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

Alexandros Beskos[1], Dan Crisan[2], Ajay Jasra[3], and Nick Whiteley[4]

[1] *Department of Statistical Science, University College London, London WC1E 6BT, UK.*
E-Mail: `alex@stats.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 a recent paper [3], the Sequential Monte Carlo (SMC) sampler introduced in [12, 19, 24] has been shown to be asymptotically stable in the dimension of the state space d at a cost that is only polynomial in d, when N the number of Monte Carlo samples, is fixed. More precisely, it has been established that the effective sample size (ESS) of the ensuing (approximate) sample and the Monte Carlo error of fixed dimensional marginals will converge as d grows, with a computational cost of $\mathcal{O}(Nd^2)$. In the present work, further results on SMC methods in high dimensions are provided as $d \to \infty$ and with N 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. We also establish marginal propagation of chaos properties of the algorithm. The accuracy in high-dimensions of some approximate SMC-based filtering schemes is also discussed.

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

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 [12, 24]). 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. [16]), resampling is adopted, sometimes performed when the 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 ([2, 4, 28]) 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, [3] have established that the SMC sampler technique can be stabilized at a cost that is only polynomial in d. It was shown in [3] 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 d also has been analyzed recently [30].

The objective of this article is to provide a more complete understanding of SMC algorithms in high dimensions, complementing and building upon the results of [3]. A variety of results are presented, addressing some generic theoretical properties of the algorithms and some issues which arise from specific classes of application.

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. Resampling at the very final time-step is often of importance in real applications; see e.g. [14].

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, 11]. Among the interesting results that have been proved in the literature is the unbiased 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 [11, 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 distribution of any 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 [3, 30], this establishes that the ergodicity of the Markov kernels used in the algorithm can provide stability of the algorithm, even in high dimensions if the number of artificial densities is scaled appropriately with d.

The final issue we address is the problem of filtering (see Section 5 for a description). Ultimately, we do not provide any analysis of a practical SMC algorithm which stabilizes as d increases. However, it is shown that when one inserts an SMC sampler in-between the arrival of each data-point and updates the entire collection of states, then the algorithm stabilizes as d grows at a cost which is $\mathcal{O}(n^2 N d^2)$, n being the time parameter. This is of limited practical use, as the computational storage costs increase with n. Motivated by the SMC sampler results, we consider some strategies which could be attempted to stabilize high-dimensional SMC filtering algorithms. In particular, we address two strategies which insert an SMC sampler in-between the arrival of each data-point. The first, which only updates the state at the current time-point, fails to stabilize as the dimension grows, unless the particles increase exponentially in the dimension. The second, which uses a marginal SMC approach (see e.g. [27]) also exhibits the same properties. At present we are not aware of any online (i.e. one which has a fixed computational cost per time step) SMC algorithm which can provably stabilize with *d for any model*, unless N is exponential in d. It is remarked, as noted in [4], that there exist statistical models for which SMC methods can work quite well in high-dimensions. In relation to this, we then investigate the SMC simulation of a filter based upon Approximate Bayesian Computation (ABC) [20]. The ABC approximation induces bias which cannot be removed in practice. We show here that the simulation error stabilizes as the dimension grows, but we argue that the bias is likely to explode as d grows. This discussion is also relevent for the popular ensemble Kalman filter (EnKF) employed in high dimensional filtering problems in physical sciences (e.g. [17]).

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]; this latter study connects with our discussion in Section 5 on filtering. Section 4 considers the practical implications of our main results with numerical simulations. The filtering problem is addressed in Section 5. We conclude with a summary in Section 6. 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 $\mathbb{V}\mathrm{ar}$ the variance. Finally, $(\Rightarrow)$ denotes convergence in distribution.

2 Framework

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 unnormalized densities associated to Π_n.

One can sample from $\{\Pi_n\}$ using an SMC sampler that targets the sequence of densities:

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

with domain $(E^d)^n$ of dimension that increases with $n = 1, \ldots, p$; here, $\{L_n\}$ is a sequence of artificial backward Markov kernels that can, in principle, be arbitrarily selected ([12]). Let $\{K_n\}$ be a sequence of Markov kernels of invariant density $\{\Pi_n\}$ and Υ a distribution; assuming the weights appearing in the statement of the algorithm are well-defined Radon Nikodym derivatives, the SMC algorithm we will ultimately explore is the one defined in Figure 1. It is remarked that our analysis is not necessarily constrained to the case of resampling according to ESS.

---

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

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

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

1. If $n \le p$, for each i sample $X_n^i \mid X_{n-1}^i = x_{n-1}^i$ from $K_n(x_{n-1}^i, \cdot)$ 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 $w_{l:n}^i \equiv 1$, $1 \le i \le N$;

    let $\check{x}_n^1, \ldots, \check{x}_n^N$ now denote the resampled particles.

Set $n = n + 1$.

Return to the start of Step 1.

Figure 1: The SMC samplers algorithm analyzed in this article.

For simplicity, we will henceforth assume that $\Upsilon \equiv \Pi_0$. It should be noted that when Υ is different from Π_0, one can modify the sequence of densities to a bridging scheme which moves from Υ to Π. However, in practice, one can make Π_0 as simple as possible so we do not consider this possibility; see [30] for more discussion and analysis when d is fixed (note that our results for SMC samplers will hold, with some modifications, also in this scenario). Note, that we only consider here the multinomial resampling method.

We will investigate the stability of SMC estimates associated to the algorithm in Figure 1. To obtain analytical results we will need to simplify the structure of the algorithm. In particular, we will consider an i.i.d. target:

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

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

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

It is remarked that this assumption is made for mathematical convenience: see [3] for a discussion on this. A further assumption that will facilitate the mathematical analysis is to apply 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 [30]) the case when the annealing sequence is derived via a more general non-decreasing Lipschitz function; see Section 3.2.1. As in [3], 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 [3], 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 [3] 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 [3]. We remark that all results of [3] 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 $s \in (\phi_0, 1]$ the state-space E is $(1, \theta, \varsigma)$-small with respect to k_s.

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

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

Note that the statement that E is $(1, \theta, \varsigma)$-small w.r.t. to k_s means that E is a one-step small set for the Markov kernel, with minorizing distribution $\varsigma \in \mathscr{P}(E)$ and parameter $\theta \in (0, 1)$ (i.e. $k_s(x, A) \ge \theta\,\varsigma(A)$ for each $(x, A) \in E \times \mathscr{E}$).

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 in Figure 1) and provide a mathematically convenient framework for understanding the impact of resampling on the properties of the algorithm. Examples of such times can be found in [3, 13]; 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 deterministic ones. We do not consider that here as it would follow a similar proof to [3], depending upon how the resampling times are defined. (An alternative procedure of treating dynamic resampling times is to use the construction in [1]; this is not considered here.) For simplicity, we will henceforth assume that d is large enough so that $m^*(d) \equiv m^*$.

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 $1 \leq i \leq N$. The work in [3] illustrates stability of the normalised weights as $d \to \infty$. Define the standardised log-weights:

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

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:

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

and in particular the variances:

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

The following weak limit can be derived from the proof of Theorem 4.1 of [3].

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

$$\left(\frac{1}{d} \sum_{j=1}^{d} G_{k,j}^i \right)_{i=1}^{N} \Rightarrow (Z^i)_{i=1}^{N} \ ,$$

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

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.

3.1   Asymptotic Results as $d \to \infty$

3.1.1 $\mathbb{L}_2-$Error

The first result of the paper pertains to the Monte-Carlo error from estimates derived via the SMC method. 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 $t = 1$. Below, recall that the $\check{X}$-notation is for resampled particles. 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:*

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

for

$$M(\sigma_{t_{m^*-1}:1}^2) = e^{\sigma_{t_{m^*-1}:1}^2} + M \, e^{17\sigma_{t_{m^*-1}:1}^2} \frac{1}{N^{1/6}}$$

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

Remark 3.1. *Compared to the i.i.d. sampling scenario, the upper bound contains the additional term $\mathbb{V}\mathrm{ar}_\pi[\varphi] \frac{1}{N} M(\sigma_{t_{m^*-1}:1}^2)$. This is a bound on the cost induced due to the dependence of the particles.*

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:

$$c_d := \frac{\int_{E^d} \Pi(x)dx}{\int_{E^d} \Pi_{\phi_0}(x)dx} \ . \tag{8}$$

We first consider the SMC sampler in Figure 1 with no resampling. Define:

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

where $\bar{G}_j^i = (1-\phi_0) \sum_{n=0}^{d-1} g(X_{n,j}^i)$. From standard properties of SMC we have $\mathbb{E}\left[\gamma_d^N(1) \right] = \gamma_d(1) \equiv c_d$. Now, consider the relative $\mathbb{L}_2$-error:

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

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 $N \geq 1$:*

$$\lim_{d\to\infty} \mathbb{V}_2(\gamma_d(1)) = \tfrac{e^{\sigma_{\phi_{0:1}}^2} - 1}{N} .$$

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 $\sigma_{\phi_{0:1}}^2$. As in [3], this is a critical quantity, which helps to measure the rate of convergence of the algorithm.

