On the Stability of Sequential Monte Carlo Methods in High Dimensions

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

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

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

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

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

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

E-Mail: staja@nus.edu.sg

Abstract

We investigate the stability of a Sequential Monte Carlo (SMC) method applied to the problem of sampling from a target distribution on $\mathbb{R}^d$ for large d. It is well known [7, 13, 48] that using a *single* importance sampling step, one produces an approximation for the target that deteriorates as the dimension d increases, unless the number of Monte Carlo samples N increases at an exponential rate in d. We show that this degeneracy can be avoided by introducing a sequence of artificial targets, starting from a 'simple' density and moving to the one of interest, using an SMC method to sample from the sequence (see e.g. [18, 22, 31, 41]). Using this class of SMC methods with a fixed number of samples, one can produce an approximation for which the effective sample size (ESS) converges to a random variable ε_N as $d \to \infty$ with $1 < \varepsilon_N < N$. The convergence is achieved with a computational cost proportional to Nd^2. If $\varepsilon_N \ll N$, we can raise its value by introducing a number of resampling steps, say m (where m is independent of d). In this case, ESS converges to a random variable $\varepsilon_{N,m}$ as $d \to \infty$ and $\lim_{m\to\infty} \varepsilon_{N,m} = N$. Also, we show that the Monte Carlo error for estimating a fixed dimensional marginal expectation is of order $\frac{1}{\sqrt{N}}$ *uniformly* in d. The results imply that, in high dimensions, SMC algorithms can efficiently control the variability of the importance sampling weights and estimate fixed dimensional marginals at a cost which is less than exponential in d and indicate that resampling leads to a reduction in the Monte Carlo error and increase in the ESS. All of our analysis is made under the assumption that the target density is i.i.d..

Key words: Sequential Monte Carlo, High Dimensions, Resampling, Functional CLT.

AMS 2000 Subject Classification: Primary 82C80; Secondary 60F99, 62F15

1 Introduction

Sequential Monte Carlo (SMC) methods can be described as a collection of techniques that approximate a sequence of distributions, known up-to a normalizing constant, of increasing dimension. Typically, the complexity of these distributions is such that one cannot rely upon standard simulation approaches. SMC methods are applied in a wide variety of applications, including engineering, economics and biology, see [28] and Chapter VIII in [20] for an overview. They combine importance sampling and resampling to approximate distributions. The idea is to introduce a sequence of proposal densities and sequentially simulate a collection of $N > 1$ samples, termed particles, in parallel from these proposals. In most scenarios it is not possible to use the distribution of interest as a proposal. Therefore, one must correct for the discrepancy between proposal and target via importance weights. In almost all cases of practical interest, the variance of these importance weights increases with algorithmic time (e.g. [34]); this can, to some extent, be dealt with via resampling. This consists of sampling with replacement from the current samples using the weights and resetting them to $1/N$. The variability of the weights is often measured by the effective sample size ([37]) and one often resamples when this drops below a threshold (dynamic-resampling).

There are a wide variety of convergence results for SMC methods, most of them concerned with the accuracy of the particle approximation of the distribution of interest as a function of N. A less familiar context, related with this paper, arises in the case when the difference in the dimension of the consecutive densities becomes large. Whilst in filtering there are several studies on the stability of SMC as the time step grows (see e.g. [19, 21, 25, 26, 30, 35]) they do not consider this latter scenario. In addition, there is a vast literature on the performance of high-dimensional Markov chain Monte Carlo (MCMC) algorithms e.g. [10, 43, 44]; our aim is to obtain a similar analytical understanding about the effect of dimension on SMC methods. The articles [4, 7, 13, 48] have considered some problems in this direction. In [7, 13, 48] the authors show that, for an i.i.d. target, as the dimension of the state grows to infinity then one requires, for some stability properties, a number of particles which grows exponentially in dimension (or 'effective dimension' in [48]); the algorithm considered is standard importance sampling. We discuss these results below.

