Error Bounds and Normalizing Constants for Sequential Monte Carlo Samplers in High Dimensions

BY ALEXANDROS BESKOS[1], DAN CRISAN[2], AJAY JASRA[3] & NICK WHITELEY[4]

[1]Department of Statistical Science, University College, London, WC1E 6BT, UK.

E-Mail: *a.beskos@ucl.ac.uk*

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

E-Mail: *d.crisan@ic.ac.uk*

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

E-Mail: *staja@nus.edu.sg*

[4]Department of Mathematics, University of Bristol, Bristol, BS8 1TW, UK

E-Mail: *nick.whiteley@bristol.ac.uk*

Abstract

In this article we develop a collection of results associated to the analysis of the Sequential Monte Carlo (SMC) samplers algorithm, in the context of high-dimensional i.i.d. target probabilities. The SMC samplers algorithm can be designed to sample from a single probability distribution, using Monte Carlo to approximate expectations w.r.t. this law. Given a target density in $d-$dimensions our results are concerned with $d \to \infty$, while the number of Monte Carlo samples, N, remains fixed. We deduce an explicit bound on the Monte-Carlo error for estimates derived using the SMC sampler and the exact asymptotic relative $\mathbb{L}_2$-error of the estimate of the normalizing constant associated to the target. We also establish marginal propagation of chaos properties of the algorithm. These results are deduced when the cost of the algorithm is $\mathcal{O}(Nd^2)$.

Key words: Sequential Monte Carlo, High Dimensions, Propagation of Chaos, Normalizing Constants.

1 Introduction

High-dimensional probability distributions are increasingly of interest in a wide variety of applications. In particular, one is concerned with the estimation of expectations with respect to such distributions. Due to the high-dimensional nature of the probability laws, such integrations cannot typically be carried out analytically; thus practitioners will often resort to Monte Carlo methods.

An important Monte Carlo methodology is Sequential Monte Carlo samplers (see [11, 20]). This is a technique designed to approximate a sequence of densities defined on a common state-space. The method works by simulating a collection of $N \geq 1$ weighted samples (termed particles) in parallel. These particles are propagated forward in time via Markov chain Monte Carlo (MCMC), using importance sampling (IS) to correct, via the weights, for the discrepancy between target distributions and proposals. Due to the weight degeneracy problem (see e.g. [15]), resampling is adopted, sometimes performed when the Effective Sample Size (ESS) drops below some threshold. Resampling generates samples with replacement from the current collection of particles using the importance weights, resetting un-normalized weights to 1 for each sample. The ESS is a number between 1 and N and indicates, approximately, the number of useful samples. For SMC samplers one is typically interested in sampling a single target density on $\mathbb{R}^d$, but due to some complexity, a collection of artificial densities are introduced, starting at some easy to sample distribution and creating a smooth path to the final target.

Recently ([3, 5, 26]) it was shown that some IS methods will not stabilize, in an appropriate sense, as the dimension of target densities in a particular class grows, unless N grows exponentially fast with dimension d. In later work, [4] have established that the SMC sampler technique for *i.i.d.* targets can, in some scenarios, be stabilized at a cost that is only polynomial in d. It was shown in [4] that ESS and the Monte Carlo error of fixed dimensional marginals stabilize as d grows, with a cost of $\mathcal{O}(Nd^2)$. This corresponds to introducing d artificial densities between an initial distribution and the one of interest. The case of fixed dimension (d) also has been analyzed recently [27], while other results associated to the dimension can be found in [16, 25].

The objective of this article is to provide a more complete understanding of SMC samplers in high dimensions, complementing and building upon the results of [4]. A variety of results are presented, addressing some generic theoretical properties of the algorithms and some issues which arise from specific classes of applications. As stated above, the results in [4] only hold for i.i.d. targets and further work is required to investigate more general model structures; in this article we focus explicitly only on i.i.d. target distributions.

1.1 Problems Addressed

The first issue investigated is the increase in error of estimating fixed-dimensional marginals using SMC samplers relative to i.i.d. sampling. Considering the case when one resamples at the very final time-step we show that the $\mathbb{L}_2$-error increases only by a factor of $\mathcal{O}(N^{-1})$ uniformly in d. This result helps to establish the uniform in d performance of the algorithm, coupled, with a term that describes the additional loss of accuracy due to the dependence of the particles (relative to i.i.d. sampling). Resampling at the very final time-step is often of importance in real applications; see e.g. [13].

The second issue we address is the estimation of ratios of normalizing constants approximated using SMC samplers. This is critical in many disciplines, including Bayesian or classical statistics, physics and rare events. In particular, for Bayesian model comparison, Bayes factors are associated to statistical models in high-dimensional spaces, and these Bayes factors need to be estimated by numerical techniques such as SMC. The normalizing constant in SMC methods has been well-studied: see [8, 10]. Among the interesting results that have been proved in the literature is the unbiasedness property. However, to our knowledge, no results have been proved in the context of asymptotics in dimension d. In this article we provide an expression for fixed N of the relative $\mathbb{L}_2$-error of the SMC estimate of a ratio of normalizing constants. The algorithm can include resampling, whereby the expression differs. The rate of convergence is $\mathcal{O}(N^{-1})$, when the computational cost is $\mathcal{O}(Nd^2)$. The results also allow us compare between different sequences of densities used within the SMC method.

The third issue we investigate is asymptotic independence properties of the particles when one resamples: propagation of chaos - see [10, Chapter 8]. This issue has practical implications which we discuss below. It is shown that, in between any two resampling times, any fixed dimensional marginal distributions over a fixed block of $1 \leq q \leq N$ particles among the N particles are asymptotically independent with the correct marginal. This result is established as d grows with N fixed, whilst the classical results require N to grow. As in [4, 27], this establishes that the ergodicity of the Markov kernels used in the algorithm can provide stability in high dimensions if the number of artificial densities is scaled appropriately with d.

The paper is structured as follows: In Section 2 we describe the SMC sampler algorithm together with our mathematical assumptions. In Section 3 our main results are given. In addition, we introduce a general annealing scheme, coupled with a consideration of stability results for data-point tempering [9]. Section 4 considers the practical implications of our main results with numerical simulations. We conclude with a summary in Section 5. Most of the proofs of our results are given in the Appendix.

1.2 Notation

Let $(E, \mathscr{E})$ be a measurable space and $\mathscr{P}(E)$ the set of probability measures on it. For μ a $\sigma-$finite measure on $(E, \mathscr{E})$ and f a measurable function, we set $\mu(f) = \int_E f(x)\mu(dx)$. For $\mu \in \mathscr{P}(E)$ and P a Markov kernel on $(E, \mathscr{E})$, we use the integration notation $P(f)(x) = \int_E P(x, dy)f(y)$ and $\mu P(f) = \int_E \mu(dx)P(f)(x)$. In addition, $P^n(f)(x) := \int_{E^{n-1}} P(x, dx_1)P(x_1, dx_2) \times \cdots \times P(f)(x_{n-1})$. The total variation difference norm for $\mu, \lambda \in \mathscr{P}(E)$ is $\|\mu - \lambda\|_{tv} := \sup_{A \in \mathscr{E}} |\mu(A) - \lambda(A)|$. The class of bounded (resp. continuous and bounded) measurable functions $f : E \to \mathbb{R}$ is written $\mathcal{B}_b(E)$ (resp. $\mathcal{C}_b(E)$). For $f \in \mathcal{B}_b(E)$, we write $\|f\|_\infty := \sup_{x \in \mathbb{R}} |f(x)|$. We will denote the $\mathbb{L}_\varrho$-norm of random variables as $\|X\|_\varrho = \mathbb{E}^{1/\varrho} |X|^\varrho$ with $\varrho \geq 1$. For a given vector $(x_1, \ldots, x_p)$ and $1 \leq q \leq s \leq p$ we denote by $x_{q:s}$ the sub-vector $(x_q, \ldots, x_s)$. For a measure μ the N-fold product is written $\mu^{\otimes N}$. For any collection of functions $(f_k)_{k \geq 1}$, $f_k : E \to \mathbb{R}$, we write $f_1 \otimes \cdots \otimes f_k : E^k \to \mathbb{R}$ for their tensor product. Throughout M is used to denote a constant whose meaning may change, depending upon the context; important dependencies are written as $M(\cdot)$. In addition, all of our results hold on probability space $(\Omega, \mathscr{F}, \mathbb{P})$, with $\mathbb{E}$ denoting the expectation operator and $\mathbb{V}$ar the variance. Finally, $(\Rightarrow)$ denotes convergence in distribution.

2 Framework

We first set-up the context within which we will derive our analytical results.

2.1 Algorithm and Set-Up

We consider the scenario when one wishes to sample from a target distribution with density Π on E^d ($E \subseteq \mathbb{R}$) with respect to Lebesgue measure, known point-wise up to a normalizing constant. In order to sample from Π, we introduce a sequence of 'bridging' densities which start from an easy to sample target and evolve toward Π. In

particular, we will consider the densities:

$$\Pi_n(x) \propto \Pi(x)^{\phi_n} , \quad x \in E^d , \tag{1}$$

for $0 \leq \phi_0 < \cdots < \phi_{n-1} < \phi_n < \cdots < \phi_p = 1$. Below, we use the short-hand Γ_n to denote un-normalized densities associated to Π_n.

One can sample from $\{\Pi_n\}$ using an SMC sampler that can be designed to target the sequence of densities:

$$\widetilde{\Pi}_n(x_{1:n}) = \Pi_n(x_n) \prod_{j=1}^{n-1} \frac{\Pi_{j+1}(x_j)K_{j+1}(x_j, x_{j+1})}{\Pi_{j+1}(x_{j+1})}$$

with domain $(\mathbb{R}^d)^n$ of dimension that increases with $n = 1, \dots, p$, Let $\{K_n\}$ be a sequence of Markov kernels of invariant density $\{\Pi_n\}$ and Υ an (unnormalised) density; assuming the weights appearing in the statement of the algorithm are well-defined Radon Nikodym derivatives, the SMC sampler we will ultimately explore is the one defined in Algorithm 1.

---

0. *Sample* $X_0^1, \dots X_0^N$ *i.i.d. from* Υ *and compute the weights for each particle* $i \in \{1, \dots, N\}$:

$$w_{0:0}^i = \frac{\Gamma_0(x_0^i)}{\Upsilon(x_0^i)} .$$

Set $n = 1$ *and* $l = 0$.

1. *If* $n \leq p$, *for each* i *sample* $X_n^i \mid X_{n-1}^i = x_{n-1}^i$ *from* $K_n(x_{n-1}^i, \cdot)$ *(i.e. conditionally independently) and calculate the weights:*

$$w_{l:n}^i = \frac{\Gamma_n(x_{n-1}^i)}{\Gamma_{n-1}(x_{n-1}^i)} w_{l:(n-1)}^i .$$

Calculate the Effective Sample Size (ESS):

$$ESS_{l:n}(N) = \frac{\left(\sum_{i=1}^N w_{l:n}^i\right)^2}{\sum_{i=1}^N (w_{l:n}^i)^2} . \tag{2}$$

If $ESS_{l:n}(N) < a$:
resample particles according to their normalised weights

$$\overline{w}_{l:n}^i = \frac{w_{l:n}^i}{\sum_{j=1}^N w_{l:n}^j} ; \tag{3}$$

set $l = n$ *and re-initialise the weights by setting* $w_{l:n}^i \equiv 1$, $1 \leq i \leq N$;
let $\check{x}_n^1, \dots, \check{x}_n^N$ *now denote the resampled particles and set* $(x_n^1, \dots, x_n^N) = (\check{x}_n^1, \dots, \check{x}_n^N)$
Set $n = n + 1$.
Return to the start of Step 1.

Algorithm 1: The SMC samplers algorithm. We consider multinomial resampling, so resampled particles are i.i.d. samples from the discrete law determined by the particles just before resampling and their weights.

For simplicity, we will henceforth assume that $\Upsilon \equiv \Gamma_0$. It should be noted that when Υ is different from Γ_0, one can modify the sequence of densities to a bridging scheme which moves from Υ to Γ_p (an appropriately modified version of our results in this paper are expected to hold in this scenario). However, in practice, one can make Π_0 as simple as possible so we do not consider this possibility; see [27] for more discussion and analysis when the dimension is fixed. Note, that we only consider here the multinomial resampling method.