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

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

where $k \in \{1,\ldots,m^*+1\}$ and $\bar{G}_{k,j}^i$ as defined in Section 2.3. As in the non-resampling case, we again have the unbiasedness property for the estimate of c_d: $\mathbb{E}\left[\prod_{k=1}^{m^*+1} \gamma_{d,k}^N(1) \right] = \prod_{k=1}^{m^*+1} \gamma_{d,k}(1) \equiv c_d$. 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 $N \geq 1$:*

$$\lim_{d\to\infty} \mathbb{V}_2\Big(\prod_{k=1}^{m^*+1} \gamma_{d,k}(1) \Big) = e^{-\sigma_{\phi_{0:1}}^2} \prod_{k=1}^{m^*+1} \left[\tfrac{1}{N} e^{2\sigma_{t_{k-1}:t_k}^2} + \left(1 - \tfrac{1}{N}\right) e^{\sigma_{t_{k-1}:t_k}^2} \right] - 1.$$

Remark 3.2. *In comparison with the no resampling scenario, the limiting expression here depends upon the incremental variance expressions. On writing the limit in the form:*

$$e^{-\sigma_{\phi_{0:1}}^2} \prod_{k=1}^{m^*+1} e^{\sigma_{t_{k-1}:t_k}^2} \left[1 + \tfrac{1}{N}\{ e^{\sigma_{t_{k-1}:t_k}^2} - 1 \} \right] - 1$$

if $N > (m^*+1)(e^{\bar{\sigma}^2} - 1)$, with $\bar{\sigma}^2 = \max_k \sigma_{t_{k-1}:t_k}^2$, then using similar manipulations to [8, Corollary 5.2], we have that

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

hence, the no resampling scenario has a lower error if this is less than the limit in Theorem 3.2. The upper-bound also shows that the error seems to grow with number of times one resamples. The limit in

Theorem 3.3 corresponds to the behavior of $\mathbb{E}_{\pi^{\otimes Nd}_{t_{k-1}(d)}}[\gamma^N_{d,k}(1)^2/\gamma_{d,k}(1)^2]$ between each resampling time. 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 of the following can be found 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:*

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

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, demonstrating the asymptotic (in the classical setting) independence of any fixed block of q of N particles as N grows. The following scenario, with the SMC sampler with resampling (at the 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 $s(d) \to s$, 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

In this section, we 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). However, 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 Theorem 2.1 (and others that will follow) generalised to the choice of $\phi_{n,d}$ considered here would involve the variances:

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

in the place of $\sigma_{s:t}^2$ 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 prefers an annealing scheme which discretizes $\phi(s)$ versus one which discretizes $\nu(s)$ (differentiable monotonic increasing Lipschitz function with $\nu(0) = \nu_0 > 0$, $\nu(1) = 1$) for estimating normalizing constants if

$$\frac{e^{\sigma_{0:1}^{2,\phi}} - 1}{N} \leq \frac{e^{\sigma_{0:1}^{2,\nu}} - 1}{N} \quad \Longleftrightarrow \quad \sigma_{0:1}^{2,\phi} \leq \sigma_{0:1}^{2,\nu} .$$

In practice, however, one has to numerically approximate $\sigma_{0:1}^{2,\phi}$ and $\sigma_{0:1}^{2,\nu}$, lessening the practical impact of this result. Similar to the scenario of Theorem 3.2, one can use Theorem 3.3 to compare annealing schemes $\phi(s)$ and $\nu(s)$ (which potentially generate a different collection and number of limiting times $0 < t_1^\phi < \cdots < t_{m_\phi^*}^\phi \leq 1$, m_ϕ^* and $0 < t_1^\nu < \cdots < t_{m_\nu^*}^\nu \leq 1$, m_ν^* respectively) via the inequality

$$e^{-\sigma_{0:1}^{2,\phi}} \prod_{k=1}^{m_\phi^*+1} \left[\frac{1}{N}e^{2\sigma_{t_{k-1}^\phi:t_k^\phi}^{2,\phi}} + \frac{N-1}{N}e^{\sigma_{t_{k-1}^\phi:t_k^\phi}^{2,\phi}}\right] \leq e^{-\sigma_{0:1}^{2,\nu}} \prod_{k=1}^{m_\nu^*+1} \left[\frac{1}{N}e^{2\sigma_{t_{k-1}^\nu:t_k^\nu}^{2,\nu}} + \frac{N-1}{N}e^{\sigma_{t_{k-1}^\nu:t_k^\nu}^{2,\nu}}\right]$$

but again, both quantities are difficult to calculate.

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 $y_1, \ldots, y_p$. 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:

$$\Pi(x) \propto \exp\left\{\sum_{k=1}^p \sum_{j=1}^d g(y_k, x_j)\right\}, \quad x_j \in E ,$$

that is a density that is i.i.d. in both the data and dimension. In this scenario one could then adopt a sequence of densities of the form:

$$\Pi_n(x) \propto \exp\left\{ \sum_{k=1}^n \sum_{j=1}^d g(y_k, x_j) \right\}, \quad 1 \leq n \leq p .$$

As noted also in Remark 3.5 of [3], for $d \to \infty$ one cannot stabilize the associated SMC algorithm (described in Figure 1) as the ratio Π_{n+1}/Π_n explodes for increasing d. To stabilize the algorithm as d grows one can insert $\lfloor d/p \rfloor$ annealing steps between consecutive data points, thus forming the densities:

$$\Pi_k^{(n)}(x) \propto \exp\left\{ \phi\left(\tfrac{k}{\lfloor d/p \rfloor} \right) \sum_{j=1}^d g(y_n, x_j) \right\} \times \Pi_{n-1}(x) , \quad n \in \{1, \ldots, p\}, \ k \in \left\{1, \ldots, \lfloor \tfrac{d}{p} \rfloor\right\} ,$$

where $\phi(s)$ is as in Section 3.2.1 with $\phi_0 = 0$. Then one can adopt Markov kernels $K_s^{(n)}$, $1 \leq n \leq p$, of product form with each component kernel $k_s^{(n)}$ having invariant measure

$$\pi_s^{(n)}(x) \propto \exp\left\{ s \cdot g(y_n, x) + \sum_{k=1}^{n-1} g(y_k, x) \right\}, \quad x \in E .$$

We consider the scenario where there is no resampling and denote by $\mathrm{ESS}_{(0,d)}(n, N)$ the effective sample size with n data after $\lfloor d/p \rfloor$ steps of the n-th SMC sampler. Throughout the data are taken as fixed. We have the following result, which follows directly from Theorem 3.1. of [3] and illustrates the stability of $\mathrm{ESS}_{(0,d)}(n, N)$ as $d \to \infty$.

Proposition 3.3. *Assume conditions (A1-2) for the kernels $k_s^{(n)}$, $n \in \{1, \ldots, p\}$. Suppose that for each data index $n \in \{1, \ldots, p\}$, $g(y_n, \cdot) \in \mathcal{B}_b(E)$. 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

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

In particular,

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

As for the case with annealing densities, there is a direct extension to the case where one resamples (see [3]). In addition, one can easily extend the results in Section 3 in the data-point tempering case examined here. In connection to the filtering scenario, 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 hints at algorithms for filtering which may stabilize as the dimension grows. As we shall see in Section 5, it is not straight-forward to do this.

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:

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

We will in fact consider two annealing schemes:

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

These are graphically displayed in Figure 2 (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 2 (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 $\pi_{\phi(u)}(\widehat{g}^2_{\phi(u)} - k_{\phi(u)}(\widehat{g}_{\phi(u)})^2)$ in (9). In this scenario it is simple to show that, under the choice (10), we have that:

$$\sigma^{2,\phi}_{0:1} = \frac{1}{2} \int_0^1 \left[\frac{1}{\phi(u)}\frac{d\phi(u)}{du}\right]^2 du \; .$$

Figure 2 (b) now plots the analytically available variances $\sigma^{2,\phi}_{0:1}$ and $\sigma^{2,\nu}_{0:1}$ (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

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 $\nu(s)$.

