true for the former one. Initially, we will focus on the term $\bar{\Xi}_{n,j}(x)$ 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:

$$M_{n,\xi_j}(\varphi)(x) = \mathbb{E}\,[\,\varphi(x_{d+1}')\,|\,x\,] = \varphi(x_{d+1}) + \mathbb{E}\,[\,a(x_j,\xi_j,Z_j)\,|\,x_j\,]\,\Delta\varphi(x_{d+1}) \qquad (43)$$

where we have set $\Delta\varphi(x_{d+1}) = \mathbb{E}\,[\,\varphi(x_{d+1}') - \varphi(x_{d+1})\,|\,x_{d+1}\,]$; x_{d+1}' denotes the Metropolis-Hastings proposal for the $(d+1)$-th co-ordinate as specified in (39) ; $a(x_j,\xi_j,Z_j)$ denotes the Metropolis-Hastings acceptance probability which depends only on the current position x_j, the (arbitrary) scaling choice ξ_j and the noise $Z_j \sim \mathcal{N}(0,1)$ for simulating the proposal for the j-th co-ordinate assuming a scaling ξ_j (that is, we have $x_j' = \rho x_j + \sqrt{1-\rho^2}\,\xi_j^{1/2}\,Z_j$). 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 co-ordinate, we have that $\mathbb{E}_{\eta_{n-1}}\,[\,\Delta\varphi(x_{d+1})\,] = 0$. Recall that $\bar{\Xi}_{n,j}(x) = \partial_{\xi_j} M_{n,\xi_j}(\varphi)(x)|_{\xi_j=\eta_{n-1}(\xi_{n,j})}$, thus to check for the differentiability of the mapping $\xi_j \mapsto \mathbb{E}\,[\,a(x_j,\xi_j,Z_j)\,|\,x_j\,]$ we can only resort to analytical calculations, starting from the fact that (after some algebraic manipulations):

$$a(x_j,\xi_j,Z_j) = 1 \wedge \exp\Big\{ -\tfrac{1}{2}\big(\xi_j^{-1} - \sigma_j^{-2}\big)\big(x_j^2 - \{\rho\, x_j + \sqrt{1-\rho^2}\,\xi_j^{1/2} Z_j\}^2\big)\Big\}\ .$$

After a lot of cumbersome analytical calculations (which are omitted for brevity) we can integrate out Z_j and find that i) the derivative $D(x_j,\eta_{n-1}(\xi_{n,j})) = \partial_{\xi_j}\mathbb{E}\,[\,a(x_j,\xi_j,Z_j)\,|\,x_j\,]|_{\xi_j=\eta_{n-1}(\xi_{n,j})}$

exists; ii) $D(x_j,\eta_{n-1}(\xi_{n,j}))$, with $x_j\sim\mathcal{N}(0,\sigma_j^2)$, has a finite second moment. Thus, continuing from (43) we have:

$$\bar{\Xi}_{n,j}(x) = \partial_{\xi_j} M_{n,\xi_j}(\varphi)(x)|_{\xi_j=\eta_{n-1}(\xi_{n,j})} = D(x_j,\eta_{n-1}(\xi_{n,j}))\,\Delta\varphi(x_{d+1})\ . \tag{44}$$

The factorisation in (44) will be exploited in the remaining calculations.

Continuing from (42), we now have that:

$$\begin{aligned}\|\tfrac{N}{\sqrt{d}}\widetilde{B}_1(N)\|_2^2 &= \tfrac{1}{d}\sum_{j=1}^{d} N^2\,\mathbb{E}\left[\,\{\eta_{n-1}^N(\bar{\Xi}_{n,j})\}^2\,\{\eta^N(\bar{\xi}_{n,j})\}^2\right] \\ &\quad + \tfrac{1}{d}\sum_{\substack{j,k=1,2,\ldots,d\\ j\neq k}} N^2\,\mathbb{E}\left[\,\eta_{n-1}^N(\bar{\Xi}_{n,j})\,\eta_{n-1}^N(\bar{\xi}_{n,j})\,\eta_{n-1}^N(\bar{\Xi}_{n,k})\,\eta_{n-1}^N(\bar{\xi}_{n,k})\,\right] \\ &=: T_1+T_2\ . \end{aligned} \tag{45}$$

The following zero-expectations obtained for terms involved in T_1, T_2 are a direct consequence of the fact that $\bar{\xi}_{n,j}(x)$ only depends on x_j and has zero expectation under η_{n-1}, and that $\bar{\Xi}_{n,j}(x)$ only depends on x_j, x_{d+1} through the product form in (44) 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 $\mathbb{E}\left[\,\{\eta_{n-1}^N(\bar{\Xi}_{n,j})\}^2\,\{\eta_{n-1}^N(\bar{\xi}_{n,j})\}^2\right]$ 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 the ones that involve cross-products of the form $\{\bar{\Xi}_{n,j}(x_{n-1}^i)\}^2\times\{\bar{\xi}_{n,j}(x_{n-1}^{i'})\}^2$, thus:

$$T_1 = \tfrac{1}{d}\sum_{j=1}^{d} N^2\cdot\tfrac{1}{N^4}\cdot\mathcal{O}(N^2) = \mathcal{O}(1)\ . \tag{46}$$