We will investigate the stability of SMC estimates associated to the method in Algorithm 1. The actual analysis is performed on a modified version of Algorithm 1 which will not perform dynamic resampling described in Section 2.2. To obtain analytical results we will need to simplify the structure of the algorithm. In particular, we will consider an i.i.d. target:

$$\Pi(x) = \prod_{j=1}^d \pi(x_j) ; \quad \pi(x_j) \propto \exp\{g(x_j)\} , \tag{4}$$

with $x_j \in E$, for some $g : E \mapsto \mathbb{R}$. In such a case all bridging densities are also i.i.d.:

$$\Pi_n(x) \propto \prod_{j=1}^{d} \pi_n(x_j) ; \quad \pi_n(x_j) \propto \exp\{\phi_n \, g(x_j)\} .$$

Thus, in this context we will have:

$$\Gamma_n(x) = \exp\Big\{\phi_n \sum_{j=1}^{d} g(x_j)\Big\} .$$

It is remarked that this i.i.d. assumption is made for mathematical convenience, as it will allow for analytical tractability of asymptotic results as $d \to \infty$. This is exactly in agreement with the i.i.d. set-up followed for deriving powerful asymptotic results as $d \to \infty$ in the MCMC literature (see e.g. [22, 23] or the review in [24] and the references therein). Going beyond the i.i.d. context when investigating $d \to \infty$ in the more established MCMC literature requires a lot of technicalities for each particular modelling scenario (see e.g. [2, 6]) so we will leave this for future work for our SMC context. Using our result, some reasonable conjectures can still be made when looking at targets used in practice; see [4] for an extensive discussion on potential extensions of our results to non-i.i.d. contexts.

A further assumption that will facilitate the mathematical analysis is to apply (conditionally) independent kernels along the different co-ordinates. That is, we will assume:

$$K_n(x, dx') = \prod_{j=1}^{d} k_n(x_j, dx'_j) ,$$

where each transition kernel $k_n(\cdot, \cdot)$ preserves $\pi_n(x)$; that is, $\pi_n k_n = \pi_n$. We study the case when one selects cooling constants $\phi_n = \phi_n(d)$ and $p = p(d)$ as below:

$$p = d ; \quad \phi_n(= \phi_{n,d}) = \phi_0 + \tfrac{n(1-\phi_0)}{d} , \quad 0 \le n \le d , \tag{5}$$

with $0 \le \phi_0 < 1$ given and fixed with respect to d. It is possible, with only notational changes, to consider (as in [27]) the case when the annealing sequence is derived via a more general non-decreasing Lipschitz function; see Section 3.2.1. As in [4], it will be convenient to consider the continuum of invariant densities and kernels on the whole of the time interval $[\phi_0, 1]$. So, we will set:

$$\pi_s(x) \propto \pi(x)^s = \exp\{s \, g(x)\} , \quad s \in [\phi_0, 1] .$$

Similarly $k_s(x, dx')$ with $s \in (\phi_0, 1]$ is the continuous-time version of the kernels $k_n(x, dx')$. As in [4], the mapping $l_d(s) = \lfloor \frac{d\,(s - \phi_0)}{1 - \phi_0} \rfloor$ is used to move between continuous and discrete time.

2.2 Conditions

We state the conditions under which we will derive our results. We will require that $E \subset \mathbb{R}$ with E being *compact*. The conditions below correspond to a simplification of the weaker conditions in [4] under the scenario of the compact state space E that we consider here. We note that imposing compactness has been done mainly to simplify proofs and keep them at a reasonable length. The numerical examples later on are executed on unbounded state spaces, and do not seem to invalidate our conjecture that several of the results in the sequel will also hold on unbounded spaces under appropriate geometric ergodicity conditions, as it was the case for the stability results as $d \to \infty$ in [4]. We remark that all results of [4] also hold under the assumptions stated here.

(**A1**) *Stability of $\{k_s\}$ - Uniform Ergodicity.*

There exists a constant $\theta \in (0, 1)$ and some $\varsigma \in \mathscr{P}(E)$ such that for each $s \in (\phi_0, 1]$ and any $(x, A) \in E \times \mathscr{E}$

$$k_s(x, A) \ge \theta \, \varsigma(A).$$

(**A2**) *Perturbations of $\{k_s\}$.*

There exists an $M < \infty$ such that for any $s, t \in (\phi_0, 1]$ we have $\sup_{x \in E} \|k_s(x, \cdot) - k_t(x, \cdot)\|_{tv} \le M \, |s - t| .$

4

In the context of our analysis, we will consider an SMC algorithm that resamples at the *deterministic* times $t_1(d), \ldots, t_{m^*(d)}(d) \in [\phi_0, 1]$ (i.e. resamples after $n = l_d(t_k(d))$ steps for $k = 1, 2, \ldots, m^*(d)$) such that $t_0(d) = \phi_0$ and $t_0(d) < t_1(d) < \cdots < t_{m^*(d)}(d) < t_{m^*(d)+1}(d) = 1$, with $l_d(t_{m^*}(d)) < d$. We will also assume that as $d \to \infty$ we have that $m^*(d) \to m^*$ and $t_k(d) \to t_k$ for $t_k \in [\phi_0, 1]$ for all relevant k. Such deterministic times are meant to mimic the behaviour of randomised ones (i.e. as for the case of the original Algorithm 1 which involves dynamic resampling at stochastic times) and provide a mathematically more convenient framework (but still a very challenging one for deriving analytical results) than dynamic resampling for understanding the impact of resampling on the properties of the algorithm. Examples of particular specifications of such times so that they connect with the randomined ones can be found in [4, 12]; the results therein provide an approach for converting the results for deterministic times, to randomized ones. In particular, they show that with a probability converging to 1 as $N \to \infty$ the randomized times essentially coincide with the particular deterministic ones. We do not consider such results here as we will focus on other aspects of the algorithm, and point the readers to [4] for more details. For simplicity, we will henceforth assume that d is large enough so that $m^*(d) \equiv m^*$.

The analysis considered here makes rather strong assumptions. The strong mixing assumptions, which can hold for sampling problems on compact state-spaces, may lead to one thinking that SMC would not be needed in such scenarios. However, we argue that there are problems where one would use SMC samplers even if MCMC mixes extremely well (see Section 3.2.2) to which our analysis extends. In addition, the proofs and ideas here provides a first analysis on which to base proofs with weaker assumptions. We also remark that the use of weaker assumptions in the stability analysis of SMC methods is quite recent; see for example [4, 27, 28].

2.3 Log-Weight-Asymptotics

Given the set-up (5) and the resampling procedure at the deterministic times $t_1(d), \ldots, t_{m^*}(d) \in [\phi_0, 1]$, and due to the i.i.d. structure described above, we have the following expression for the particle weights:

$$\log(w^i_{l_d(t_{k-1}(d)):l_d(t_{k(d)})}) = \tfrac{1}{d} \sum_{j=1}^d \bar{G}^i_{k,j} \ ,$$

where $\bar{G}^i_{k,j} = (1 - \phi_0) \sum_{n=l_d(t_{k-1}(d))}^{l_d(t_k(d))-1} g(X^i_{n,j})$ for $1 \le i \le N$ and $X^i_{n,j}$ is the jth component of X^i_n. The work in [4] illustrates stability of the normalised weights as $d \to \infty$. Define the standardised log-weights:

$$G^i_{k,j} = (1 - \phi_0) \sum_{n=l_d(t_{k-1}(d))}^{l_d(t_k(d))-1} \big(g(X^i_{n,j}) - \mathbb{E}_{\pi_{t_{k-1}(d)}}[g(X^i_{n,j})] \big) \ . \tag{6}$$

The notation $\mathbb{E}_{\pi_{t_{k-1}(d)}}[g(X^i_{n,j})]$ refers to an expectation under the initial dynamics $X^i_{l_d(t_{k-1}(d)),j} \sim \pi_{t_{k-1}(d)}$; after that, $X^i_{n,j}$ will evolve according to the Markov transitions k_n. We also use the notation

$$\mathbb{E}_{\pi^{\otimes Nd}_{t_{k-1}(d)}}[\cdot]$$

when imposing similar initial dynamics, but now independently over all co-ordinates and particles; such dynamics differ of course from the actual particle dynamics of the SMC algorithm. In what follows, we use the Poisson equation:

$$g(x) - \pi_u(g) = \widehat{g}_u(x) - k_u(\widehat{g}_u)(x)$$

(i.e. $\widehat{g}_u$ is the solution of the Poisson equation associate to kernel k_u with invariant measure π_u - this exists under our assumptions) and in particular the variances:

$$\sigma^2_{s:t} = (1 - \phi_0) \int_s^t \pi_u(\widehat{g}_u^2 - k_u(\widehat{g}_u)^2) du \ , \quad \phi_0 \le s < t \le 1 \ . \tag{7}$$

The following weak limit can be derived from the proof of Theorem 4.1 of [4].

Remark 2.1 (Log-Weight-Asymptotics). *Assume (A1-2) and $g \in \mathcal{B}_b(E)$. For any $N \ge 1$ we have:*

$$\Big(\tfrac{1}{d} \sum_{j=1}^d G^i_{k,j} \Big)_{i=1}^N \Rightarrow (Z^i)_{i=1}^N \ ,$$

where the Z^i's are i.i.d. copies from $N(0, \sigma^2_{t_{k-1}:t_k})$.

The result illustrates that the consideration of $\mathcal{O}(d)$ Markov chain steps between resampling times stabilise the particle standardised log-weights as $d \to \infty$.

3   Main Results

We now present the main results of the article. The emphasis here will be to illustrate that Monte-Carlo errors for important estimates will be stable as $d \to \infty$, for fixed N.

3.1   Asymptotic Results as $d \to \infty$

The first result of the paper pertains to the Monte-Carlo error from estimates derived via the SMC method.

3.1.1 $\mathbb{L}_2-$Error

We will consider mean squared errors and obtain $\mathbb{L}_2$-bounds with resampling carried out also 'at the end', that is when one resamples also at time $t = 1$. Recall Algorithm 1 that the $\check{X}$-notation is used for resampled particles once they have been resampled. Resampling at time $t = 1$ is required when one wishes to obtain un-weighted samples. We have the following result, with proof in Appendix A.2.

Theorem 3.1. *Assume (A1-2) and $g \in \mathcal{B}_b(E)$. Then for any $N \geq 1$, $\varphi \in \mathcal{C}_b(E)$ we have:*

$$\lim_{d \to \infty} \left\| \left(\tfrac{1}{N} \sum_{i=1}^{N} \left[\varphi(\check{X}^i_{d,1}) - \pi(\varphi) \right] \right) \right\|_2^2 \leq \mathbb{V}\mathrm{ar}_\pi[\varphi] \, \tfrac{1}{N} \left(1 + M(\sigma^2_{t_{m*}:1}) \right)$$

for

$$M(\sigma^2_{t_{m*}:1}) = e^{\sigma^2_{t_{m*}:1}} + M \, e^{17\sigma^2_{t_{m*}:1}} \, \tfrac{1}{N^{1/6}}$$

with $M < \infty$ independent of N and $\sigma^2_{t_{m*}:1}$.

Remark 3.1. *Compared to an i.i.d. sampling scenario, the upper bound obtained in Theorem 3.1 contains the extra term $\mathbb{V}\mathrm{ar}_\pi[\varphi] \, \tfrac{1}{N} \, M(\sigma^2_{t_{m*}:1})$. This is is due to the dependence of the particles from resampling.*

3.1.2   Normalizing Constants

The second main result of the paper is the stability of estimating normalising constants in high dimensions. The quantity of interest here is the ratio of normalising constants:

$$c_d := \frac{\int_{E^d} \Gamma_d(x) dx}{\int_{E^d} \Gamma_0(x) dx} \ . \tag{8}$$

We first consider the SMC sampler in Algorithm 1 without the resampling step. Define:

$$\gamma_d^N(1) = \tfrac{1}{N} \sum_{i=1}^{N} w^i_{0:d} \equiv \tfrac{1}{N} \sum_{i=1}^{N} e^{\frac{1}{d} \sum_{j=1}^{d} \bar{G}^i_j} \ ; \quad \gamma_d(1) = \mathbb{E}\left[e^{\frac{1}{d} \sum_{j=1}^{d} \bar{G}^1_j} \right] \ ,$$

where $\bar{G}^i_j = (1 - \phi_0) \sum_{n=0}^{d-1} g(X^i_{n,j})$. By [10, Proposition 7.4.1] we have $\mathbb{E}\left[\gamma_d^N(1) \right] = \gamma_d(1) \equiv c_d$. Now, consider the relative $\mathbb{L}_2$-error:

$$\mathbb{V}_2(\gamma_d^N(1)) = \mathbb{E}\left[\left(\tfrac{\gamma_d^N(1)}{c_d} - 1 \right)^2 \right] \ .$$

We then have the following result, proven in Appendix A.3.

Theorem 3.2. *Assume (A1-2) and $g \in \mathcal{B}_b(E)$. Then for any $N \geq 1$:*

$$\lim_{d \to \infty} \mathbb{V}_2(\gamma_d^N(1)) = \frac{e^{\sigma^2_{\phi_0:1}} - 1}{N} \ .$$

The result establishes a $\mathcal{O}(N^{-1})$ rate of convergence at a computational cost of $\mathcal{O}(Nd^2)$. The information in the limit is in terms of the expression $\sigma^2_{\phi_0:1}$. As in [4], this is a critical quantity, which helps to measure the rate of convergence of the algorithm.