1.1 Contribution of the Article

We investigate the stability of an SMC algorithm in high dimensions used to produce a sample from a sequence of probabilities on a common state-space. This problem arises in a wide variety of applications including many encountered in Bayesian statistics. For some Bayesian problems the posterior density can be very 'complex', that is, multi-modal and/or with high correlations between certain variables in the target ('static' inference, see e.g. [33]). A commonly used idea is to introduce a simple distribution, which is more straightforward to sample from, and to interpolate between this distribution and the actual posterior by introducing intermediate distributions from which one samples sequentially. Whilst this problem departs from the standard ones in the SMC literature, it is possible to construct SMC methods to approximate this sequence; see [18, 22, 31, 41]. The methodology investigated here is applied in many practical contexts: financial modelling [32], regression [46] and approximate Bayesian inference [23]. In addition, high-dimensional problems are of practical importance and normally more challenging than their low dimensional counterparts. The question we look at is whether such algorithms, as the dimension d of the distributions increases, are stable in any sense. That is, whilst d is *fixed* in practice, we would like identify the computational cost of the algorithm for large d, to ensure that the algorithm is stable. Within the SMC context described here, we quote the following statement made in [13]:

> 'Unfortunately, for truly high dimensional systems, we conjecture that the number of intermediate steps would be prohibitively large and render it practically infeasible.'

One of the objectives of this article is to investigate the above statement from a theoretical perspective. In the sequel we show that for i.i.d. target densities:

- The SMC algorithm analyzed, with computational cost $\mathcal{O}(Nd^2)$ is stable. Analytically, we prove that ESS converges weakly to a non-trivial random variable ε_N as d grows and the number of particles *is kept fixed*. In addition, we show that the Monte Carlo error of the estimation of fixed dimensional marginals, for a fixed number of particles N is of order $1/\sqrt{N}$ *uniformly* in d. The algorithm can *include* dynamic resampling at some particular deterministic times. In this case, the algorithm will resample $\mathcal{O}(1)$ times. Our results indicate that estimates will improve when one resamples.

- The dynamically resampling SMC algorithm (with stochastic times and some minor modifications) will, with probability greater than or equal to $1-M/\sqrt{N}$, where M is a positive constant independent of N, also exhibit these properties.

- Our results are proved for $\mathcal{O}(d)$ steps in the algorithm. If one takes $\mathcal{O}(d^{1+\delta})$ steps with any $\delta > 0$, then ESS converges in probability to N and the Monte Carlo error is the same as with i.i.d. sampling. If $-1 < \delta < 0$ then ESS will go-to zero (Corollary 5.1). That is, $\mathcal{O}(d)$ steps are a critical order for the stability of the algorithm in our scenario.

Going into more details, we informally summarise below our main results (throughout the costs of the SMC algorithms are $\mathcal{O}(Nd^2)$ for i.i.d. target densities):

- Theorem 3.1 shows that, for the algorithm that does not resample, the evolution of the log-weight of a particle stabilizes via convergence to a time-changed Brownian motion, as $d \to \infty$ with N fixed.

- Theorem 3.2 shows, in the context of no resampling, that the ESS converges to a non-trivial random variable as d grows with N fixed.

- Theorem 3.3 shows, in the context of no resampling, that the Monte Carlo error associated to expectations w.r.t. a fixed dimensional marginal of the target is $\mathcal{O}(N^{-1/2})$ as $d \to \infty$.

- Theorem 4.1 shows, for the algorithm which resamples at some deterministic times (which are not available in practice), that the ESS converges to a non-trivial random variable as d grows with N fixed.

- Theorem 4.2 shows, for the SMC algorithm which resamples at deterministic times, that the Monte Carlo error associated to expectations w.r.t. a fixed dimensional marginal of the target is $\mathcal{O}(N^{-1/2})$ as $d \to \infty$.

- Theorem 4.3 which shows that as $d \to \infty$, a practical algorithm which resamples at stochastic times (and under a modification) inherits the same properties as the algorithm which resamples at deterministic times, with a probability that is lower-bounded by $1 - M/\sqrt{N}$, for some $M < \infty$.