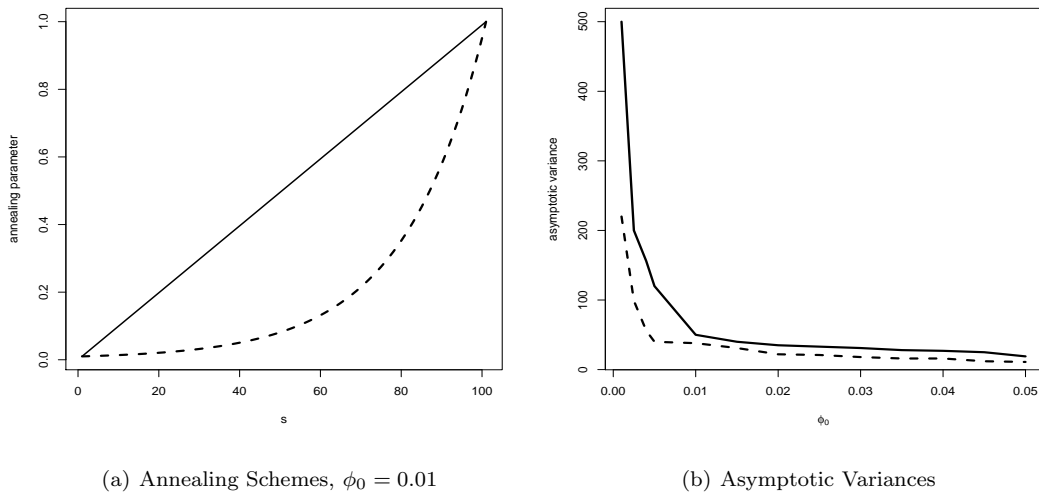(a) Annealing Schemes, $\phi_0 = 0.01$ (b) Asymptotic Variances

Figure 2: 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 $\sigma^{2,\phi}_{0:1}$ and $\sigma^{2,\nu}_{0:1}$ (broken line) against ϕ_0 for the case of exact sampling with $k_s(x, dx') \equiv \pi_s(dx')$ and the scenario (10).

of the algorithm, and calculated the corresponding realisations of the log Ratio:

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

(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.

4.2 Bayesian Linear Model

We now consider the implications of main results in the context of a Bayesian linear regression model (see [15] for a book-length introduction as well as a wealth of practical applications). This is a statistical model that associates a p-vector of responses, say Y, to a $p \times d$-matrix of explanatory variables, X,

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 repetitions 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.

for some $p \geq 1$, $d \geq 1$. In particular:

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

where β is a d-vector of unknown regression coefficients and $\epsilon \sim N_p(0, \mathbf{1}_p)$, with $\mathbf{1}_p$ the $p \times p$ 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 $d = 50$, $p = 50$ 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 $\lfloor 10d/p \rfloor$ steps between the p 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 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 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   Filtering

An important application of SMC methods is filtering. In the following we will look at the effect of dimension for several filtering algorithms.

5.1   Set-Up

Consider the discrete-time filtering problem; we have fixed observations $y_1, \ldots, y_n$, with $y_k \in \mathbb{R}^{d_y}$ and a hidden Markov chain $X_1, \ldots, X_n$, with $X_k \in E^d$ such that the Y_k's are conditionally independent of all other variables, given X_k. We assume that the density w.r.t. Lebesgue measure is

$$g(y_k|x_k) = \exp\left\{ \sum_{j=1}^{d} h(y_k, x_{k,j}) \right\} \tag{13}$$

with $x_{k,j} \in E$ and $h : \mathbb{R}^{d_y} \times E \to \mathbb{R}$. The hidden Markov chain is taken to be time-homogeneous with transition density w.r.t. Lebesgue measure:

$$F(x_k|x_{k-1}) = \prod_{j=1}^{d} f(x_{k,j}|x_{k-1,j}) \ , \quad k \geq 1 \ ,$$

where $x_0 = (x_{0,1}, x_{0,2}, \ldots, x_{0,d}) \in E^d$ is a given fixed point and $\int_E f(x'|x)dx' = 1$. This is certainly a very specific model structure chosen, as in the previous part of the paper, for mathematical convenience. Clearly, our results depend on the given structure; it is certainly the case that for some other classes of state-space models standard SMC methods could stabilize with the dimension of the problem (as could be the case for instance when the lihelihood in (13) involved only a few of the co-ordinates of x).

The objective in filtering is to compute for a π-integrable function φ:

$$\mathbb{E}\left[\, \varphi(X_n)\,|\,y_{1:n}\,\right] = \int_E \varphi(x)\pi_n(x_n|y_{1:n})dx_n \tag{14}$$

where

$$\pi_n(x_n|y_{1:n}) = \frac{\int_{E^{(n-1)d}} \prod_{k=1}^{n} g(y_k|x_k)F(x_k|x_{k-1})dx_{1:n-1}}{\int_{E^{nd}} \prod_{k=1}^{n} g(y_k|x_k)F(x_k|x_{k-1})dx_{1:n}} \ .$$

It should be noted that one can re-write the filter via the standard prediction-updating formula:

$$\pi_n(x_n|y_{1:n}) = \frac{g(y_n|x_n)\pi_{n|n-1}(x_n|y_{1:n-1})}{p(y_n|y_{1:n-1})} \ ;$$

$$\pi_{n|n-1}(x_n|y_{1:n-1}) = \int_{E^d} F(x_n|x_{n-1})\pi_{n-1}(x_{n-1}|y_{1:n-1})dx_{n-1} \ ;$$

$$p(y_n|y_{1:n-1}) = \frac{\int_{E^{nd}} \prod_{k=1}^{n} g(y_k|x_k)F(x_k|x_{k-1})dx_{1:n}}{\int_{E^{(n-1)d}} \prod_{k=1}^{n-1} g(y_k|x_k)F(x_k|x_{k-1})dx_{1:n-1}} \ .$$

Typically, one cannot calculate (14), so we resort to particle filtering. The most basic approach is to perform an SMC algorithm, which approximates the sequence of densities

$$\pi(x_{1:n}|y_{1:n}) \propto \prod_{k=1}^{n} g(y_k|x_k)F(x_k|x_{k-1})$$

by using the prior dynamics, characterised by F, as a proposal. This yields an un-normalized incremental weight at time step n of the algorithm which is $g(y_n|x_n)$. One can then resample or not. We consider the scenario as the dimension of the state increases, when the data record is fixed (i.e. we keep the time parameter n and data fixed).

In reference to the works [2, 4, 6, 28], it is clear that the standard particle filter cannot be used in general to approximate filters with high-dimensional states. An alternative, as mentioned for the data-point tempering scenario in Section 3.2.2 (see also [4, 18]) is to insert an annealing SMC sampler between consecutive filtering steps, updating the entire trajectory $x_{1:n} \in E^{nd}$.

Assuming the i.i.d. structure above, the model dynamics decompose over d independent co-ordinates. One could use MCMC kernels for the SMC samplers between arrivals of consecutive data-points, with each univariate kernel (the product of d of them forming the complete kernel) having invariant density:

$$\pi_p^{(n)}(x_{1:n}) \propto \exp\left\{ \phi\!\left(\tfrac{p}{d}\right) h(y_n, x_n) + \log(f(x_n|x_{n-1})) + \sum_{i=1}^{n-1}\{\, h(y_i, x_i) + \log(f(x_i|x_{i-1}))\,\} \right\},$$

for $x_{1:n} \in E^n$, $\phi(0) = 0$, and $0 \le p \le d$. We write these marginal target densities at data-time n on the continuum as $\pi_s^{(n)}$ with associated Markov kernels $k_s^{(n)}$ (which operate on spaces of increasing dimension). No resampling is added to the algorithm, but easily could be. The ESS (see (2) in Figure 1) is denoted $\mathrm{ESS}_{(0,d)}(n, N)$. It is a straight-forward application of Theorem 3.1 of [3] to get that, under assumptions (A1-2) for the kernels $k_s^{(n)}$, $n \ge 1$, and the condition that for each $n \ge 1$, $h(y_n; x_n) \in \mathcal{B}_b(E)$ (with the associated solution to Poisson's equation written $\hat{g}_s^{(n)}$), for any fixed $N > 1$, $n \ge 1$, $\mathrm{ESS}_{(0,d)}(n, N)$ converges in distribution to

$$\frac{[\sum_{i=1}^{N} e^{Z^i}]^2}{\sum_{i=1}^{N} e^{2Z^i}}$$

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

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

In particular,

$$\lim_{d\to\infty} \mathbb{E}\left[\,\mathrm{ESS}_{(0,d)}(n, N)\,\right] = \mathbb{E}\left[\frac{[\sum_{i=1}^{N} e^{Z^i}]^2}{\sum_{i=1}^{N} e^{2Z^i}}\right]. \tag{15}$$

Thus the cost for this algorithm to be stable as $d \to \infty$, is $\mathcal{O}(n^2 d^2 N)$. This result is not surprising; one updates the whole state-trajectory and the stability proved in [3] is easily imported into the algorithm. However, this algorithm is not online and will be of limited practical significance unless one has access to substantial computational power. The result is also slightly misleading: it assumes that the MCMC kernels have a uniform mixing with respect to the time parameter of the HMM (see condition A1). This is unlikely to hold unless one increases the computational effort, associated to the MCMC kernel, with n.