We now consider the SMC sampler in Algorithm 1 with resampling at the deterministic times $\{t_k(d)\}$ described in Section 2.2. We make the following definitions:

$$\gamma_{d,k}^N(1) = \tfrac{1}{N} \sum_{i=1}^{N} e^{\frac{1}{d} \sum_{j=1}^{d} \bar{G}^i_{k,j}} \ ; \quad \gamma_{d,k}(1) = \mathbb{E}_{\pi^{\otimes Nd}_{t_{k-1}(d)}}\left[e^{\frac{1}{d} \sum_{j=1}^{d} \bar{G}^1_{k,j}} \right] \ ,$$

where $k \in \{1, \ldots, m^*+1\}$ and $\bar{G}^i_{k,j}$ as defined in Section 2.3. As in the non-resampling case, we again have the unbiasedness property for the estimate of c_d:

$$\mathbb{E}\Big[\prod_{k=1}^{m^*+1} \gamma^N_{d,k}(1)\Big] = \prod_{k=1}^{m^*+1} \gamma_{d,k}(1) \equiv c_d .$$

We have the following result whose proof is in Appendix A.3.

Theorem 3.3. *Assume (A1-2) and $g \in \mathcal{B}_b(E)$. Then for any $N \geq 1$:*

$$\lim_{d\to\infty} \mathbb{V}_2\Big(\prod_{k=1}^{m^*+1} \gamma_{d,k}(1)\Big) = e^{-\sigma^2_{\phi_0:1}} \prod_{k=1}^{m^*+1} \Big[\tfrac{1}{N} e^{2\sigma^2_{t_{k-1}:t_k}} + \big(1 - \tfrac{1}{N}\big) e^{\sigma^2_{t_{k-1}:t_k}} \Big] - 1.$$

Remark 3.2. *It should be remarked that the estimation of normalizing constants without resampling as in Theorem 3.2 is usually not feasible as the asymptotic variance in the theorem can be extremely large. However, making an analytical comparison to Theorem 3.2 the limiting expression here depends upon the incremental variance expressions. On writing the limit in the form:*

$$\prod_{k=1}^{m^*+1} \Big[1 + \tfrac{1}{N}\{e^{\sigma^2_{t_{k-1}:t_k}} - 1\}\Big] - 1$$

if $N > (m^*+1)(e^{\bar{\sigma}^2} - 1)$, with $\bar{\sigma}^2 = \max_k \sigma^2_{t_{k-1}:t_k}$, then using the inequality $e^x \leq 1 + 2x$, $x \leq 1$, we have that:

$$\lim_{d\to\infty} \mathbb{V}_2\Big(\prod_{k=1}^{m^*+1} \gamma_{d,k}(1)\Big) \leq \tfrac{2(m^*+1)(e^{\bar{\sigma}^2} - 1)}{N} .$$

Contrasting this bound with the limit in the case without resampling in Theorem 3.2, it involves one term that increases linearly with the number of resampling times and an exponential term that decreases with the number of resampling times as the σ^2-terms are calculated over smaller time intervals. Following the proof in the Appendix, the stated limit in Theorem 3.3 is determined by the behavior of

$$\mathbb{E}_{\pi^{\otimes Nd}_{t_{k-1}(d)}}\big[\gamma^N_{d,k}(1)^2/\gamma_{d,k}(1)^2\big]$$

between resampling times. In effect, the ergodicity of the system takes over, and breaks up the error in estimation of the ratio of normalizing constants to different tours between resampling times (see Proposition 3.1).

To provide some intuition for Theorem 3.3, we give the following result. It is associated to the asymptotic independence, between resampling times, of the log-weights $\sum_{j=1}^d \frac{1}{d} G^i_{k,j}$ in (6). The proof is in Appendix A.3.

Proposition 3.1. *Assume (A1-2) and that $g \in \mathcal{B}_b(E)$. Then for any $N \geq 1$, $i \in \{1, \ldots, N\}$, $c_{1:k} \in \mathbb{R}$ and $k \in \{1, \ldots, m^*+1\}$, we have that:*

$$\lim_{d\to\infty} \mathbb{E}\big[e^{\sum_{l=1}^k c_l \frac{1}{d} \sum_{j=1}^d G^i_{l,j}}\big] = \prod_{l=1}^k \mathbb{E}\big[e^{c_l Z^l}\big] ,$$

where $Z^l \sim N(0, \sigma^2_{t_{l-1}:t_l})$ independently over $1 \leq l \leq k$.

3.1.3 Propagation of Chaos

Finally we deduce a rather classical result in the analysis of particle systems: propagation of chaos, which in the standard setting with fixed dimension d, refers to the asymptotic independence of any fixed block of q out of N particles as N grows. The following scenario, with the SMC sampler with resampling (at times $\{t_k(d)\}$), is considered: let $s(d)$ be a sequence such that $s(d) \in (t_{k-1}(d), t_k(d))$ for some $1 \leq k \leq m^*+1$, with limit $s \in (t_{k-1}(d), t_k(d))$. Denote by

$$\mathbb{P}^{(q,N)}_{s(d),j}$$

the marginal law of any of the q particles out of N at time $s(d)$ and in dimension $j \in \{1, \ldots, d\}$. By construction, particles are considered at a time when they are not resampled. We have the following Propagation of Chaos result, whose proof is in Appendix A.4.

Proposition 3.2. *Assume (A1-2) and that $g \in \mathcal{B}_b(E)$. Then, for any fixed $j \in \{1, \ldots, d\}$ and any $1 \leq k \leq m^* + 1$, a sequence $s(d) \in (t_{k-1}(d), t_k(d))$ with $s(d) \rightarrow s$, and any $N \geq 1$, with $1 \leq q \leq N$ fixed:*

$$\lim_{d \to \infty} \| \mathbb{P}^{(q,N)}_{s(d),j} - \pi_s^{\otimes q} \|_{tv} = 0 \ .$$

The result establishes the asymptotic independence of the marginals of the particles, between any two resampling times, as d grows. This is in contrast to the standard scenario where the particles only become independent as N grows. Critically, the MCMC steps provide the effect that the marginal particle distributions converge to the target $\pi_s^{\otimes q}$. Thus, Proposition 3.2 establishes that it is essentially the ergodicity of the system which helps to drive the stability properties of the algorithm. It should be noted that if one considers the particles just after resampling, one cannot obtain an asymptotic independence in d. Here, as in classical results for particles methods, one has to rely on increasing N.

3.2 Other Sequences of Densities

We now discuss some issues associated with the selection of the sequence of chosen bridging densities $\{\Pi_n\}$.

3.2.1 Annealing Sequence

Recall that we use the equidistant annealing sequence ϕ_n in (5). One could also consider a general differentiable, increasing Lipschitz function $\phi(s)$, $s \in [0, 1]$ with $\phi(0) = \phi_0 \geq 0$, $\phi(1) = 1$, and use the construction $\phi_{n,d} = \phi(n/d)$; then the asymptotic result in Remark 2.1 (and then the main results in Section 3) generalised to the choice of $\phi_{n,d}$ considered here would involve the variances:

$$\sigma^{2,\phi}_{s:t} = \int_s^t \pi_{\phi(u)}(\widehat{g}^2_{\phi(u)} - k_{\phi(u)}(\widehat{g}_{\phi(u)})^2) \left[\frac{d\phi(u)}{du} \right] d\phi(u) \ , \quad 0 \leq s < t \leq 1 \ , \tag{9}$$

in the place of $\sigma^2_{s:t}$ in (7). Our proofs in this paper are given in terms of the annealing sequence (5), corresponding to a linear choice of $\phi(\cdot)$, but it is straightforward to modify them to the above scenario.

This point is illuminated by our main results. For example Theorem 3.2 helps to compare various annealing schemes for estimating normalizing constants, via the limiting quantity (9). That is, if we are only concerned with variance, one would prefer an annealing scheme $\phi(s)$ giving a lower $(e^{\sigma^{2,\phi}_{0:1}} - 1)/N$, equivalently a lower $\sigma^{2,\phi}_{0:1}$. In practice, however, one has to numerically approximate $\sigma^{2,\phi}_{0:1}$ lessening the practical impact of this result. Similar considerations apply for the result in Theorem 3.3.

3.2.2 Data Point Tempering

An interesting sequence of densities introduced in [9] arises in the scenario when Π is associated with a batch data-set $y_1, \ldots, y_L$. The idea is to construct the sequence of densities so that arriving data-points are added sequentially to the target as the time parameter of the algorithm increases. More concretely, we will assume here that the target Π corresponds to the posterior distribution for the quantity of interest x:

$$\Pi(x) = p(x \mid y_1, y_2, \ldots, y_L) \propto p(x) p(y_1 \mid x) p(y_2 \mid y_1, x) \cdots p(y_L \mid y_{1:(L-1)}, x)$$

$$\propto \exp \left\{ \sum_{l=0}^{L} \sum_{j=1}^{d} g^{(l)}(y_{1:l}, x_j) \right\}, \quad x_j \in E \ ,$$

so that in the final expression we again impose a simplified context when all factors in the specification of Π have an isomeric product structure over the x_j's, via the consideration of the functionals $g^{(l)} = g^{(l)}(y_{1:l}, x_j)$; we use the convention that $y_{1:0} = \emptyset$. In this scenario one could then adopt a sequence of densities of the form:

$$\Pi^{(l)}(x) \propto \exp \left\{ \sum_{k=0}^{l} \sum_{j=1}^{d} g^{(k)}(y_{1:k}, x_j) \right\}, \quad 0 \leq l \leq L \ .$$

Clearly, for $d \to \infty$ one cannot stabilize the associated SMC algorithm (Algorithm 1) as applied on the sequence of targets $\Pi^{(0)}, \Pi^{(1)}, \ldots, \Pi^{(L)}$ as $\Pi^{(l+1)}/\Pi^{(l)}$ explodes for increasing d. To stabilize the algorithm as d grows one can insert $\lfloor d/L \rfloor$ annealing steps between consecutive data points, thus forming the densities:

$$\Pi^{(l)}_n(x) \propto \exp \left\{ \phi\left(\frac{n}{\lfloor d/L \rfloor} \right) \sum_{j=1}^{d} g^{(l)}(y_{1:l}, x_j) \right\} \times \Pi^{(l-1)}(x) \ , \quad l \in \{1, \ldots, L\}, \ n \in \{1, \ldots, \lfloor \frac{d}{L} \rfloor\} \ ,$$

where $\phi(s)$ is as in Section 3.2.1 with $\phi_0 = 0$. Then one can adopt Markov kernels $K_n^{(l)}$, $0 \leq l \leq L$, of product form with each component kernel $k_n^{(l)}$ having invariant measure

$$\pi_n^{(l)}(x) \propto \exp\left\{\phi\left(\frac{n}{\lfloor d/L \rfloor}\right) g^{(l)}(y_{1:l}, x) + \sum_{k=0}^{l-1} g^{(k)}(y_{1:k}, x)\right\}, \quad x \in E .$$

The continuous-index versions $\pi_s^{(l)}$ and $k_s^{(l)}$ are defined accordingly.

With a small abuse of notation, we consider now the application of Algorithm 1 upon the sequence of targets $\Pi_n^{(l)}$ defined above, with particles moved according to the kernels $K_n^{(l)}$. We consider the scenario where there is no resampling and denote here by $\mathrm{ESS}_{(0,d)}(N)$ the effective sample size after all data points have been incorporated in the sampler. Throughout the data are taken as fixed. We have the following result, which is a small generalisation of Theorem 3.1 of [4] and follows directly from it's proof; it illustrates stability of $\mathrm{ESS}_{(0,d)}(N)$ as $d \to \infty$. (For simplicity we write simply $g^{(l)}(x)$ instead of $g^{(l)}(y_{1:l}, x)$.)

Proposition 3.3. *Assume conditions (A1-2) for kernels $k_s^{(l)}$, $l \in \{1, \ldots, L\}$. Suppose that for each $l \in \{1, \ldots, L\}$ we have $g^{(l)} \in \mathcal{B}_b(E)$. Then, for any fixed $N > 1$ and $n \geq 1$, $\mathrm{ESS}_{(0,d)}(n, N)$ converges in distribution to:*

$$\frac{\left(\sum_{i=1}^N e^{Z^i}\right)^2}{\sum_{i=1}^N e^{2Z^i}}$$

where $Z^i \overset{i.i.d.}{\sim} N(0, \sigma^2)$ with

$$\sigma^2 = \sum_{l=1}^L \int_0^1 \pi_{\phi(u)}^{(l)}\left((\widehat{g}_{\phi(u)}^{(l)})^2 - k_{\phi(u)}^{(l)}(\widehat{g}_{\phi(u)}^{(l)})^2\right)\left[\frac{d\phi(u)}{du}\right]d\phi(u) ,$$

where $\widehat{g}_s^{(l)}$ is the solution to the Poisson equation $g^{(l)}(x) - \pi_s^{(l)}(g^{(l)}) = \widehat{g}_s^{(l)}(x) - k_s^{(l)}(\widehat{g}_s^{(l)})(x)$. In particular,

$$\lim_{d \to \infty} \mathbb{E}\left[\mathrm{ESS}_{(0,d)}(N)\right] = \mathbb{E}\left[\frac{\left(\sum_{i=1}^N e^{Z^i}\right)^2}{\sum_{i=1}^N e^{2Z^i}}\right] .$$

As for the case with annealing densities, there is a direct extension to the case where one resamples. In addition, one can easily extend the results in Section 3 in the data-point tempering case examined here. In connection to filtering, this is a class of densities that falls into the scenario of a state-space model with a deterministic dynamic on the hidden state (i.e. only the initial state is stochastic and propagated deterministically; see e.g. [7]). This is a scenario where one would want to use SMC methods, even if MCMC techniques work very well; this is due to the fact that the problem is sequential by its nature - see [18] for real examples in such a context.