Then, moving on to the T_2-term, notice that all 4-way products in the expectation term $\mathbb{E}\left[\,\eta_{n-1}^N(\bar{\Xi}_{n,j})\,\eta_{n-1}^N(\bar{\xi}_{n,j})\,\eta_{n-1}^N(\bar{\Xi}_{n,k})\,\eta_{n-1}^N(\bar{\xi}_{n,k})\,\right]$ have expectation 0, except for the products involving the same particles $\bar{\Xi}_{n,j}(x_{n-1}^i)\,\bar{\xi}_{n,j}(x_{n-1}^i)\,\bar{\Xi}_{n,k}(x_{n-1}^i)\,\bar{\xi}_{n,k}(x_{n-1}^i)$. Thus, we have that:

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

Thus, overall we have that:

$$\|\widetilde{B}_1(N)\|_2 = \mathcal{O}(\tfrac{\sqrt{d}}{N}) + \mathcal{O}(\tfrac{d}{N^{3/2}})\ . \tag{47}$$

Results (46), (47), used within (45) complete the proof.

References

[1] Andrieu, C., & Moulines É. (2006). On the ergodicity properties of some adaptive MCMC algorithms. *Ann. Appl. Prob.*, **16**, 1462–1505.

[2] Beskos, A., Crisan, D. & Jasra, A. (2014). On the stability of sequential Monte Carlo methods in high dimensions. *Ann. Appl. Probab.* (to appear).

[3] Beskos, A., Roberts, G. & Stuart, A. (2009). Optimal scalings for local Metropolis-Hastings chains on nonproduct targets in high dimensions, *Ann. Appl. Probab.*, **19**, 863–898.

[4] Cérou, F., Del Moral, P. & Guyader, A. (2011). A non-asymptotic variance theorem for unormalized Feynman-Kac particle models. *Ann. Inst. Henri Poincare*, **47**, 629–649.

[5] Chopin, N. (2002). A sequential particle filter for static models. *Biometrika*, **89**, 539–552.

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

[7] Cotter, S., Roberts, G., Stuart, A. & White, D. (2013). MCMC methods for functions: modifying old algorithms to make them faster. *Statist. Sci.*, **28**, 424-446.

[8] Crisan, D. & Doucet, A. (2000). Convergence of sequential Monte Carlo methods. Technical Report CUED/F-INFENG/, University of Cambridge.

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

[10] Del Moral, P., Doucet, A. & Jasra, A. (2006). Sequential Monte Carlo samplers. *J. R. Statist. Soc. B*, **68**, 411–436.

[11] Del Moral, P., Doucet, A. & Jasra, A. (2012). On adaptive resampling strategies for sequential Monte Carlo methods. *Bernoulli*, **18**, 252–278.

[12] Del Moral, P., Doucet, A. & Jasra, A. (2012). An adaptive sequential Monte Carlo method for approximate Bayesian computation. *Statist. Comp.*, **22**, 1009–1020.

[13] Doering, C. & Gibbon, J. (1995). *Applied Analysis of the Navier-Stokes Equations.* Cambridge University Press: Cambridge.

[14] Doucet, A. & Johansen, A. (2011). A tutorial on particle filtering and smoothing: Fifteen years later. In *Handbook of Nonlinear Filtering* (eds. D. Crisan et B. Rozovsky), Oxford University Press: Oxford.

[15] Gelman, A., & Meng, X. L. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science*, 163-185.

[16] Giraud, F., Minvielle, P. & Del Moral, P. (2013). Non-asymptotic analysis of adaptive and annealed Feynman-Kac particle models. arXiv:1209.5654 [math.PR].

[17] Jasra, A., Stephens, D. A., Doucet, A. & Tsagaris, T. (2011). Inference for Lévy driven stochastic volatility models using adaptive sequential Monte Carlo. *Scand. J. Statist.*, **38**, 1–22.

[18] Kantas, N., Beskos, A., & Jasra, A. (2013). Sequential Monte Carlo for inverse problems: a case study for the Navier Stokes equation. arXiv:1307.6127 [stat.CO].

[19] Pillai, N. S., Stuart, A. M., & Thiery, A. H. (2014). Gradient Flow from a Random Walk in Hilbert Space, *Stochastic Partial Differential Equations: Analysis and Computations* (to appear).

[20] Schäfer, C. & Chopin, N. (2013). Adaptive Monte Carlo on binary sampling spaces. *Statist. Comp.*, **23**, 163–184.