One apparent generalization would be to use SMC samplers at each data-point time to sample from the annealed smoothing densities, except freezing the first $n-1$ co-ordinates; it is then easily seen that one does not have to store the trajectory. However, one can use the following intuition as to why this procedure will not work well (the following is based upon personal communication with Prof. A. Doucet). In the idealized scenario, one samples exactly from the final target density of the SMC sampler. In this case, the final target density is exactly the conditionally optimal proposal (see [16]) and the incremental weight is:

$$\int_{E^d} g(y_n|x_n)F(x_n|x_{n-1})dx_n = \prod_{j=1}^{d}\int_E e^{h(y_n,x_{n,j})}f(x_{n,j}|x_{n-1,j})dx_{n,j} \ ,$$

which will typically have exponentially increasing variance in d. We conjecture that similar issues arise for advanced SMC approaches such as [10].

5.2 Marginal Algorithm

Due to the obvious instability of the above procedure, we consider another alternative, presented e.g. in [27]. This algorithm would proceed as follows. When targeting the filter at time 1, one adopts an SMC sampler, which as discussed before, will stabilize with the dimension. Then at subsequent time-steps, to consider the initial density of the SMC sampler:

$$\sum_{l=1}^{N}\overline{w}_d^{l,(n-1)}\prod_{j=1}^{d}f(x_{0,j}^{l',(n)}|x_{d,j}^{l,(n-1)}) \tag{16}$$

where we have defined:

$$\overline{w}_{0:d-1}^{l,(n-1)} \propto \exp\left\{\tfrac{1}{d}\sum_{j=1}^{d}\sum_{i=0}^{d-1}h(y_{n-1},x_{i,j}^{l,(n-1)})\right\} \ ; \quad \sum_{l=1}^{N}\overline{w}_d^{l,(n-1)}=1 \ ,$$

as the normalized weight. One could then resample and apply SMC samplers on the sequence of target distributions (e.g. with $n=2$)

$$\pi_k^{(n)}(x_{1:d}) \propto \exp\left\{\phi\big(\tfrac{k}{d}\big)\sum_{j=1}^{d}h(y_n,x_j)\right\}\Big[\sum_{l=1}^{N}\prod_{j=1}^{d}f(x_j|\check{x}_{d,j}^{l,(n-1)})\Big] \ , \quad k\in\{0,\dots,d\} \ , \tag{17}$$

where $\check{x}_{d,j}^{l,(n-1)}$ is the resampled particle when sampling from (16).

Using simple intuition, one might expect that this SMC algorithm may stabilize as the dimension grows. For example, if one could sample exactly from $\pi_p^{(d)}(x_{1:d})$, then the importance weight is exactly 1; there is no weight degeneracy problem: As proved in [3], under ergodicity assumptions, the SMC sampler will asymptotically produce a sample from the final target density.

However, the following result suggests that the algorithm will collapse unless the number of particles grows exponentially fast with the dimension. Consider the case with N_d particles, where N depends on d here, and these are samples exactly from the previous filter; denote the samples $\check{X}_{1:d}^{1:N_d}$. Conditionally

upon $\check{X}_{1:d}^{1:N_d}$, sample $X_{1:d}^{1:N_d}$ exactly from the approximation $\pi_d^{(n)}$ (17). This presents the most optimistic scenario one could hope for. Denote below, $\overline{\varphi}_n(x) = \varphi(x) - \pi_n(\varphi)$. We have the following result.

Proposition 5.1. *Consider the algorithm above so that for any $(x, x') \in E^2$ we have $f(x|x') \in (\underline{f}, \overline{f})$ for constants $0 < \underline{f} < \overline{f} < \infty$ and for any $(y, x) \in \mathbb{R}^{d_y} \times E$, $h(y|x) \in (\underline{h}, \overline{h})$ for $0 < \underline{h} < \overline{h} < \infty$. Suppose $\varphi \in \mathcal{B}_b(E)$. Then, there exist an $M < \infty$ and $\kappa > 1$ such that for any $d \geq 2$, $j \in \{1, \ldots, d\}$ and $N_d \geq 1$ we have*

$$\mathbb{E}\left[\left(\tfrac{1}{N_d}\sum_{i=1}^{N_d} \overline{\varphi}_n(X_j^i)\right)^2\right] \leq \frac{M\kappa^d}{N_d} \ .$$

Proof. Throughout the proof, write $\pi_d^{(n)}$ as the approximated filter at time n. Then we have the simple decomposition:

$$\mathbb{E}\left[\left(\tfrac{1}{N_d}\sum_{i=1}^{N_d} \varphi(X_j^i) - \pi_n(\varphi)\right)^2\right] = \mathbb{E}\left[\left(\tfrac{1}{N_d}\sum_{i=1}^{N_d} \varphi(X_j^i) - \pi_d^{(n)}(\varphi) + \pi_d^{(n)}(\varphi) - \pi_n(\varphi)\right)^2\right] \ .$$

On applying the C_2-inequality, we can decompose the error into the one of the Monte Carlo error of approximating expectations w.r.t. $\pi_d^{(n)}$ and that of approximating the filter.

Consider the first error. Conditioning on the i.i.d. samples drawn from the filter at time $n-1$, one may apply the Marcinkiewicz-Zygmund inequality and use the i.i.d. property of the algorithm to obtain the upper bound:

$$\left(\tfrac{M}{N_d}\right)\mathbb{E}\left[\frac{\tfrac{1}{N_d}\sum_{i=1}^{N_d} f(\varphi^2 e^h)(\check{X}_j^i)\prod_{l \neq j} f(e^h)(\check{X}_l^i)}{\tfrac{1}{N_d}\sum_{i=1}^{N_d}\prod_{l=1}^d f(e^h)(\check{X}_l^i)} - \left\{\frac{\tfrac{1}{N_d}\sum_{i=1}^{N_d} f(\varphi e^h)(\check{X}_j^i)\prod_{l \neq j} f(e^h)(\check{X}_l^i)}{\tfrac{1}{N_d}\sum_{i=1}^{N_d}\prod_{l=1}^d f(e^h)(\check{X}_l^i)}\right\}^2\right] \ .$$

As φ and h are bounded, it is easily seen that the function in the expectation is uniformly bounded in d and hence that one has an upper-bound of the form M/N_d.

Now to deal with the second error, this can be written:

$$\mathbb{E}\left[\left(\frac{\tfrac{1}{N_d}\sum_{i=1}^{N_d} f(\varphi e^h)(\check{X}_j^i)\prod_{l \neq j} f(e^h)(\check{X}_l^i)}{\tfrac{1}{N_d}\sum_{i=1}^{N_d}\prod_{l=1}^d f(e^h)(\check{X}_l^i)} - \frac{\pi_{n-1}f(\varphi e^h)\pi_{n-1}f(e^h)^{d-1}}{\pi_{n-1}f(e^h)^d}\right)^2\right] \ .$$

The bracket can be decomposed into the form:

$$\left(\frac{\tfrac{1}{N_d}\sum_{i=1}^{N_d} f(\varphi e^h)(\check{X}_j^i)\prod_{l \neq j} f(e^h)(\check{X}_l^i)}{\{\pi_{n-1}f(e^h)^d\}\tfrac{1}{N_d}\sum_{i=1}^{N_d}\prod_{l=1}^d f(e^h)(\check{X}_l^i)}\right)\left(\pi_{n-1}f(e^h)^d - \tfrac{1}{N_d}\sum_{i=1}^{N_d}\prod_{l=1}^d f(e^h)(\check{X}_l^i)\right) +$$

$$\frac{1}{\pi_{n-1}f(e^h)^d}\left(\tfrac{1}{N_d}\sum_{i=1}^{N_d} f(\varphi e^h)(\check{X}_j^i)\prod_{l \neq j} f(e^h)(\check{X}_l^i) - \pi_{n-1}f(\varphi e^h)\pi_{n-1}f(e^h)^{d-1}\right) \ .$$

Applying the C_2-inequality again, we can break up the two terms. Using the lower bound on h and upper-bounds on φ and h the $\mathbb{L}_2$-error of the first term is upper-bounded by:

$$\frac{\|\varphi\|_\infty^2 e^{2\overline{h}}}{\pi_{n-1}f(e^h)^{2d}e^{2\underline{h}}}\mathbb{E}\left[\left(\pi_{n-1}f(e^h)^d - \tfrac{1}{N_d}\sum_{i=1}^{N_d}\prod_{l=1}^d f(e^h)(\check{X}_l^i)\right)^2\right] = \frac{\|\varphi\|_\infty^2 e^{2\overline{h}}}{N_d\pi_{n-1}f(e^h)^{2d}e^{2\underline{h}}}\mathbb{V}\text{ar}\left[\prod_{l=1}^d f(e^h)(\check{X}_l^1)\right] \ .$$