4 Numerical Simulations

We now present two numerical examples, to illustrate the practical implications of our theoretical results. It is noted that the state-space E is not compact here, yet the impact of our results can still be observed.

4.1 Comparison of Annealing Schemes

We consider a target distribution comprised of d i.i.d. $N(0, 1)$ co-ordinates. The bridging densities are in this case:

$$\pi_{\phi(s)}(x) \propto \exp\left\{-\tfrac{1}{2}\phi(s)x^2\right\} . \tag{10}$$

Two annealing schemes are considered:

$$\phi(s) = \phi_0 + (1 - \phi_0)s ;$$
$$\nu(s) = \tfrac{\phi_0 e^{\vartheta} - 1}{e^{\vartheta} - 1} + \left(\tfrac{1 - \phi_0}{e^{\vartheta} - 1}\right)e^{\vartheta s} . \tag{11}$$

These are graphically displayed in Figure 1 (a), with $\vartheta = 5$.

The purpose of investigating the two annealing schemes is as follows. In practical applications of SMC samplers we have observed that algorithms with slow initial annealing schemes can often out-perform those with faster ones

(see Figure 1 (a)). Thus, we expect scheme $\nu(s)$ to perform better than $\phi(s)$ w.r.t. the expression for the asymptotic variance (9), hence deliver a lower relative $\mathbb{L}_2$-error for the estimation of the normalizing constant in high dimensions. To obtain some analytically computable proxies for the asymptotic variances (9) we use the variances that one would obtain when $k_s(x, dx') \equiv \pi_s(dx')$, that is we substitute $\pi_{\phi(u)}(g^2) - \pi_{\phi(u)}(g)^2$ for $\pi_{\phi(u)}(\widehat{g}_{\phi(u)}^2 - k_{\phi(u)}(\widehat{g}_{\phi(u)})^2)$ in (9). In this scenario it is simple to show that, under the choice (10), we have that:

$$\sigma_{\phi_0:1}^{2,\phi} = \frac{1}{2} \int_{\phi_0}^1 \left[\frac{1}{\phi(u)} \frac{d\phi(u)}{du} \right]^2 du \ .$$

Figure 1 (b) now plots the analytically available variances $\sigma_{\phi_0:1}^{2,\phi}$ and $\sigma_{\phi_0:1}^{2,\nu}$ (broken line) against ϕ_0. The graph indeed provides some evidence that that the scheme $\nu(s)$ should give better results. This is particularly evident when ϕ_0 is small; this is unsurprising as one initializes from Π_{ϕ_0}, hence if ϕ_0 is closer to 1 one expects a constant increase in the annealing parameter to be preferable to a slow initial evolution.

We ran SMC samplers with both annealing schemes with $N = 10^4$ particles and different dimension values $d \in \{10, 25, 50\}$. The choice $\phi_0 = \frac{1}{d}$ is used for both annealing schemes. We used a Markov kernel $k_s(x, dx')$ corresponding to a Random-Walk Metropolis with proposal $y = x + N(0, \frac{1}{25\phi_0})$, thus the proposal variance is $1/25$ times the variance of the starting distribution of the bridge $N(0, \frac{1}{\phi_0})$; this is a choice that gave good acceptance probabilities over all d bridging steps of the sampler. Multinomial resampling was used when the effective sample size dropped below $\frac{N}{2}$. We made 50 independent runs of the algorithm, and calculated the corresponding realisations of the log Ratio:

$$\sum_{k=2}^{m^*(d)+1} \log \left(\frac{\gamma_{d,k}^N(1)}{\gamma_{d,k-1}^N(1)} \right) \Big/ \log \left(\frac{\gamma_{d,k}(1)}{\gamma_{d,k-1}(1)} \right) \tag{12}$$

(note that now the resampling times and their number are random) and their sample variance. This experiment was carried out for choices of dimension $d \in \{10, 25, 50\}$ and for both annealing schemes $\phi(s)$ and $\nu(s)$. The ratio of the obtained variances for the annealing sequence $\phi(s)$ over $\nu(s)$ are shown in Table 1. The results confirm our theoretical findings above for the superiority of $\nu(s)$ over $\phi(s)$ based on the analytical expression for the asymptotic variance even for moderate d.

Dimension d	10	25	50
Ratio of Variances (for $\phi(s)$ over $\nu(s)$)	2.32	3.47	7.05

Table 1: The empirical variances (oven 50 independent runs for the SMC sampler) of the log Ratio (12) using the annealing $\phi(s)$ over the corresponding variances for the annealing sequence $\nu(s)$.

4.2 Bayesian Linear Model

We now consider the implications of main results in the context of a Bayesian linear regression model (see [14] for a book-length introduction as well as a wealth of practical applications). This is a statistical model that associates a L-vector of responses, say Y, to a $L \times d$-matrix of explanatory variables, X, for some $L \geq 1$, $d \geq 1$. In particular:

$$Y = X\beta + \epsilon$$

where β is a d-vector of unknown regression coefficients and $\epsilon \sim N_L(0, \mathbf{1}_L)$, with $\mathbf{1}_L$ the $L \times L$ identity matrix. A prior density on β is taken as $N_d(0, \mathbf{1}_d)$ which yields a posterior density found to be the d-dimensional Gaussian $N_d((\mathbf{1}_d + X'X)^{-1}X'Y, (\mathbf{1}_d + X'X)^{-1})$ where X' denotes transpose. This is the target distribution for our SMC sampler.

The objective is to investigate the bound in Theorem 3.1 and the implications of Proposition 3.2. Note that the target distribution is not of product structure here. The data-point tempering method (see Section 3.2.2) is also compared with annealing. We consider the case $d = 50$, $L = 50$ with $N = 10^3$; the data are all simulated. The annealing scheme ν in (11) is adopted as well as the data-point tempering method with $\lfloor 10d/L \rfloor$ steps between the L data point arrivals. Particles are propagated along the bridging densities via Markov kernels corresponding to Random-Walk Metropolis within Gibbs: the proposal for a univariate co-ordinate x conditionally on the rest is $y = x + N(0, \frac{1}{16})$. Dynamic resampling according to the ESS is employed (threshold $\frac{N}{2}$) as well as resampling at the last time step (see Theorem 3.1). For the annealing scheme the number of SMC steps is scaled as a multiple of
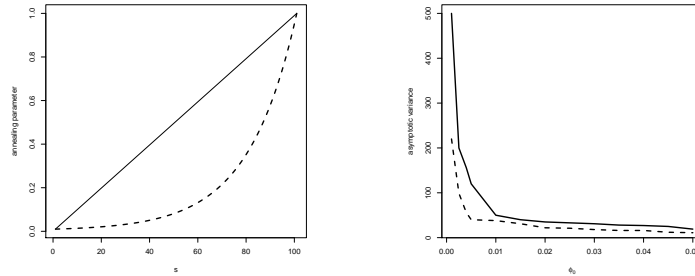
(a) Annealing Schemes, $\phi_0 = 0.01$ (b) Asymptotic Variances

Figure 1: Panel (a): Plot of the annealing parameters $\phi(s)$ and $\nu(s)$ (broken line) defined in (11) against time when $\phi_0 = 0.01$ and $\vartheta = 5$. Panel (b): A Plot of $\sigma^{2,\phi}_{\phi_0:1}$ and $\sigma^{2,\nu}_{\phi_0:1}$ (broken line) against ϕ_0 for the case of exact sampling with $k_s(x, dx') \equiv \pi_s(dx')$ and the scenario (10).

Time-Steps	d	$5d$	$10d$
Relative Error	4.75	4.47	3.9

Table 2: The mean square error when estimating $\mathbb{E}[\beta_1 | Y]$ for the (annealed) SMC sampler over 100 repeatitions relative to i.i.d. sampling. 1000 particles are run and we resample at the final time step. The error is calculated for different choices of number of time-steps for the SMC sampler.

d. This increase at the number of time steps aims at illustrating the propagation of chaos (Proposition 3.2). We fixed $d = 50$ for computational cost considerations, but the SMC algorithms will easily stabilize for much larger d.

Each SMC method employed is repeatedly applied 100 times. We calculate the mean square error for the estimation of $\mathbb{E}[\beta_1 | Y]$ (analytically available here) over the 100 replications and we compare with the corresponding error under i.i.d. sampling of the posterior of β_1; the results are reported in Table 2. In the table we can observe the increase in mean square error of the (annealed) SMC algorithm relatively to i.i.d. simulation. The increase here is not substantial as indicated by Theorem 3.1, although one may need to take d very large (and have an i.i.d. target) before the bound in Theorem 3.1 is realized. As the number of time-steps increases we can observe an improvement. This is due to increase in diversity of the population, which improves the SMC estimate even when resampling at the end. For the data-point tempering method (the CPU time is roughly comparable with the case of $10d$ time-steps of the annealed SMC) the corresponding value of the relative mean square error is 7.5, which is slightly worse than the annealing scheme. In general, it is difficult to draw a definite conclusion on which scheme may be better.

5   Summary

In this paper we have considered the stability of SMC methods in high-dimensions. In particular: the $\mathbb{L}_2$-error of marginal estimates, the $\mathbb{L}_2$-relative error of the normalizing constants and propagation of chaos properties. The main focus has been illustrating stability of estimates in high dimension d for fixed number of particles N. Some directions for future work are as follows.

Firstly, in the context of normalizing constants, one direction is the consideration of rare events problems, e.g. as described in [8]. Following [8], it is possible to obtain computational complexity results for some rare events problems. Relevant to the material in this paper, one can pose some rare-events problems in terms of the dimensionality (that is, the event of interest becomes less likely as the dimension grows). Our results would, in many cases, not apply to this scenario and an extension to this case is important.

Secondly, for normalizing constants we have only considered the relative $\mathbb{L}_2$-error. It would be of interest to consider e.g. higher-order errors. In addition, by studying the normalizing constant, we have only considered one particular important functional that changes with d. More generally, when one can perform estimation with direct Monte Carlo, with a cost which is less than exponential in d, is it possible to do this also with SMC methods?

Thirdly, one important research direction will be to investigate the relevance of results in high dimensions in the context of SMC methods used in filtering (e.g. particle filtering). This appears to be the main interest in the literature and it may be that there is no SMC algorithm which will always perform well in high-dimensions for any model. One potentially important algorithm in this direction can be found in [21].

Finally, one could considerably weaken the the hypotheses made in this article; as noted earlier in the article, the use of weaker conditions in the stability analysis of SMC methods is rather recent. Given the work of [28] and the number of exponential moments that we need to treat, it seems that multiplicative drift conditions [19] could be adopted.

Acknowledgements

A   Proofs

A.1   Preliminary Results

We summarize in Lemmas A.1 and A.2 below some results required in the proofs obtained in [4] or implied directy from results in that paper. Recall the definition of $G_{k,j}^i$ from (6).

Lemma A.1 (G-Asymptotics). *Assume (A1-2) and $g \in \mathcal{B}_b(E)$.*

i) *Under the starting distribution $X_{l_d(t_{k-1}(d)),1:d}^{1:N} \sim \pi_{t_{k-1}(d)}^{\otimes Nd}$ we have that as $d \to \infty$:*

$$\frac{G_{k,j}^i}{\sqrt{d}} \Rightarrow N(0, \sigma_{t_{k-1}:t_k}^2) \; ; \quad \frac{1}{d} \sum_{j=1}^d G_{k,j}^i \Rightarrow N(0, \sigma_{t_{k-1}:t_k}^2) \; .$$

ii) *We have that $| \mathbb{E}_{\check{X}_{l_d(t_{k-1}(d)),j}^i} [G_{k,j}^i] | \le M$, and for any $p \ge 2$:*

$$\mathbb{E}_{\check{X}_{l_d(t_{k-1}(d)),j}^i} |G_{k,j}^i|^p \le M \, d^{\frac{p}{2} \vee 1} \; .$$

iii) *Under either $\pi_{t_{k-1}(d)}^{\otimes Nd}$ as in i) or the actual particle distribution we have that as $d \to \infty$:*

$$\mathbb{E} \big[e^{\frac{c}{d} \sum_{j=1}^d G_{k,j}^i} \big] \to \mathbb{E} \big[e^{c \, N(0, \sigma_{t_{k-1}:t_k}^2)} \big] \equiv e^{\frac{1}{2} c^2 \sigma_{t_{k-1}:t_k}^2} \; .$$

iv) *We have that as $d \to \infty$:*

$$\frac{1}{d} \sum_{j=1}^d \mathbb{E}_{\check{X}_{l_d(t_{k-1}(d)),j}^i} [G_{k,j}^i] \to 0 \; , \quad in \;\; \mathbb{L}_1 \; ;$$

$$\frac{1}{d^2} \sum_{j=1}^d \mathbb{E}_{\check{X}_{l_d(t_{k-1}(d)),j}^i} \big[\big(G_{k,j}^i - \mathbb{E}_{\check{X}_{l_d(t_{k-1}(d)),j}^i} [G_{k,j}^i] \big)^2 \big] \to \sigma_{t_{k-1}:t_k}^2 \; , \quad in \;\; \mathbb{L}_1 \; .$$

Proof.

i) Both weak limit follows from the proof of Theorem 3.2 of [4]. Notice, that a minor difference is that instead of the *fixed* times ϕ_0 and 1 considered in Theorem 3.2 of [4] we now sum terms between the *varying* time instances $t_{k-1}(d)$ and $t_k(d)$. However, the proof for this case follows trivially from the proof for the fixed times due to the limits $t_{k-1}(d) \to t_{k-1}$ and $t_k(d) \to t_k$.

ii) All these results follow directly from Theorem A.1 of [4].

iii) This follows from the CLT's in parts i) and ii) and the uniform integrability result obtained in Lemma A.6.

iv) The first result corresponds to Proposition C.4 of [4]. The second result is shown in the proof of Theorem 4.1 of [4]. □

Lemma A.2. *(Convergence of Marginal Laws) Assume (A1-2) and $g \in \mathcal{B}_b(E)$. Then we have:*

i) *For a sequence of times $s(d) \in (\phi_0, 1)$ with $t_{k-1}(d) < s(d)$ and $s(d) \to s \in (t_{k-1}, 1)$ and the collection of time steps $u(d) = (l_d(t_{k-1}(d)) + 1) : l_d(s(d))$ we have that as $d \to \infty$:*

$$\|k_{u(d)}(\check{X}^i_{l_d(t_{k-1}(d)),1}) - \pi_{t_{k-1}(d)}k_{u(d)}\|_{tv} \to 0 , \quad in \quad \mathbb{L}_1 ;$$

$$\|\pi_{t_{k-1}(d)}k_{u(d)} - \pi_{s(d)}\|_{tv} \to 0 .$$

ii) *For a sequence of times $s(d) \in (\phi_0, 1)$ with $t_{k-1}(d) < s(d)$ and $s(d) \to s \in (t_{k-1}, 1)$ and the collection of time steps $u(d) = l_d(t_{k-1}(d)) : l_d(s(d))$ we have that:*

$$\left(\overline{w}^{1:N}_{u(d)}, X^{1:N}_{l_d(s(d)),1} \right) \Rightarrow \left(\frac{e^{Z^{1:N}}}{\sum_{l=1}^N e^{Z^l}}, Y^{1:N} \right)$$

where $\{Z^i\}_{i=1}^N$ are i.i.d. copies from $N(0, \sigma^2_{t_{k-1}:s})$ and, independently, $\{Y^i\}_{i=1}^N$ are i.i.d. copies from π_s.

Proof.

i) The first result follows by the proof of Proposition C.4 of [4]; the second result from Proposition A.1 of [4].

ii) The weak convergence of the weights is analytically illustrated in the proof of Theorem 4.1 of [4]. The weak convergence of the positions of the Markov chain is proven in Proposition A.1 of [4]. The independence between the $Z^{1:N}$ and $Y^{1:N}$ limiting variables follows trivially from the fact that any single co-ordinate has a vanishing effect on the weights as $d \to \infty$. □

A.2 $\mathbb{L}_2$-Error

Proof of Theorem 3.1. We begin by noting that, due to exhangeability of the particles:

$$\mathbb{E}\left[\left(\frac{1}{N} \sum_{i=1}^N [\varphi(\check{X}^i_{d,1}) - \pi(\varphi)] \right)^2 \right] = \frac{1}{N} \mathbb{E}\left[\{ \overline{\varphi}(\check{X}^1_{d,1}) \}^2 \right] + \left(\frac{N-1}{N} \right) \mathbb{E}\left[\overline{\varphi}(\check{X}^1_{d,1}) \overline{\varphi}(\check{X}^2_{d,1}) \right] \tag{13}$$

where we have set $\overline{\varphi}(x) = \varphi(x) - \pi(\varphi)$. Starting with the first term on the R.H.S. of (13), and averaging over the noise introduced by resampling, one has

$$\frac{1}{N} \mathbb{E}\left[\{ \overline{\varphi}(\check{X}^1_{d,1}) \}^2 \right] = \frac{1}{N} \sum_{i=1}^N \mathbb{E}\left[\overline{w}^i_{u(d)} \{ \overline{\varphi}(X^i_{d,1}) \}^2 \right]$$

where we have set $u(d) = l_d(t_{m^*}(d)) : d$. Recall that $\overline{w}^i_{u(d)}$ denote the normalized weights. By the asymptotic independence result in Lemma A.2(ii) we have that

$$\lim_{d \to \infty} \frac{1}{N} \sum_{i=1}^N \mathbb{E}\left[\overline{w}^i_{u(d)} \{ \overline{\varphi}(X^i_{d,1}) \}^2 \right] = \frac{1}{N} \mathbb{E}\left[\sum_{i=1}^N \frac{e^{Z^i}}{\sum_{l=1}^N e^{Z^l}} \{ \overline{\varphi}(Y^i) \}^2 \right] = \frac{\mathbb{V}\mathrm{ar}_\pi[\varphi]}{N} ,$$

where $\{Z^i\}_{i=1}^N$ are i.i.d. from $N(0, \sigma^2_{t_{m^*}:1})$ and, independently, $Y^1, \ldots, Y^N$ i.i.d. from π. We now look at the second term on the R.H.S. of (13). Averaging over the resampling index and invoking again the asymptotic independence result of Lemma A.2(ii) we have:

$$\mathbb{E}\left[\overline{\varphi}(\check{X}^1_{d,1}) \overline{\varphi}(\check{X}^2_{d,1}) \right] = \sum_{i=1}^N \mathbb{E}\left[\overline{\varphi}^2(X^i_{d,1}) (\overline{w}^i_{u(d)})^2 \right] + \sum_{i \neq l} \mathbb{E}\left[\overline{\varphi}(X^i_{d,1}) \overline{\varphi}(X^l_{d,1}) \overline{w}^i_{u(d)} \overline{w}^l_{u(d)} \right] \to$$

$$\pi(\overline{\varphi}^2) \mathbb{E}\left[\sum_{i=1}^N \frac{e^{2Z^i}}{(\sum_{l=1}^N e^{Z^l})^2} \right] + 0 \equiv N \pi(\overline{\varphi}^2) \mathbb{E}\left[\frac{e^{2Z^1}}{(\sum_{l=1}^N e^{Z^l})^2} \right] \tag{14}$$

13

for random variables $\{Z^i\}_{i=1}^N$ as defined above (in the last calculation we took advantage of exhangeablity). We have the decomposition (writing $\sigma^2 \equiv \sigma^2_{t_{m^*-1}:1}$ for notational convenience):

$$\frac{e^{2Z^1}}{(\sum_{l=1}^N e^{Z^l})^2} = \frac{1}{N^2}\frac{e^{2Z^1}}{e^{\sigma^2}} + e^{-\sigma^2}\frac{e^{2Z^1}}{(\sum_{l=1}^N e^{Z^l})^2}\Big(e^{\sigma^2} - \big(\tfrac{\sum_{l=1}^N e^{Z^l}}{N}\big)^2\Big) .$$

We concentrate on the second term. Using Holder inequality we have:

$$\mathbb{E}\Big[\Big|\frac{e^{2Z^1}}{(\sum_{l=1}^N e^{Z^l})^2}\Big(e^{\sigma^2} - \big(\tfrac{\sum_{l=1}^N e^{Z^l}}{N}\big)^2\Big)\Big|\Big] \leq \mathbb{E}^{\frac{2}{3}}\Big[\frac{e^{3Z^1}}{(\sum_{l=1}^N e^{Z^l})^3}\Big]\,\mathbb{E}^{\frac{1}{3}}\Big[\Big|e^{\sigma^2} - \big(\tfrac{\sum_{l=1}^N e^{Z^l}}{N}\big)^2\Big|^3\Big]$$

Setting $Z^{(1)} := \min_{1\leq i\leq N} Z^i$ we get that (using also Cauchy-Schwarz):

$$\mathbb{E}\Big[\frac{e^{3Z^1}}{(\sum_{l=1}^N e^{Z^l})^3}\Big] \leq \tfrac{1}{N^3}\mathbb{E}\big[e^{3Z^1 - 3Z^{(1)}}\big] \leq \tfrac{1}{N^3}\mathbb{E}^{\frac{1}{2}}\big[e^{6Z^1}\big]\mathbb{E}^{\frac{1}{2}}\big[e^{-6Z^{(1)}}\big] .$$

By standard results on order statistics the pdf of $Z^{(1)}$ is upper bounded by N times the pdf of $N(0,\sigma^2)$. So, we have that:

$$\mathbb{E}\big[e^{-6Z^{(1)}}\big] \leq Ne^{18\sigma^2} .$$

By adding and subtracting e^{σ^2} in the summand and multiplying the square, one can use Minkowski and the Marcinkiewicz Zygmund inequality to obtain:

$$\mathbb{E}^{\frac{1}{3}}\Big[\Big|e^{\sigma^2} - \big(\tfrac{\sum_{l=1}^N e^{Z^l}}{N}\big)^2\Big|^3\Big] \leq \frac{Me^{6\sigma^2}}{N^{1/2}} ,$$

for some $M < \infty$ that does not depend upon N or σ^2. Putting together the above arguments, we have shown that the right-hand part of the R.H.S. of (14), when $d \to \infty$, is upper-bounded by the quantity $\mathbb{V}\mathrm{ar}_\pi(\varphi)(\frac{1}{N}e^{\sigma^2} + Me^{17\sigma^2}\frac{1}{N^{7/6}})$ which completes the proof. □

A.3 Normalizing Constants

Proof of Theorem 3.2. By the expression of the normalized variance (and the fact that the different particles are i.i.d.), one can re-center to rewrite:

$$\mathbb{V}_2(\gamma_d(1)) = \mathbb{E}\Big[\big(\tfrac{\overline{\gamma}_d^N(1)}{\overline{\gamma}_d(1)} - 1\big)^2\Big]$$

with

$$\overline{\gamma}_d^N(1) = \tfrac{1}{N}\sum_{i=1}^N e^{\frac{1}{d}\sum_{j=1}^d G_j^i} ; \quad \overline{\gamma}_d(1) = \mathbb{E}\big[e^{\frac{1}{d}\sum_{j=1}^d G_j^1}\big] ,$$

where we have now set

$$G_j^i = (1-\phi_0)\sum_{n=0}^{d-1}\big(g(X_{n,j}^i) - \mathbb{E}\big[g(X_{n,j}^i)\big]\big) \tag{15}$$

and $i \in \{1,\ldots,N\}$, $j \in \{1,\ldots,d\}$. We have that:

$$\mathbb{E}\Big[\big(\tfrac{\overline{\gamma}_d^N(1)}{\overline{\gamma}_d(1)} - 1\big)^2\Big] = 1 - \tfrac{2}{\overline{\gamma}_d(1)}\mathbb{E}\big[\overline{\gamma}_d^N(1)\big] + \tfrac{1}{\overline{\gamma}_d(1)^2}\mathbb{E}\big[\overline{\gamma}_d^N(1)^2\big]$$
$$\equiv -1 + \tfrac{1}{\overline{\gamma}_d(1)^2}\mathbb{E}\big[\overline{\gamma}_d^N(1)^2\big] \tag{16}$$

where we have used the unbiasedness property (i.e. $\mathbb{E}\big[\overline{\gamma}_d^N(1)\big] = \overline{\gamma}_d(1)$) of the normalizing constant, see e.g. [10]. We define $Z_d^i = \tfrac{1}{d}\sum_{j=1}^d G_j^i$ for G_j^i defined in (15) and $1 \leq i \leq N$. Thus, due to Z_d^i's being i.i.d., we have:

$$\mathbb{E}\big[\overline{\gamma}_d^N(1)^2\big] = \tfrac{1}{N}\mathbb{E}\big[e^{2Z_d^1}\big] + (1-\tfrac{1}{N})\mathbb{E}^2\big[e^{Z_d^1}\big] .$$

By Lemma A.1(iii), applied when $t_{k-1}(d) \equiv \phi_0$ and $t_k(d) \equiv 1$, one has that:

$$\mathbb{E}\big[e^{2Z_d^1}\big] \to \exp\{2\sigma^2_{\phi_0:1}\} ; \quad \mathbb{E}\big[e^{Z_d^1}\big] \to \exp\{\tfrac{1}{2}\sigma^2_{\phi_0:1}\} .$$