As the variance on the R.H.S. is easily seen to be equal to $\pi_{n-1}(f(e^h)^{2d}) - \pi_{n-1}(f(e^h))^{2d}$ one yields the upper-bound:

$$\frac{M}{N_d}\left[\frac{\pi_{n-1}(f(e^h)^{2d})}{\pi_{n-1}(f(e^h))^{2d}} - 1\right].$$

For the second term one can follow similar arguments to yield the upper-bound:

$$\frac{1}{N_d \pi_{n-1}[f(e^h)]^{2d}}\left[\pi_{n-1}[f(\varphi e^h)|^2]\pi_{n-1}[f(e^h)^2]^{(d-1)} - \pi_{n-1}[f(\varphi e^h)]^2\pi_{n-1}[f(e^h)]^{2(d-1)}\right]$$

from which one can easily conclude. $\qquad\square$

Remark 5.1. *On inspection of the proof, it is easily seen that one can write the error as $\frac{M}{N_d}(1+\kappa^d)$ which represents two sources of error. The first is the Monte Carlo error due to estimating the marginal expectation w.r.t. the approximation. This appears to be controllable for any N_d converging to infinity. The second source of error is in approximating the filter, which seems to require a number of particles which will increase exponentially in the dimension; this is the drawback of this algorithm.*

Remark 5.2. *We remark that this is only an upper-bound, but we can be even more precise; if one considers the relative $\mathbb{L}_2$-error of the estimate of $p(y_n|y_{1:n-1})$ then this is equal to*

$$\frac{1}{N_d}\left[\frac{\pi_{n-1}(f(e^h)^2)^d}{\pi_{n-1}(f(e^h))^{2d}} - 1\right]$$

which will explode in the dimension, unless N_d grows at an exponential rate. This is in contrast to the SMC sampler case in Section 3, where one can obtain an estimate of the normalizing constant, whose relative $\mathbb{L}_2$-error stabilizes for any N.

5.3   Approximate Bayesian Computation (ABC)

In this section we consider SMC methods in the context of an ABC filter - an approximate filtering scheme which is of practical interest when evaluation of the likelihood function in the state-space model is intractable. We note in passing the connection of the ABC methods to the ensemble Kalman filter [25], a full treatment of the latter is well beyond the scope of the present work.

The idea of this approach, which is primarily adopted when:

- The function $g(y|x_{1:d})$ is intractable, that is, one cannot evaluate it point-wise.

- It is possible to simulate from $g(\cdot|x_{1:d})$ for any $x_{1:d} \in E^d$.

In this scenario, standard SMC methods can be used to sample from an approximation of the smoothing density, of the form (for some $\epsilon > 0$):

$$\pi_\epsilon(x_{1:n}, u_{1:n}|y_{1:n}) \propto \prod_{k=1}^{n} \mathbb{I}_{\{u_k:|y_k-u_k|<\epsilon\}}(u_k)\, g(u_k|x_k)\, \overline{f}(x_k|x_{k-1})\ . \tag{18}$$

Here, the idea is to sample, at each time-point, pseudo-data $u_k \in \mathbb{R}^{d_y}$; the density is non-zero when all of the simulated pseudo data lie within ϵ of the observed data (in $\mathbb{L}_1$-distance). Adopting an SMC algorithm with proposals $g \otimes \overline{f}$ yields an un-normalized incremental weight of the form $\mathbb{I}_{\{u_k:|y_k-u_k|<\epsilon\}}(u_k)$, which circumvents the evaluation of g. In the context of high-dimensional models, as discussed here, the SMC algorithm will collapse using the approaches in the previous sections. However, when d_y is not large, one would expect that indeed, the SMC approximation of the ABC filter should be reasonably stable (in some sense). We quantify this with the following result.

We will assume here conditions (A1-A3) of [20]. In particular, that:

$$\sup_{x_{1:d}} \|g(\cdot, x_{1:d})\|_\infty = e^{d \sup_{x_{1:d}} \|h(\cdot, x_{1:d})\|_\infty} \; ; \quad \sup_{x_{1:d}} |g(y, x_{1:d}) - g(u, x_{1:d})| \leq M |y - u| \ .$$

The latter assumption will typically only hold when $E \subset \mathbb{R}$ with E being compact. We consider only the scenario where no resampling is performed. The expectation with respect to the SMC algorithm conditioned on the fixed data (which is suppressed from the notation) is written as $\mathbb{E}$. Also, we write simply $\mathbb{I}_{A_{y_{1:n},\epsilon}}$ in the place of $\mathbb{I}_{\{u_{1:n}^{1:N} : \sum_{j=1}^N \prod_{k=1}^n \mathbb{I}_{\{u_k:|y_k-u_k|<\epsilon\}}(u_k^j) > 0\}}$.

Proposition 5.2. *Given the set-up above, one has that for any $n \geq 1$, $p \geq 1$, $\varphi \in \mathcal{B}_b(E)$, $\epsilon > 0$ there exists an $M(n, p, \varphi, \epsilon) > 0$, and for $d \geq 1$ there exists $\kappa_n(d, \varphi) > 0$ which does not depend upon p, ϵ such that for any $N \geq 1$:*

$$\mathbb{E}\left[\left| \sum_{i=1}^N \frac{\prod_{k=1}^n \mathbb{I}_{\{u_k:|y_k-u_k|<\epsilon\}}(u_k^i)}{\sum_{j=1}^N \prod_{k=1}^n \mathbb{I}_{\{u_k:|y_k-u_k|<\epsilon\}}(u_k^j)} \overline{\varphi}_n(X_{n,1}^i) \right|^p \mathbb{I}_{A_{y_{1:n},\epsilon}} \right]^{1/p} \leq \frac{M(n,p,\varphi,\epsilon)}{\sqrt{N}} + \kappa_n(d, \varphi)\epsilon \qquad (19)$$

where $\kappa_n(d, \varphi)$, as $d \to \infty$, converges to zero or diverges to infinity.

Proof. Write

$$\pi_{n,\epsilon}(\varphi) := \int \varphi(x_{n,1}) \pi_\epsilon(x_{1:n}, u_{1:n}|y_{1:n}) dx_{1:n} du_{1:n} \ .$$

Then, one can add and subtract this term in the $|\cdot|^p$ and apply Minkowski, leading to:

$$\mathbb{E}\left[\left| \sum_{i=1}^N \frac{\prod_{k=1}^n \mathbb{I}_{\{u_k:|y_k-u_k|<\epsilon\}}(u_k^i)}{\sum_{j=1}^N \prod_{k=1}^n \mathbb{I}_{\{u_k:|y_k-u_k|<\epsilon\}}(u_k^j)} \varphi(X_{n,1}^i) - \pi_{n,\epsilon}(\varphi) \right|^p \mathbb{I}_{A_{y_{1:n},\epsilon}} \right]^{1/p} + |\pi_{n,\epsilon}(\varphi) - \pi_n(\varphi)| \ .$$

The first term is easily dealt with using standard proof techniques in Monte Carlo computation; see for example part of the proof of Theorem 3.3. of [3]. Hence we need only treat the bias term. Following the arguments of Theorem 1 of [20], one can obtain that an upper bound on the bias is:

$$\kappa_n(d, \varphi) = \frac{\epsilon \|\varphi\|_\infty}{\left(\int_{\mathbb{R}^2} e^{h(y_n,x)} f(x|x') \pi_{n-1,1}(x'|y_{1:n-1}) dx dx' \right)^d} \left(2L + \kappa_{n-1}(d, \varphi) e^{d \sup_x \|h(y_n,x)\|_\infty} \right)$$

where $\pi_{n-1,1}$ is the filter at time $n - 1$ marginalized to its first component and $\kappa_0(d, \varphi) = 0$. $\qquad \square$

Remark 5.3. *The above result is of interest for high-dimensional filtering. Essentially it establishes that the SMC approximation of the ABC approximation is stable for any $d \geq 1$, with computational cost of $\mathcal{O}(Nd)$; this is the first term on the R.H.S. of (19). However, the deterministic component of*

the ABC approximation of the filter is likely to deteriorate as $d \to \infty$. For any $n \geq 1$ the sequence $(\kappa_n(d, \varphi))_{d \geq 1}$ is likely to diverge, as for example when $n = 1$ it is proportional to

$$\left(\int_E e^{h(y_1, x_1)} f(x_1 | x_{0,1}) dx_1 \right)^{-d}.$$

Whilst this is only an upper-bound, we will see in Section 5.4 that the error seems to increase with d in simulations.