Using these limits in (16) and recalling also that $\overline{\gamma}_d(1) \equiv \mathbb{E}\big[e^{Z_d^1}\big]$, gives the required result. □

14

Proof of Theorem 3.3. Denote:

$$\overline{\gamma}_{d,k}^N(1) = \frac{1}{N}\sum_{i=1}^N e^{\frac{1}{d}\sum_{j=1}^d G_{k,j}^i} \; ; \quad \overline{\gamma}_{d,k}(1) = \mathbb{E}_{\pi_{t_{k-1}(d)}^{\otimes Nd}}\left[e^{\frac{1}{d}\sum_{j=1}^d G_{k,j}^1} \right] , \tag{17}$$

for the standardised $G_{k,j}^i$ in (6). We look at the relative $\mathbb{L}_2$-error:

$$\mathbb{V}_2\Big(\prod_{k=1}^{m^*+1}\gamma_{d,k}(1) \Big) = \mathbb{E}\left[\Big(\prod_{k=1}^{m^*+1}\frac{\overline{\gamma}_{d,k}^N(1)}{\overline{\gamma}_{d,k}(1)} - 1 \Big)^2 \right] .$$

Using the unbiased property of normalising constants, see e.g. [10], we have:

$$\mathbb{E}\left[\big(\prod_{k=1}^{m^*+1}\frac{\overline{\gamma}_{d,k}^N(1)}{\overline{\gamma}_{d,k}(1)} - 1 \big)^2 \right] = \mathbb{E}\left[\prod_{k=1}^{m^*+1}\frac{\overline{\gamma}_{d,k}^N(1)^2}{\overline{\gamma}_{d,k}(1)^2} \right] - 1 .$$

For notational convenience, we set:

$$\Delta_{k,d} := \frac{\overline{\gamma}_{d,k}^N(1)^2}{\overline{\gamma}_{d,k}(1)^2} \; ; \quad \delta_{k,d} := \mathbb{E}_{\pi_{t_{k-1}(d)}^{\otimes Nd}}\left[\frac{\overline{\gamma}_{d,k}^N(1)^2}{\overline{\gamma}_{d,k}(1)^2} \right] \; ; \quad \Delta_{1:k,d} = \prod_{q=1}^k\Delta_{q,d} \; ; \quad \delta_{1:k,d} = \prod_{q=1}^k\delta_{q,d} .$$

Following the definitions of $\overline{\gamma}_{d,k}^N(1)$ and $\overline{\gamma}_{d,k}(1)$ in (17), and exploiting independence among particles under $\pi_{t_{k-1}(d)}^{\otimes Nd}$, we have that:

$$\mathbb{E}_{\pi_{t_{k-1}(d)}^{\otimes Nd}}\left[\frac{\overline{\gamma}_{d,k}^N(1)^2}{\overline{\gamma}_{d,k}(1)^2} \right] = \frac{\frac{1}{N}\mathbb{E}\left[e^{\frac{2}{d}\sum_{j=1}^d G_{k,j}^1} \right] + \left(1 - \frac{1}{N}\right)\mathbb{E}^2\left[e^{\frac{1}{d}\sum_{j=1}^d G_{k,j}^1} \right]}{\mathbb{E}^2\left[e^{\frac{1}{d}\sum_{j=1}^d G_{k,j}^1} \right]}$$

$$\rightarrow e^{-\sigma_{t_{k-1}:t_k}^2}\left[e^{2\sigma_{t_{k-1}:t_k}^2}\frac{1}{N} + \left(1 - \frac{1}{N}\right)e^{\sigma_{t_{k-1}:t_k}^2} \right] ,$$

with the limit obtained from Lemma A.1(iii). Therefore:

$$\delta_{1:(m^*+1),d} \rightarrow e^{-\sigma_{\phi_0:1}^2}\prod_{k=1}^{m^*+1}\left[\frac{1}{N}e^{2\sigma_{t_{k-1}:t_k}^2} + \left(1 - \frac{1}{N}\right)e^{\sigma_{t_{k-1}:t_k}^2} \right] .$$

Thus, it suffices to show that the following difference goes to zero as $d \rightarrow \infty$:

$$A_d := \big| \mathbb{E}\left[\Delta_{1:(m^*+1),d}\right] - \delta_{1:(m^*+1),d} \big| .$$

Now, note that simple calculations give that:

$$A_d = \left| \sum_{k=1}^{m^*+1}\mathbb{E}\left[\Delta_{1:(k-1),d}\big(\mathbb{E}[\Delta_{k,d}|\mathscr{F}_{t_{k-1}(d)}^N] - \delta_{k,d} \big) \right] \cdot \delta_{(k+1):(m^*+1),d} \right| ,$$

under the conventions that $\Delta_{1:0,d} = \delta_{(m^*+2):(m^*+1)} = 1$. Applying Cauchy-Schwarz yields the following upper-bound:

$$\mathbb{E}\left[\Delta_{1:(k-1),d}\big| \mathbb{E}[\Delta_{k,d}|\mathscr{F}_{t_{k-1}(d)}^N] - \delta_{k,d}\big| \right] \le$$
$$\mathbb{E}^{1/2}\left[\Delta_{1:(k-1),d}^2 \right] \cdot \mathbb{E}^{1/2}\left[\big| \mathbb{E}[\Delta_{k,d} \mid \mathscr{F}_{t_{k-1}(d)}^N] - \delta_{k,d}\big|^2 \right] .$$

Via Lemma A.3 the second of the terms in the bottom line vanishes in the limit, so it suffices to show that the first term in the bottom line is upper bounded uniformly in d. Using the Cauchy-Schwarz inequality, we have that:

$$\mathbb{E}\left[\Delta_{1:(k-1),d}^2 \right] \le \prod_{q=1}^{k-1}\mathbb{E}^{1/2}\left[\Delta_{q,d}^4 \right] .$$

Recalling the definition of $\Delta_{k,d} = \frac{\overline{\gamma}_{d,k}^N(1)^2}{\overline{\gamma}_{d,k}(1)^2}$ from (17), using triangle inequality for norms we have:

$$\mathbb{E}\left[\overline{\gamma}_{d,q}^N(1)^8\right] \le \Big(\frac{1}{N}\sum_{i=1}^N\mathbb{E}^{1/8}\left[e^{\frac{8}{d}\sum_{j=1}^d G_{q,j}^i} \right] \Big)^8$$

Now, we complete via Lemma A.6. □

15

Proof of Proposition 3.1. To simplify the notation we drop i for the particle number and define:

$$\mathcal{G}_{l,d} = \sum_{j=1}^{d} G_{l,j} \ ,$$

for $1 \le l \le k$. Our proof proceeds by induction. For $k = 1$, the result follows by Lemma A.1(iii). Assume that the result holds at time $k - 1 \ge 1$. Then we have the simple decomposition:

$$\mathbb{E}\big[e^{\sum_{l=1}^{k} c_l\,\mathcal{G}_{l,d}/d}\big] = \mathbb{E}\Big[\mathbb{E}\big[e^{c_k\,\mathcal{G}_{k,d}/d}\,\big|\,\mathscr{F}^N_{t_{k-1}(d)}\big]\big\{e^{\sum_{l=1}^{k-1} c_l\,\mathcal{G}_{l,d}/d} - \mathbb{E}\big[e^{\sum_{l=1}^{k-1} c_l\,\mathcal{G}_{l,d}/d}\big]\big\}\Big]$$
$$+ \mathbb{E}\big[e^{\sum_{l=1}^{k-1} c_l\,\mathcal{G}_{l,d}/d}\big]\,\mathbb{E}\big[e^{c_k\,\mathcal{G}_{k,d}/d}\big]\ . \tag{18}$$

We begin by dealing with the first term on the R.H.S. of (18). By Lemma A.4 we have that:

$$\mathbb{E}\big[e^{c_k\,\mathcal{G}_k/d}\,\big|\,\mathscr{F}^N_{t_{k-1}(d)}\big] - \mathbb{E}_{\pi^{\otimes d}_{t_{k-1}(d)}}\big[e^{c_k\mathcal{G}_k/d}\big] \xrightarrow{\mathbb{P}} 0 \ ,$$

whereas from Lemma A.1(iii) we have:

$$\mathbb{E}_{\pi^{\otimes d}_{t_{k-1}(d)}}\big[e^{c_k\mathcal{G}_k/d}\big] \to e^{\frac{1}{2}c_k^2\,\sigma^2_{t_{k-1}:t_k}} \ . \tag{19}$$

Moreover, by the induction hypothesis:

$$\big\{e^{\sum_{l=1}^{k-1} c_l\,\mathcal{G}_{l,d}/d} - \mathbb{E}\big[e^{\sum_{l=1}^{k-1} c_l\,\mathcal{G}_{l,d}/d}\big]\big\} \Rightarrow e^{\sum_{l=1}^{k-1} c_l X_l} - e^{\frac{1}{2}\sum_{l=1}^{k-1} c_l^2 \sigma^2_{t_{l-1}:t_l}} \ .$$

The expression in the expectation of the first term of (18) is uniformly integrable: indeed, careful and repeated (but otherwise straightforward) use of Hölder and Jensen inequalities will eventually give that:

$$\Big|\mathbb{E}\big[e^{c_k\,\mathcal{G}_{k,d}/d}\,\big|\,\mathscr{F}^N_{t_{k-1}(d)}\big]\big\{e^{\sum_{l=1}^{k-1} c_l\,\mathcal{G}_{l,d}/d} - \mathbb{E}\big[e^{\sum_{l=1}^{k-1} c_l\,\mathcal{G}_{l,d}/d}\big]\big\}\Big|_{\mathbb{L}_{1+\epsilon}}$$
$$\le M \prod_{l=1}^{k-1}\big(\mathbb{E}\big[e^{\sum_{l=1}^{k} c'_l\,\mathcal{G}_{l,d}/d}\big]\big)^{1/(1+\delta_l)}$$

for positive constants $c'_{1:k}$, $\delta_{1:k}$, M independent of d. As a consequence, convergence in distribution implies also convergence of expectations:

$$\mathbb{E}\Big[\mathbb{E}\big[e^{c_k\,\mathcal{G}_{k,d}/d}\,\big|\,\mathscr{F}^N_{t_{k-1}(d)}\big]\big\{e^{\sum_{l=1}^{k-1} c_l\,\mathcal{G}_{l,d}/d} - \mathbb{E}\big[e^{\frac{1}{d}\sum_{l=1}^{k-1} c_l\,\mathcal{G}_{l,d}/d}\big]\big\}\Big]$$
$$\to \mathbb{E}\big[e^{c_k^2\sigma^2_{t_{k-1}:t_k}/2}\big\{e^{\sum_{l=1}^{k-1} c_l X_l} - e^{\frac{1}{2}\sum_{l=1}^{k-1} c_l^2\sigma^2_{t_{l-1}:t_l}}\big\}\big] \equiv 0 \ .$$

Now turning to the second term on the R.H.S. of (18), we work as follows:

$$\mathbb{E}\big[e^{c_k\,\mathcal{G}_{k,d}/d}\big] = \mathbb{E}\Big[\mathbb{E}\big[e^{c_k\,\mathcal{G}_{k,d}/d}\,\big|\,\mathscr{F}^N_{t_{k-1}(d)}\big] - \mathbb{E}_{\pi^{\otimes d}_{t_{k-1}(d)}}\big[e^{c_k\,\mathcal{G}_{k,d}/d}\big]\Big] + \mathbb{E}_{\pi^{\otimes d}_{t_{k-1}(d)}}\big[e^{c_k\,\mathcal{G}_{k,d}/d}\big]$$
$$\to 0 + e^{\frac{1}{2}c_k^2\,\sigma^2_{t_{k-1}:t_k}} \ ,$$

from Lemma A.4 and (19). We can thus deduce by the induction hypothesis that

$$\mathbb{E}\big[e^{\sum_{l=1}^{k-1} c_l\,\mathcal{G}_{l,d}/d}\big]\,\mathbb{E}\big[e^{c_k\,\mathcal{G}_k/d}\big] \to e^{\frac{1}{2}\sum_{l=1}^{k} c_l^2\sigma^2_{t_{l-1}:t_l}} \equiv \prod_{l=1}^{k}\mathbb{E}\big[e^{c_l Z^l}\big]$$

which completes the proof. $\qquad\square$

Lemma A.3. *Assume (A1-2) and $g \in \mathcal{B}_b(E)$. Then for any $\epsilon > 0$, $N \ge 1$ and $1 \le k \le m^* + 1$:*

$$\mathbb{E}\big[\tfrac{\overline{\gamma}^N_{d,k}(1)^2}{\overline{\gamma}_{d,k}(1)^2}\,\big|\,\mathscr{F}^N_{t_{k-1}(d)}\big] - \mathbb{E}_{\pi^{\otimes Nd}_{t_{k-1}(d)}}\big[\tfrac{\overline{\gamma}^N_{d,k}(1)^2}{\overline{\gamma}_{d,k}(1)^2}\big] \to 0 \ , \quad in \ \ \mathbb{L}_{1+\epsilon} \ .$$