Remark 5.4. *Due to the link between ABC and EnKF [25] and the bias of the EnKF [23], we conjecture that the EnKF will be subject to a similar behavior as for ABC (for non-linear models). That is, one can numerically approximate the approximation of the filter in high dimensions, but that the approximation collapses as the dimension of the state grows.*

5.4 Numerical Example: Linear Gaussian State-Space Model

In the following example we consider the ABC approximation error of linear Gaussian state-space model, $k \geq 1$:

$$Y_k = H X_k + V_k \; ;$$
$$X_k = X_{k-1} + W_k \; ,$$

where $d_y = 1$, $H = (1, \ldots, 1)$ is a $1 \times d$ vector, $V_k \overset{\text{i.i.d.}}{\sim} N_1(0, 1)$, $X_0 = (0, \ldots, 0)'$ and $W_k \overset{\text{i.i.d.}}{\sim} N_d(0, \mathbf{1}_d)$. 200 data points are generated from the model.

We run the SMC-based ABC algorithm in [20] with the parameter ϵ (see (18) for the approximation of the smoothing density) fixed at 5, with $d \in \{10, 40, 200\}$. The first moment in the first dimension is estimated and the quantity on the L.H.S. of (19) (associated to this) is estimated with $N = 1000$ with 50 repeats. The results can be observed in Figure 5.4. where the estimate of the L.H.S. of (19) over times 1 to 200 for $d = 200$ against $d = 10$ (black) and $d = 40$ (broken blue) is plotted. It appears that the error grows with d. It is remarked that in general as the model is not i.i.d. (in dimension) one cannot guarantee a uniform per-time step increase in the error. However, the relative increase in the error, w.r.t. dimension, is consistent with our empirical experience with applying the algorithm.

5.5 Discussion on High-Dimensional Filtering

One of the motivations of our work was to investigate the issue of stability in high dimensions of SMC algorithms for filtering. As can be seen, there is still much scope for future work and this issue is far from resolved. In particular and in relation to this issue (as established in Section 5.3) what is currently missing in the literature is a concerted effort from probabilists, statisticians and applied mathematicians on the analysis of algorithms used in data assimilation (at least one exception is [23]).

In relation to the above discussion, one potentially very fruitful starting point, is the filtering of the Navier stokes equation; see [5]. In this context, one is given the 2-dimensional Navier-Stokes equation
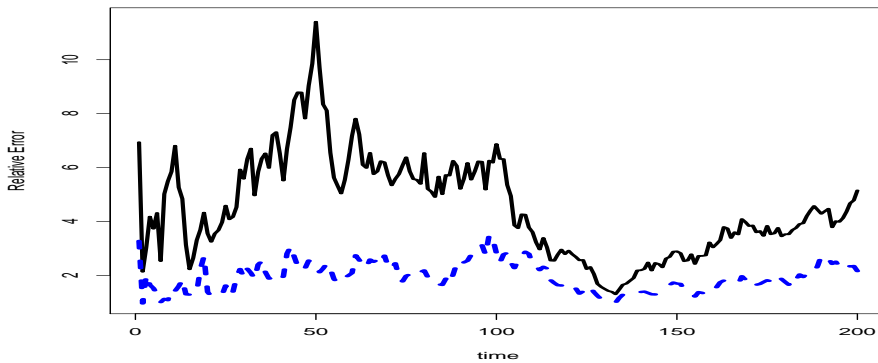
Figure 3: A Plot of the Relative $\mathbb{L}_2$-Error of the ABC Filter Algorithm. The black line is the $\mathbb{L}_2$-error for estimating the first moment in the first dimension for 200 dimensions against 10 dimensions over 200 time steps. The broken blue line is similar except for 200 dimensions versus 40 dimensions.

on a torus and the objective is to infer the initial condition of the PDE, given access to noisy data. In mathematical terms, given a Gaussian prior on x the initial condition, on Hilbert space $\mathcal{H}$ one seeks to deal with the density of the filter w.r.t. the prior (see [5, Theorem 3.2])

$$\pi_k(x|y_{1:k}) \propto \exp\{-\Phi_k(y_{1:k}, x)\}$$

where $x \in \mathcal{H}$ and $\Phi_k : \mathbb{R}^k \times \mathcal{H} \to \mathbb{R}$ a potential function associated to observed data $y_{1:k} \in \mathbb{R}$. This is an example of trying to filter a state with deterministic dynamics, albeit in infinite dimensions. Given the analysis in Section 3.2.2, it is likely that, by defining a data-point tempering SMC algorithm on the infinite dimensional space, one can consider the associated finite-dimensional algorithm (with state-vector in $\mathbb{R}^d$) and establish the stability as d grows. This is assuming one has sufficiently mixing MCMC kernels; see [26] for some ideas. This issue is a subject of current research jointly with Prof. A. Stuart.

More generally, one is interested in the issue of when the dynamics of the hidden state are Markovian. As noted here and in more details in [4], the underlying structure of the state-space model is important for establishing some sort of stability in high dimensions of a numerical filtering algorithm, SMC or otherwise. Akin to the standard problem of filter-stability in time (e.g. [29]), perhaps considerable research is needed with regards to dimension of the deterministic (true) filter, before a full analysis of numerical algorithms can be undertaken. However, it is not obvious how that analysis can be undertaken.

6 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 stability of some SMC-based filtering algorithms have also been investigated. Some directions for future work are as follows.

Firstly, in the context of normalizing constants, one direction is the consideration of rare events problems. Following [8], it is possible to obtain computational complexity results for some rare events problems. However, one can pose some rare-events problems in terms of the dimensionality. Our results would, in many cases, not apply to this scenario and an extension to this case is important. In the case where one uses SMC samplers to sample from 'twisted' target densities (see [21]) the analysis adopted here can be applied. However, one would still need to verify that the path-sampling-based estimate will stabilize as d grows.

Secondly, for normalizing constants, we have only considered the relative $\mathbb{L}_2$-error. It would be of interest to consider e.g. logarithmic efficiency or higher-order errors. In addition, we have only considered one particular important functional that grows 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, and rather importantly, is it possible to find any online SMC algorithm to solve the filtering problem in general, whose cost does not increase exponentially in the dimension? At present, our only suggestion is the accept/reject scheme in [11]. We are currently investigating the stability properties of this algorithm. It could be that in general, as noted above, one cannot obtain a stability result as in [3].

Finally, one could considerably weaken the the hypotheses made in this article. Given the number of exponential moments that we need to treat, it seems that multiplicative drift conditions [22] could be adopted; see [31].

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 [3] 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 $X^{1:N}_{l_d(t_{k-1}(d)),1:d} \sim \pi^{\otimes Nd}_{t_{k-1}(d)}$ we have that:*

$$\frac{G^i_{k,j}}{\sqrt{d}} \Rightarrow N(0, \sigma^2_{t_{k-1}:t_k}) \ ; \quad \frac{1}{d} \sum_{j=1}^{d} G^i_{k,j} \Rightarrow N(0, \sigma^2_{t_{k-1}:t_k}) \ .$$

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

$$\mathbb{E}_{\check{X}^i_{l_d(t_{k-1}(d)),j}} |G^i_{k,j}|^p \leq M \, d^{\frac{p}{2} \vee 1} \ .$$

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

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

iv) *We have that:*

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

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

Proof.

i) Both weak limit follows from the proof of Theorem 3.2 of [3]. Notice, that a minor difference is that instead of the *fixed* times ϕ_0 and 1 considered in Theorem 3.2 of [3] 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 [3].

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 [3]. The second result is shown in the proof of Theorem 4.1 of [3]. $\qquad\qquad\square$

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 $d \to \infty$:*

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

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

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:*

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

where $\{Z^i\}_{i=1}^N$ are i.i.d. copies from $N(0, \sigma_{t_{k-1}:s}^2)$ 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 [3]; the second result from Proposition A.1 of [3].

ii) The weak convergence of the weights is analytically illustrated in the proof of Theorem 4.1 of [3]. The weak convergence of the positions of the Markov chain is proven in Proposition A.1 of [3]. 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:

$$\mathbb{E}\left[\left(\frac{1}{N} \sum_{i=1}^N [\varphi(\check{X}_{d,1}^i) - \pi(\varphi)] \right)^2 \right] = \frac{1}{N} \mathbb{E}\left[\{ \overline{\varphi}(\check{X}_{d,1}^1) \}^2 \right] + \left(\frac{N-1}{N} \right) \mathbb{E}\left[\overline{\varphi}(\check{X}_{d,1}^1) \overline{\varphi}(\check{X}_{d,1}^2) \right] \tag{20}$$

where we have set $\overline{\varphi}(x) = \varphi(x) - \pi(\varphi)$. Starting with the first term on the R.H.S. of (20), and averaging over the resampling time, one has

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

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

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

where $\{Z^i\}_{i=1}^N$ are i.i.d. from $N(0, \sigma_{t_{m^*-1}:1}^2)$ 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 (20). Averaging over the resampling index and invoking again the asymptotic independence result of Lemma A.2(ii) we have:

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

$$\pi(\overline{\varphi}^2) \mathbb{E}\left[\sum_{i=1}^N \frac{e^{2Z^i}}{(\sum_{l=1}^N e^{Z^l})^2} \right] + 0 \equiv N \pi(\overline{\varphi}^2) \mathbb{E}\left[\frac{e^{2Z^1}}{(\sum_{l=1}^N e^{Z^l})^2} \right] \tag{21}$$

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 $\sigma^2 \equiv \sigma_{t_{m^*-1}:1}^2$ for notational convenience):

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

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

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

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

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

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:

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

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:

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

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 (21), 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. $\qquad\square$

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:

$$\mathbb{V}_2(\gamma_d(1)) = \mathbb{E}\left[\left(\frac{\overline{\gamma}_d^N(1)}{\overline{\gamma}_d(1)} - 1\right)^2\right]$$

with

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

where we have now set

$$G_j^i = (1-\phi_0)\sum_{n=0}^{d-1}\left(g(x_{n,j}^i) - \mathbb{E}\left[g(X_{n,j}^i)\right]\right) \tag{22}$$

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

$$\mathbb{E}\left[\left(\frac{\overline{\gamma}_d^N(1)}{\overline{\gamma}_d(1)} - 1\right)^2\right] = 1 - \frac{2}{\overline{\gamma}_d(1)}\mathbb{E}\left[\overline{\gamma}_d^N(1)\right] + \frac{1}{\overline{\gamma}_d(1)^2}\mathbb{E}\left[\overline{\gamma}_d^N(1)^2\right]$$

$$\equiv -1 + \frac{1}{\overline{\gamma}_d(1)^2}\mathbb{E}\left[\overline{\gamma}_d^N(1)^2\right] \tag{23}$$

where we have used the unbiasedness property (i.e. $\mathbb{E}\left[\overline{\gamma}_d^N(1)\right] = \overline{\gamma}_d(1)$) of the normalizing constant, see e.g. [11]. We define $Z_d^i = \frac{1}{d}\sum_{j=1}^d G_j^i$ for G_j^i defined in (22) and $1 \leq i \leq N$. Thus, due to Z_d^i's being i.i.d., we have:

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

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

$$\mathbb{E}\left[e^{2Z_d^1}\right] \to \exp\{2\sigma_{\phi_0:1}^2\} ; \quad \mathbb{E}\left[e^{Z_d^1}\right] \to \exp\{\tfrac{1}{2}\sigma_{\phi_0:1}^2\} .$$

Using these limits in (23) and recalling also that $\overline{\gamma}_d(1) \equiv \mathbb{E}\left[e^{Z_d^1}\right]$, gives the required result. $\qquad\square$

Proof of Theorem 3.3. Denote:

$$\overline{\gamma}_{d,k}^N(1) = \frac{1}{N}\sum_{i=1}^N e^{\frac{1}{d}\sum_{j=1}^d G_{k,j}^i} ; \quad \overline{\gamma}_{d,k}(1) = \mathbb{E}_{\pi_{t_{k-1}(d)}^{\otimes Nd}}\left[e^{\frac{1}{d}\sum_{j=1}^d G_{k,j}^1}\right] , \tag{24}$$

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

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

Using the unbiased property of normalising constants, see e.g. [11], we have:

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

For notational convenience, we set:

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

Following the definitions of $\overline{\gamma}_{d,k}^N(1)$ and $\overline{\gamma}_{d,k}(1)$ in (24), and exploiting independence among particles under $\pi_{t_{k-1}(d)}^{\otimes Nd}$, we have that:

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

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

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

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

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

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

Now, note that a simple identity gives that:

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

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

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

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:

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

Recalling the definition of $\Delta_{k,d} = \frac{\overline{\gamma}_{d,k}^N(1)^2}{\overline{\gamma}_{d,k}(1)^2}$ from (24), using triangle inequality for norms we have:

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

Now, we complete via Lemma A.6. $\qquad\square$

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

$$\mathbb{E}\left[e^{\sum_{l=1}^k c_l\,\mathcal{G}_{l,d}/d} \right] = \mathbb{E}\left[\mathbb{E}\left[e^{c_k\,\mathcal{G}_{k,d}/d} \mid \mathscr{F}_{t_{k-1}(d)}^N \right] \left\{ e^{\sum_{l=1}^{k-1} c_l\,\mathcal{G}_{l,d}/d} - \mathbb{E}\left[e^{\sum_{l=1}^{k-1} c_l\,\mathcal{G}_{l,d}/d} \right] \right\} \right]$$
$$+ \mathbb{E}\left[e^{\sum_{l=1}^{k-1} c_l\,\mathcal{G}_{l,d}/d} \right] \mathbb{E}\left[e^{c_k\,\mathcal{G}_{k,d}/d} \right] \ . \tag{25}$$

We begin by dealing with the first term on the R.H.S. of (25). By Lemma A.4 we have that:

$$\mathbb{E}\left[e^{c_k\,\mathcal{G}_k/d} \mid \mathscr{F}_{t_{k-1}(d)}^N \right] - \mathbb{E}_{\pi_{t_{k-1}(d)}^{\otimes d}}\left[e^{c_k\mathcal{G}_k/d} \right] \xrightarrow{\mathbb{P}} 0 \ ,$$

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

$$\mathbb{E}_{\pi_{t_{k-1}(d)}^{\otimes d}}\left[e^{c_k\mathcal{G}_k/d} \right] \to e^{\frac{1}{2} c_k^2\,\sigma_{t_{k-1}:t_k}^2} \ . \tag{26}$$

Moreover, by the induction hypothesis:

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

The expression in the expectation of the first term of (25) is uniformly integrable: indeed, careful and repeated (but otherwise straightforward) use of Hölder and Jensen inequalities will eventually give that:

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

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

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

Now turning to the second term on the R.H.S. of (25), we work as follows:

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

from Lemma A.4 and (26). We can thus deduce by the induction hypothesis that

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

which completes the proof. $\qquad\square$

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

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

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

$$
\mathbb{E}\left[\overline{\gamma}^N_{d,k}(1)^2 \,\middle|\, \mathscr{F}^N_{t_{k-1}(d)} \right] = \tag{27}
$$
$$
= \frac{1}{N^2} \left(\mathbb{E}\left[\sum_{i=1}^{N} e^{\frac{2}{d} \sum_{j=1}^{d} G^i_{k,j}} \,\middle|\, \mathscr{F}^N_{t_{k-1}(d)} \right] + \sum_{i \neq m} \mathbb{E}\left[e^{\frac{1}{d} \sum_{j=1}^{d} G^i_{k,j}} \,\middle|\, \mathscr{F}^N_{t_{k-1}(d)} \right] \mathbb{E}\left[e^{\frac{1}{d} \sum_{j=1}^{d} G^m_{k,j}} \,\middle|\, \mathscr{F}^N_{t_{k-1}(d)} \right] \right) \, .
$$

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 $d \to \infty$:

$$
\mathbb{E}\left[e^{\frac{c}{d} \sum_{j=1}^{d} G^i_{k,j}} \,\middle|\, \mathscr{F}^N_{t_{k-1}(d)} \right] - \mathbb{E}_{\pi^{\otimes Nd}_{t_{k-1}(d)}}\left[e^{\frac{c}{d} \sum_{j=1}^{d} G^i_{k,j}} \right] \to 0 \, , \quad in \ \ \mathbb{L}_{2(1+\epsilon)} \, . \tag{28}
$$

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 (27). 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, we have that for any $N \geq 1$, $i \in \{1, \ldots, N\}$, $k \in \{1, \ldots, m^* + 1\}$ and $c \in \mathbb{R}$:*

$$\mathbb{E}\left[e^{\frac{c}{d} \sum_{j=1}^{d} G_{k,j}^i} \mid \mathscr{F}_{t_{k-1}(d)}^N \right] - \mathbb{E}_{\pi_{t_{k-1}(d)}^{\otimes Nd}}\left[e^{\frac{c}{d} \sum_{j=1}^{d} G_{k,j}^i} \right] \xrightarrow{\mathbb{P}} 0 .$$

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

$$\mathbb{E}\left[e^{\frac{c}{d} \sum_{j=1}^{d} G_{k,j}^i} \mid \mathscr{F}_{t_{k-1}(d)}^N \right] = \prod_{j=1}^{d} \mathbb{E}_{\check{X}_{l_d(t_{k-1}(d)),j}}\left[e^{\frac{c}{d} G_{k,j}^i} \right] .$$