Proof. Due to conditional independence among particles given $\mathscr{F}^N_{t_{k-1}(d)}$, we have:

$$\mathbb{E}\left[\overline{\gamma}^N_{d,k}(1)^2 \,\middle|\, \mathscr{F}^N_{t_{k-1}(d)}\right] = \tag{20}$$

$$= \frac{1}{N^2}\left(\mathbb{E}\left[\sum_{i=1}^N e^{\frac{2}{d}\sum_{j=1}^d G^i_{k,j}} \middle| \mathscr{F}^N_{t_{k-1}(d)}\right] + \sum_{i\neq m}\mathbb{E}\left[e^{\frac{1}{d}\sum_{j=1}^d G^i_{k,j}} \middle| \mathscr{F}^N_{t_{k-1}(d)}\right]\mathbb{E}\left[e^{\frac{1}{d}\sum_{j=1}^d G^m_{k,j}} \middle| \mathscr{F}^N_{t_{k-1}(d)}\right]\right).$$

Now, for any constant $c \geq 1$ we have $\sup_d \mathbb{E}_{\pi^{\otimes Nd}_{t_{k-1}(d)}}\left[e^{\frac{c}{d}\sum_{j=1}^d G^i_{k,j}}\right] < \infty$ from Lemma A.6, so it suffices to prove that for any constant $c \geq 1$, as $d \to \infty$:

$$\mathbb{E}\left[e^{\frac{c}{d}\sum_{j=1}^d G^i_{k,j}} \middle| \mathscr{F}^N_{t_{k-1}(d)}\right] - \mathbb{E}_{\pi^{\otimes Nd}_{t_{k-1}(d)}}\left[e^{\frac{c}{d}\sum_{j=1}^d G^i_{k,j}}\right] \to 0, \quad \text{in } \mathbb{L}_{2(1+\epsilon)}. \tag{21}$$

The factor of two in the norm arises as own has to use Cauchy-Schwarz to separate the product terms on the R.H.S. of (20). Now, Lemma A.4 established the above convergence in probability; this together with uniform integrability implied by Lemma A.6 establishes the result. $\qquad\square$

Lemma A.4. *Assume (A1-2) and that $g \in \mathcal{B}_b(E)$. Then, for any $N \geq 1$, $i \in \{1,\ldots,N\}$, $k \in \{1,\ldots,m^*+1\}$ and $c \in \mathbb{R}$:*

$$\mathbb{E}\left[e^{\frac{c}{d}\sum_{j=1}^d G^i_{k,j}} \middle| \mathscr{F}^N_{t_{k-1}(d)}\right] - \mathbb{E}_{\pi^{\otimes Nd}_{t_{k-1}(d)}}\left[e^{\frac{c}{d}\sum_{j=1}^d G^i_{k,j}}\right] \xrightarrow{\mathbb{P}} 0.$$

Proof. By the conditional independence along j, we have:

$$\mathbb{E}\left[e^{\frac{c}{d}\sum_{j=1}^d G^i_{k,j}} \middle| \mathscr{F}^N_{t_{k-1}(d)}\right] = \prod_{j=1}^d \mathbb{E}_{\check{X}_{l_d(t_{k-1}(d)),j}}\left[e^{\frac{c}{d}G^i_{k,j}}\right].$$

We now omit various sub/super-scripts to simplify the notation, using also $\mathbb{E}_\pi \equiv \mathbb{E}_{\pi_{t_{k-1}(d)}}$ and $\mathbb{E}_{\check{X}_{0,j}} \equiv \mathbb{E}_{\check{X}_{l_d(t_{k-1}(d)),j}}$. We can rewrite:

$$\mathbb{E}\left[e^{\frac{c}{d}\sum_{j=1}^d G_j} \middle| \mathscr{F}^N\right] - \mathbb{E}_{\pi^{\otimes Nd}}\left[e^{\frac{c}{d}\sum_{j=1}^d G_j}\right] =$$

$$= \left(\prod_{j=1}^d \mathbb{E}_\pi\left[e^{cG_j/d}\right]\right)\left[\prod_{j=1}^d\left\{\frac{\{\mathbb{E}_{\check{X}_{0,j}} - \mathbb{E}_\pi\}[e^{cG_j/d}]}{\mathbb{E}_\pi[e^{cG_j/d}]} + 1\right\} - 1\right]. \tag{22}$$

From Lemma A.1(iii) it follows that $\prod_{j=1}^d \mathbb{E}_\pi\left[e^{cG_j/d}\right] \to e^{\frac{1}{2}c^2\sigma^2_{t_{k-1}:t_k}}$, hence we can now concentrate on the second factor-term on the R.H.S. of (22). We will replace the product with a sum using logarithms. To that end define:

$$\beta_j(d) := \frac{\{\mathbb{E}_{\check{X}_{0,j}} - \mathbb{E}_\pi\}[e^{cG_j/d}]}{\mathbb{E}_\pi[e^{cG_j/d}]}.$$

Note that since $g \in \mathcal{B}_b(E)$, we have that G_j/d is bounded from above and below, so there exist an $\epsilon > 0$ and $M > 0$ such that:

$$-1 + \epsilon \leq \beta_j(d) \leq M < \infty. \tag{23}$$

We need to prove that $e^{\sum_{j=1}^d \log(1+\beta_j(d))} - 1 \xrightarrow{\mathbb{P}} 0$. We consider a second order Taylor expansion of the exponent:

$$\sum_{j=1}^d \log(1 + \beta_j(d)) = \sum_{j=1}^d\left\{\beta_j(d) - \frac{1}{2}\frac{1}{(1+\xi_j(d))^2}\beta_j^2(d)\right\} \tag{24}$$

where $\xi_j(d) \in [0 \wedge \beta_j(d), 0 \vee \beta_j(d)]$. By Lemma A.5 we have that:

$$\sum_{j=1}^d \beta_j(d) \xrightarrow{\mathbb{P}} 0; \quad \sum_{j=1}^d \beta_j^2(d) \xrightarrow{\mathbb{P}} 0.$$

Since $\xi_j(d)$'s are bounded due to (23), these two results imply via the Taylor expansion in (24) that also $\sum_{j=1}^d \log(1+\beta_j(d)) \xrightarrow{\mathbb{P}} 0$. Due to the continuity of the exponential function, this implies now that $e^{\sum_{j=1}^d \log(1+\beta_j(d))} - 1 \Rightarrow 0$ and the proof is now complete since weak convergence to a constant implies convergence in probability. $\qquad\square$

Lemma A.5. *Assume (A1-2) and $g \in \mathcal{B}_b(E)$. Then we have that for any $N \geq 1$, $i \in \{1, \ldots, N\}$, $k \in \{1, \ldots, m^*+1\}$ and $c \in \mathbb{R}$:*

$$(i) \quad \sum_{j=1}^{d} \frac{\left\{ \mathbb{E}_{\check{X}^i_{l_d(t_{k-1}(d)),j}} - \mathbb{E}_{\pi_{t_{k-1}(d)}} \right\} \left[e^{c\, G^i_{k,j}/d} \right]}{\mathbb{E}_{\pi_{t_{k-1}(d)}} \left[e^{c\, G^i_{k,j}/d} \right]} \to 0 , \quad in\ \ \mathbb{L}_1 .$$

$$(ii) \quad \sum_{j=1}^{d} \left(\frac{\left\{ \mathbb{E}_{\check{X}^i_{l_d(t_{k-1}(d)),j}} - \mathbb{E}_{\pi_{t_{k-1}(d)}} \right\} \left[e^{c\, G^i_{k,j}/d} \right]}{\mathbb{E}_{\pi_{t_{k-1}(d)}} \left[e^{c\, G^i_{k,j}/d} \right]} \right)^2 \to 0 , \quad in\ \ \mathbb{L}_1 .$$

Proof. To simplify the presentation, we drop many super/subscripts: that is, we write the quantity of interest as:

$$\frac{\left\{ \mathbb{E}_{\check{X}_{0,j}} - \mathbb{E}_{\pi} \right\} \left[e^{c\, G_j/d} \right]}{\mathbb{E}_{\pi} \left[e^{c\, G_j/d} \right]} .$$

Note that $\mathbb{E}_{\pi} \left[e^{c\, G_j/d} \right] \equiv \mathbb{E}_{\pi} \left[e^{c\, G_1/d} \right]$. Since $|g|$ is bounded, $|G_1/d|$ is also bounded, so $\mathbb{E}_{\pi} \left[e^{c\, G_1/d} \right]$ is lower and upper bounded by positive constants and can be ignored in the calculations. We will be using the second-order Taylor expansion:

$$e^{c\, G_j/d} = 1 + \frac{c\, G_j}{d} + \frac{1}{2} e^{\xi_j(d)} \left(\frac{c\, G_j}{d} \right)^2 , \tag{25}$$

where $\xi_j(d) \in \left[0 \wedge \frac{c\, G_j}{d} , 0 \vee \frac{c\, G_j}{d} \right]$.

Proof of (i):
The $\mathbb{L}_1$-norm of the variable of interest is upper bounded by (recalling that $\mathbb{E}_{\pi} \left[G_j \right] \equiv 0$):

$$\mathbb{E} \left| \sum_{j=1}^{d} \mathbb{E}_{\check{X}_{0,j}} \left[\frac{c\, G_j}{d} \right] \right| + \frac{c^2}{2} \mathbb{E} \left| \sum_{j=1}^{d} \left\{ \mathbb{E}_{\check{X}_{0,j}} - \mathbb{E}_{\pi} \right\} \left[e^{\xi_j(d)} \left(\frac{G_j}{d} \right)^2 \right] \right| .$$

The first term in this bound goes to zero by Lemma A.1(iv). Thus considering the second term, we have the trivial inequality (for convenience we set $\sigma^2 \equiv \sigma^2_{t_{n-1}:t_n}$):

$$\mathbb{E} \left| \sum_{j=1}^{d} \left\{ \mathbb{E}_{\check{X}_{0,j}} - \mathbb{E}_{\pi} \right\} \left[e^{\xi_j(d)} \left(\frac{G_j}{d} \right)^2 \right] \right| \leq \tag{26}$$

$$\mathbb{E} \left| \sum_{j=1}^{d} \mathbb{E}_{\check{X}_{0,j}} \left[(e^{\xi_j(d)} - 1) \left(\frac{G_j}{d} \right)^2 \right] \right| + \mathbb{E} \left| \sum_{j=1}^{d} \mathbb{E}_{\check{X}_{0,j}} \left[\left(\frac{G_j}{d} \right)^2 \right] - \sigma^2 \right| + \left| \sigma^2 - \frac{1}{d} \mathbb{E}_{\pi} \left[e^{\xi_1(d)} G_1^2 \right] \right| .$$

Note that

- $|\xi_1(d)| < M$ (due to the boundedness assumption on g) ;

- $\xi_1(d) \to 0$ in distribution (so also in $\mathbb{L}_p$ for any $p \geq 1$ due to the above uniform bound) ;

- $\mathbb{E}_{\pi} \left[G_1^2/d \right] \to \sigma^2$,

with the last two results following from Lemma A.1(i,ii). These results, together, imply that the last term on the R.H.S. of (26) goes to zero. For the first term on the R.H.S. of (26) we work as follows. Since for each j, $|G_j/d|$ is bounded we have that $|e^{\xi_j(d)} - 1| \leq M |\xi_j(d)| \leq M |\frac{G_j}{d}|$. As a result, using the triangular inequality and then this latter bound we have that:

$$\mathbb{E} \left| \sum_{j=1}^{d} \mathbb{E}_{\check{X}_{0,j}} \left[(e^{\xi_j(d)} - 1) \left(\frac{G_j}{d} \right)^2 \right] \right| \leq \frac{M}{d^3} \sum_{j=1}^{d} \mathbb{E} \left[\mathbb{E}_{\check{X}_{0,j}} |G_j|^3 \right] = \frac{M}{d^3} \sum_{j=1}^{d} \mathbb{E} |G_j|^3 .$$

From Lemma A.1(ii) we have that this latter term is upper-bounded by $\frac{M}{d^3} d\, d^{3/2} \to 0$. Now, for the second term on the R.H.S. of (26) we work as follows. We have that:

$$\mathbb{E}_{\check{X}_{0,j}} \left[G_j^2 \right] = \mathbb{E}_{\check{X}_{0,j}} \left[\left(G_j - \mathbb{E}_{\check{X}_{0,j}} \left[G_j \right] \right)^2 \right] + \mathbb{E}^2_{\check{X}_{0,j}} \left[G_j \right] .$$

18

Lemma A.1(ii) gives that $|\mathbb{E}_{\check{X}_{0,j}}[G_j]| \leq M$, so we have that $\frac{1}{d^2}\sum_{j=1}^{d}\mathbb{E}_{\check{X}_{0,j}}^2[G_j] \to 0$ in $\mathbb{L}_1$. The result now follows from Lemma A.1(iv).