We will now omit various subscripts/superscripts 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:

$$\mathbb{E}\left[e^{\frac{c}{d} \sum_{j=1}^{d} G_j} \mid \mathscr{F}^N \right] - \mathbb{E}_{\pi^{\otimes Nd}}\left[e^{\frac{c}{d} \sum_{j=1}^{d} G_j} \right] =$$
$$= \left(\prod_{j=1}^{d} \mathbb{E}_\pi\left[e^{c\, G_j/d} \right] \right) \left[\prod_{j=1}^{d} \left\{ \frac{\{\mathbb{E}_{\check{X}_{0,j}} - \mathbb{E}_\pi\}[e^{c\, G_j/d}]}{\mathbb{E}_\pi[e^{c\, G_j/d}]} + 1 \right\} - 1 \right] . \tag{29}$$

From Lemma A.1(iii) it follows that $\prod_{j=1}^{d} \mathbb{E}_\pi\left[e^{c\, G_j/d} \right] \to e^{\frac{1}{2} c^2 \sigma_{t_{k-1}:t_k}^2}$, hence we can now concentrate on the second factor-term on the R.H.S. of (29). We will replace the product with a sum using logarithms. To that end define:

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

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:

$$-1 + \epsilon \leq \beta_j(d) \leq M < \infty . \tag{30}$$

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:

$$\sum_{j=1}^{d} \log(1 + \beta_j(d)) = \sum_{j=1}^{d} \left\{ \beta_j(d) - \tfrac{1}{2} \tfrac{1}{(1 + \xi_j(d))^2} \beta_j^2(d) \right\} \tag{31}$$

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 (30), these two results imply via the Taylor expansion in (31) 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. $\square$

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 $c \in \mathbb{R}$:*

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

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

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

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

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

$$e^{c\,G_j/d} = 1 + \tfrac{c\,G_j}{d} + \tfrac{1}{2} e^{\xi_j(d)} \left(\tfrac{c\,G_j}{d} \right)^2 , \tag{32}$$

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

Proof of (i):

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

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

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 $\sigma^2 \equiv \sigma^2_{t_{n-1}:t_n}$):

$$\mathbb{E}\left| \sum_{j=1}^{d} \{\mathbb{E}_{\check{X}_{0,j}} - \mathbb{E}_\pi\}\left[e^{\xi_j(d)}\left(\tfrac{G_j}{d}\right)^2\right] \right| \leq \tag{33}$$

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

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) ;

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

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 (33) goes to zero. For the first term on the R.H.S. of (33) 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:

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

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 (33) we work as follows. We have that:

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

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 (32). Clearly, the $\mathbb{L}_1$-norm of the random variable of interest is bounded by:

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

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 $c \in \mathbb{R}$:*

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

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

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

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

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

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:

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

Hence, we have that:

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

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 $X^{1:q}_{s(d),1} = (X^1_{s(d),1}, \ldots, X^q_{s(d),1})$:

$$
\begin{aligned}
|\,\mathbb{E}\,[\,F(X^{1:q}_{s(d),1})\,] - \pi^{\otimes q}_s(F)\,| \leq{} & |\,\mathbb{E}\,[\,F(X^{1:q}_{s(d),1})\,] - \mathbb{E}_{\pi^{\otimes N}_{t_{k-1}(d)}}[\,F(X^{1:q}_{s(d),1})\,]\,| \\
& + |\,\mathbb{E}_{\pi^{\otimes N}_{t_{k-1}(d)}}[F(X^{1:q}_{s(d),1})] - \pi^{\otimes q}_{s(d)}(F)\,| + |\,\pi^{\otimes q}_{s(d)}(F) - \pi^{\otimes q}_s(F)\,|\,.
\end{aligned} \tag{34}
$$

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 (34) one can use conditional expectations and write it as:

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

where $\mathscr{F}^N_{t_{k-1}(d)}$ 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:

$$
\left\{k^{\otimes q}_{u(d)}(\check{X}^{1:q}_{l_d(t_{k-1}(d)),1}, \cdot\,) - \left(\pi_{t_{k-1}(d)}k_{u(d)}\right)^{\otimes q}\right\}(F) \tag{35}
$$

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:

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

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

$$
\begin{aligned}
&\left\{k^{\otimes q}_{u(d)}(\check{X}^{1:q}_{l_d(t_{k-1}(d)),1}, \cdot\,) - \left(\pi_{t_{k-1}(d)}k_{u(d)}\right)^{\otimes q}\right\}(dy_{1:q}) \\
&= \sum_{l=1}^q \left(k_{u(d)}(\check{X}^l_{l_d(t_{k-1}(d)),1}, \cdot\,) - \pi_{t_{k-1}(d)}k_{u(d)}\right)(dy_l) \otimes \mu_l(dy_{1:(l-1)}, dy_{(l+1):q})\,.
\end{aligned} \tag{36}
$$

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 (36) we have that the expression in (35) is bounded in absolute value by:

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

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 (34) goes to zero as $d \to \infty$. The second term on the R.H.S. of (34) can be treated in a similar manner. One has again the identity:

$$
\begin{aligned}
\mathbb{E}_{\pi^{\otimes N}_{t_{k-1}(d)}}\,[\,F(X^{1:q}_{s(d),1})\,] - \pi^{\otimes q}_{s(d)}(F) ={}& \\
= \sum_{l=1}^q \int_{\mathbb{R}} \{\,\pi_{t_{k-1}(d)}k_{u(d)} - \pi_{s(d)}\,\}&(dx_l) \times \left\{\pi^{\otimes(l-1)}_{s(d)} \otimes \left(\pi_{t_{k-1}(d)}k_{u(d)}\right)^{\otimes(q-l)}(F(x_l))\right\} \\
\leq q\,\|\pi_{t_{k-1}(d)}k_{u(d)} - \pi_{s(d)}&\|_{tv}\,.
\end{aligned}
$$

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

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] 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.

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

[4] 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.

[5] BRETT, C. E. A., LAM, K. F., LAW, K. J. H., MCCORMICK, D. S., SCOTT, M. R. & STUART, A. M. (2011). Stability of filters for the Navier Stokes equation. Technical Report, University of Warwick.

[6] BUI-QUANG, P., MUSSO, C. & LE GLAND, F. (2010). An insight into the issue of dimensionality in particle filtering, *Proc. 13th Internl. Conf. Information Fusion*.

[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] CHORIN, A. & TU, X. (2009). Implicit sampling from particle filters. *PNAS*, **106**, 17249–17254.

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

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

[13] DEL MORAL, P., DOUCET, A. & JASRA, A. (2012). On adaptive resampling procedures for sequential Monte Carlo methods. *Bernoulli*, (to appear).

[14] DEL MORAL, P., DOUCET, A. & JASRA, A. (2012). An adaptive sequential Monte Carlo method for approximate Bayesian computation. *Statist. Comp.*, (to appear).

[15] 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.

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

[17] EVENSEN, G. (2009). *Data Assimilation*, Springer: New York.

[18] GODSILL S., & CLAPP, T. (2001). Improvement strategies for Monte Carlo particle filters. In *Sequential Monte Carlo Methods in Practice*, 139–158. Springer: New York.

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

[20] JASRA, A., SINGH, S. S., MARTIN, J. S. & MCCOY, E. (2012). Filtering via approximate Bayesian computation. *Statist. Comp.*, (to appear).

[21] JOHANSEN, A. M., DEL MORAL, P., DOUCET, A. (2006). Sequential Monte Carlo samplers for rare events. *Proc. 6th Internl. Workshop on Rare Event Simulation*, 256-267.

[22] KONTOYIANNIS, I. & MEYN, S. P. (2005). Large deviation asymptotics and the spectral theory of multiplicatively regular Markov processes. *Elec. J. Probab.*, **10**, 61123.

[23] LE GLAND, F., MONBET, V. & TRAN, V. D. (2011). Large sample asymptotics for the ensemble Kalman filter. In *The Oxford Handbook of Nonlinear Filtering*, (Crisan, D. & Rozovskii, B. Eds), Oxford: OUP.

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

[25] NOTT, D., MARSHALL, L. & NGOC, T. M. (2012). The ensemble Kalman filter is an ABC algorithm. *Statist. Comp.*, (to appear).

[26] PILLAI, N., STUART, A. M. & THIÉRY, A. H. (2011). On the random walk Metropolis algorithm for Gaussian random field priors and the gradient flow. *Arxiv* preprint.

[27] POYIADJIS, G., DOUCET, A. & SINGH, S. S. (2011). Particle approximations of the score and observed information matrix in state-space models with application to parameter estimation. *Biometrika*, **98**, 65–80.

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

[29] VAN HANDEL, R. (2009). Uniform time average consistency of Monte Carlo particle filters. *Stoch. Proc. Appl.*, **119**, 3835–3861.

[30] WHITELEY, N. (2011). Sequential Monte Carlo samplers: Error bounds and insensitivity to initial conditions. Technical Report, University of Bristol.

[31] WHITELEY, N., KANTAS, N, & JASRA, A. (2011). Linear variance bounds for particle approximations of time homogeneous Feynman-Kac formulae. Technical Report, University of Bristol.