Proof of (ii):
We will use again the Taylor expansion (25). Clearly, the $\mathbb{L}_1$-norm of the random variable of interest is bounded by:

$$2\sum_{j=1}^{d}\mathbb{E}\left[\left(\mathbb{E}_{\check{X}_{0,j}}\left[\frac{G_j}{d}\right]\right)^2\right] + 2\sum_{j=1}^{d}\mathbb{E}\left[\left(\left\{\mathbb{E}_{\check{X}_{0,j}} - \mathbb{E}_{\pi}\right\}\left[\frac{1}{2}\left(\frac{G_j}{d}\right)^2 e^{\xi_j(d)}\right]\right)^2\right].$$

The first term goes to zero from the first result in Lemma A.1(ii) and the second from the second result in Lemma A.1(ii) applied here for $p = 4$. $\qquad\square$

Lemma A.6. *Assume (A1-2) and $g \in \mathcal{B}_b(E)$. Then we have that for any $N \geq 1$, $i \in \{1,\ldots,N\}$ and $k \in \{1,\ldots,m^*+1\}$ and any fixed $c \in \mathbb{R}$:*

$$\sup_d \mathbb{E}\left[e^{\frac{c}{d}\sum_{j=1}^{d}G_{k,j}^i}\right] < \infty.$$

Proof. To simplify the notation we rewrite the quantity of interest as

$$\mathbb{E}\left[e^{\frac{c}{d}\sum_{j=1}^{d}G_j}\right] \equiv \mathbb{E}\left[\prod_{j=1}^{d}\mathbb{E}_{\check{X}_{0,j}}\left[e^{\frac{c}{d}G_j}\right].\right]$$

Applying a second order Taylor expansion for $e^{\frac{c}{d}G_j}$ yields that the above is equal to:

$$\mathbb{E}\left[\prod_{j=1}^{d}\left(1 + c\,\mathbb{E}_{\check{X}_{0,j}}\left[\frac{G_j}{d}\right] + \frac{c^2}{2}\mathbb{E}_{\check{X}_{0,j}}\left[\left(e^{\xi_j(d)}\frac{G_j}{d}\right)^2\right]\right)\right]$$

with $\xi_j(d) \in [0 \wedge \frac{cG_j}{d}, 0 \vee \frac{cG_j}{d}]$. Using the fact that $|G_j/d|$ is upper bounded by a constant, from Lemma A.1(ii) we have:

$$\left|c\,\mathbb{E}_{\check{X}_{0,j}}\left[\frac{G_j}{d}\right]\right| \leq |c|\frac{M}{d}; \quad \frac{c^2}{2}\mathbb{E}_{\check{X}_{0,j}}\left[\left(e^{\xi_j(d)}\frac{G_j}{d}\right)^2\right] \leq c^2\frac{M}{d}.$$

Hence, we have that:

$$\mathbb{E}\left[e^{\frac{c}{d}\sum_{j=1}^{d}G_j}\right] \leq \left(1 + \frac{M}{d}\right)^d$$

with the latter upper bound converging by standard results in analysis. $\qquad\square$

A.4   Propagation of Chaos

Proof of Proposition 3.2. For simplicity, consider the first q of N particles and $j = 1$. Then, for a function $F : E^q \to [0,1]$ we have, using the notation $X_{s(d),1}^{1:q} = (X_{s(d),1}^1,\ldots,X_{s(d),1}^q)$:

$$|\mathbb{E}[F(X_{s(d),1}^{1:q})] - \pi_s^{\otimes q}(F)| \leq |\mathbb{E}[F(X_{s(d),1}^{1:q})] - \mathbb{E}_{\pi_{t_{k-1}(d)}^{\otimes N}}[F(X_{s(d),1}^{1:q})]|$$

$$+ |\mathbb{E}_{\pi_{t_{k-1}(d)}^{\otimes N}}[F(X_{s(d),1}^{1:q})] - \pi_{s(d)}^{\otimes q}(F)| + |\pi_{s(d)}^{\otimes q}(F) - \pi_s^{\otimes q}(F)|. \tag{27}$$

The last term on the R.H.S. goes to zero via the bounded convergence theorem (this follows directly from having assumed that g is upper bounded), so we consider the first two terms. For the first term on the R.H.S. of (27) one can use conditional expectations and write it as:

$$\mathbb{E}\left[\mathbb{E}[F(X_{s(d),1}^{1:q}) \mid \mathscr{F}_{t_{k-1}(d)}^N] - \mathbb{E}_{\pi_{t_{k-1}(d)}^{\otimes N}}[F(X_{s(d),1}^{1:q})]\right]$$

where $\mathscr{F}_{t_{k-1}(d)}^N$ is the filtration generated by the particle system up to (and including) the $(n-1)^{th}$ resampling time. The quantity inside the expectation can be equivalently written as:

$$\left\{k_{u(d)}^{\otimes q}(\check{X}_{l_d(t_{k-1}(d)),1}^{1:q}, \cdot) - \left(\pi_{t_{k-1}(d)}k_{u(d)}\right)^{\otimes q}\right\}(F) \tag{28}$$

where we set $u(d) = (l_d(t_{k-1}(d)) + 1) : l_d(s(d))$. For $1 \leq l \leq q$ we define the probability measures:

$$\mu_l = \mu_l(dy_{1:(l-1)}, dy_{(l+1):q}) = \left(\pi_{t_{k-1}(d)}k_{u(d)}\right)^{\otimes(l-1)} \otimes k_{u(d)}^{\otimes(q-l)}(\check{X}_{l_d(t_{k-1}(d)),1}^{(l+1):q}, \cdot).$$

19

Notice the simple identity (since intermediate terms in the sum below will cancel out):

$$\big\{k_{u(d)}^{\otimes q}(\check{X}_{l_d(t_{k-1}(d)),1}^{1:q},\,\cdot\,) - \big(\pi_{t_{k-1}(d)}k_{u(d)}\big)^{\otimes q}\big\}(dy_{1:q}) \tag{29}$$

$$= \sum_{l=1}^{q} \big(k_{u(d)}(\check{X}_{l_d(t_{k-1}(d)),1}^{l},\,\cdot\,) - \pi_{t_{k-1}(d)}k_{u(d)}\big)(dy_l) \otimes \mu_l(dy_{1:(l-1)}, dy_{(l+1):q})\ .$$

Since $|F| \leq 1$, we have $|\int \mu_l(dy_{1:(l-1)}, dy_{(l+1):q})F(y_{1:q})| \leq 1$ for any y_l. Given this property, using the identity (29) we have that the expression in (28) is bounded in absolute value by:

$$\Big| \sum_{l=1}^{q} \int_{\mathbb{R}} \big\{k_{u(d)}(\check{X}_{l_d(t_{k-1}(d)),1}^{l},\,\cdot\,) - \pi_{t_{k-1}(d)}k_{u(d)}\big\}(dy_l)\Big\{\frac{\int \mu_l(dy_{1:(l-1)}, dy_{(l+1):q})F(y_{1:q})}{\sup_{y_l\in\mathbb{R}}|\int \mu_l(dy_{1:(l-1)}, dy_{(l+1):q})F(y_{1:q})|}\Big\}\Big|$$

$$\leq \sum_{l=1}^{q} \|k_{u(d)}(\check{X}_{l_d(t_{k-1}(d)),1}^{l}) - \pi_{t_{k-1}(d)}k_{u(d)}\|_{tv}\ .$$

The above total variation bound converges to zero in $\mathbb{L}_1$ as $d \to \infty$ by Lemma A.2(i), so also the first term on the R.H.S. of (27) goes to zero as $d \to \infty$. The second term on the R.H.S. of (27) can be treated in a similar manner. One has again the identity:

$$\mathbb{E}_{\pi_{t_{k-1}(d)}^{\otimes N}}\big[\,F(X_{s(d),1}^{1:q})\,\big] - \pi_{s(d)}^{\otimes q}(F) =$$

$$= \sum_{l=1}^{q} \int_{\mathbb{R}} \big\{\,\pi_{t_{k-1}(d)}k_{u(d)} - \pi_{s(d)}\,\big\}(dx_l) \times \Big\{\pi_{s(d)}^{\otimes(l-1)} \otimes \big(\pi_{t_{k-1}(d)}k_{u(d)}\big)^{\otimes(q-l)}(F(x_l))\Big\}$$

$$\leq q\,\|\pi_{t_{k-1}(d)}k_{u(d)} - \pi_{s(d)}\|_{tv}\ .$$

This last bound which will go to zero by Lemma A.2(i). Hence we conclude. $\qquad\square$

References

[1] ARNAUD, É. & LE GLAND, F. (2009). SMC with adaptive resampling : large sample asymptotics. *Proc. of the 2009 IEEE Workshop on Statistical Signal Processing*, 481–484

[2] BÉDARD. M. (2007). Weak convergence of Metropolis algorithms for non-iid target distributions. *Ann. Appl. Probab.*, **17**, 1222–44.

[3] BENGTSSON, T., BICKEL, P., & LI, B. (2008). Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. In *Essays in Honor of David A. Freeman*, D. Nolan & T. Speed, Eds, 316–334, IMS.

[4] BESKOS, A., CRISAN, D., & JASRA A. (2011). On the stability of sequential Monte Carlo methods in high dimensions. Technical Report, Imperial College London.

[5] BICKEL, P., LI, B. & BENGTSSON, T. (2008). Sharp failure rates for the bootstrap particle filter in high dimensions. In *Pushing the Limits of Contemporary Statistics*, B. Clarke & S. Ghosal, Eds, 318–329, IMS.

[6] BREYER, L. A., PICCIONI, M. & SCARLATTI, S. (2004). Optimal scaling of MALA for nonlinear regression. *Ann. Appl. Probab.*, **14**, 1479–1505.

[7] CAMPILLO, F., CÉROU, F., LE GLAND, F., & RAKOTOZAFY, R. (1995). Particle and cell approximations for nonlinear filtering. Inria Research Report (RR-2567).

[8] CÉROU, F., DEL MORAL, P. & GUYADER, A. (2011). A non-asymptotic variance theorem for un-normalized Feynman-Kac particle models. *Ann. Inst. Henri Poincare*, **47**, 629–649.

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

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

[11] DEL MORAL, P., DOUCET, A. & JASRA, A. (2006). Sequential Monte Carlo samplers. *J. R. Statist. Soc. B*, **68**, 411–436.

[12] DEL MORAL, P., DOUCET, A. & JASRA, A. (2012). On adaptive resampling procedures for sequential Monte Carlo methods. *Bernoulli*, **18**, 252–272.

[13] DEL MORAL, P., DOUCET, A. & JASRA, A. (2012). An adaptive sequential Monte Carlo method for approximate Bayesian computation. *Statist. Comp.*, **22**, 1223–1237.

[14] DENISON, D. G. T.,HOLMES, C. C., MALLICK, B. K. & SMITH, A. F. M. (2002). *Bayesian Methods for Nonlinear Classification and Regression*, Wiley: New York.

[15] DOUCET, A., GODSILL, S. & ANDRIEU, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statist. Comp.*, **3**, 197–208.

[16] EBERLE, A., & MARINELLI, C. (2013). Quantitative approximations of evolving probability measures and sequential MCMC methods. *Prob. Theo. Rel. Fields*, **155**, 665–701.

[17] JARZYNSKI, C., 1997. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, **78**, 2690–2693.

[18] KANTAS, N., BESKOS A., JASRA, A., CRISAN, D. & STUART, A. (2013). Sequential Monte Carlo Samplers for inverse problems related to the Navier Stokes equation. In preparation.

[19] KONTOYIANNIS, I. & MEYN, S. P. (2005). Large deviation asymptotics and the spectral theory of multiplicatively regular Markov processes. *Elec. J. Probab.*, **10**, 61–123.

[20] NEAL, R. M. (2001). Annealed importance sampling. *Statist. Comp.*, **11**, 125–139.

[21] REBESCHINI, P. & VAN HANDEL, R. (2013). Can local particle filters beat the curse of dimensionality? arxiv preprint.

[22] ROBERTS, G. O., GELMAN, A. & GILKS W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.*, **7**, 110–120.

[23] ROBERTS, G.O., and ROSENTHAL, S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Statist. Soc. B*, **60**, 255–268.

[24] ROBERTS, G.O., and ROSENTHAL, S. (1998). Optimal scaling for various Metropolis-Hastings algorithms. *Statist. Sci.*, **16**, 351–367.

[25] SCHWEIZER, N. (2012). Non-asymptotic error Bounds for sequential MCMC and stability of Feynman-Kac propagators. arxiv preprint.

[26] SNYDER, C., BENGTSSON, T., BICKEL, P., & ANDERSON, J. (2008). Obstacles to high-dimensional particle filtering. *Month. Weather Rev.*, **136**, 4629–4640.

[27] WHITELEY, N. (2012). Sequential Monte Carlo samplers: Error bounds and insensitivity to initial conditions. *Stoch. Anal.*, **5**, 774-798.

[28] WHITELEY, N., KANTAS, N, & JASRA, A. (2012). Linear variance bounds for particle approximations of time homogeneous Feynman-Kac formulae. *Stoch. Proc. Appl.*, **122**, 1840–1865.