Our results show that in high-dimensional problems, one is able to control the variability of the weights; this is a minimum requirement for applying the algorithm. They also establish that one can estimate *fixed* dimensional marginals even as the dimension d increases. The results help to answer the point of [13] quoted above. In the presence of a quadratic cost and increasingly sophisticated hardware (e.g. [36]) SMC methods are applicable, in the static context, in high-dimensions. To support this, [32] presents further empirical evidence of the results presented here. In particular, it is shown there that SMC techniques are algorithmically stable for models of dimension over 1000 with computer simulations that run in just over 1 hour. Hence the SMC techniques analyzed here can certainly be used for high-dimensional *static* problems. The analysis of such methods for time-dependent applications (e.g. filtering) is subject to further research.

When there is no resampling, the proofs of our results rely on martingale array techniques. To show that the algorithm is stable we establish a functional central limit theorem (fCLT) under easily verifiable conditions, for a triangular array of non-homogeneous Markov chains. This allows one to establish the convergence in distribution of ESS (as d increases). The result also demonstrates the dependence of the algorithm on a mixture of asymptotic variances (in the Markov chain CLT) of the non-homogeneous kernels.

1.2 Structure of the Article

In Section 2 we discuss the SMC algorithm of interest and the class of target distributions we consider. In Section 3 we show that ESS converges in distribution to a non-trivial random variable as $d \to \infty$ when the algorithm does not resample. We also show that the Monte Carlo error of the estimation of fixed dimensional marginals, for a fixed number of particles N, has an upper bound of the form $M/\sqrt{N}$, where M is independent of d. We address the issue of resampling in Section 4, where it is shown that as $d \to \infty$ any dynamically resampling SMC algorithm, using the deterministic ESS (the expected ESS with one particle) will resample $\mathcal{O}(1)$ times and also exhibit convergence of the ESS and Monte Carlo error. In addition, any dynamically resampling SMC algorithm, using the empirical ESS (with some modification) will, with high probability, display the same convergence of the ESS and Monte Carlo error. Finally, we conclude in Section 5 with some remarks on $\mathcal{O}(d)$ steps being a critical order and ideas for future work. Proofs are collected in the Appendix.

1.3 Notation

Let $(E, \mathscr{E})$ be a measurable space and $\mathscr{P}(E)$ be the set of probability measures on $(E, \mathscr{E})$. For a given function $V : E \mapsto [1, \infty)$ we denote by $\mathscr{L}_V$ the class of functions $f : E \mapsto \mathbb{R}$ for which

$$|f|_V := \sup_{x \in E} \frac{|f(x)|}{V(x)} < +\infty \ .$$

For two Markov kernels, P and Q on $(E, \mathscr{E})$, we define the V-norm:

$$|||P - Q|||_V := \sup_{x \in E} \frac{\sup_{|f| \leq V} |P(f)(x) - Q(f)(x)|}{V(x)} \ ,$$

with $P(f)(x) := \int_E P(x, dy) f(y)$. The notation

$$\|P(x, \cdot) - Q(x, \cdot)\|_V := \sup_{|f| \leq V} |P(f)(x) - Q(f)(x)|$$

is also used. For $\mu \in \mathscr{P}(E)$ and P a Markov kernel on $(E, \mathscr{E})$, we adopt the notation $\mu P(f) := \int_E \mu(dx) P(f)(x)$. In addition, $P^n(f)(x) := \int_{E^{n-1}} P(x, dx_1) P(x_1, dx_2) \times \cdots \times P(f)(x_{n-1})$. $\mathscr{B}(\mathbb{R})$ is used to denote the class of Borel sets and $\mathcal{C}_b(\mathbb{R})$ the class of bounded continuous $\mathscr{B}(\mathbb{R})$-measurable functions. Denote $\|f\|_\infty = \sup_{x \in \mathbb{R}} |f(x)|$. We will also define the $\mathbb{L}_\varrho$-norm, $\|X\|_\varrho = \mathbb{E}^{1/\varrho} |X|^\varrho$, for $\varrho \geq 1$ and denote by $\mathbb{L}_\varrho$ the space of random variables such that $\|X\|_\varrho < \infty$. For $d \geq 1$, $\mathcal{N}_d(\mu, \Sigma)$ denotes the d-dimensional normal distribution with mean μ and covariance Σ; when $d = 1$ the subscript is dropped. For any vector $(x_1, \ldots, x_p)$, we denote by $x_{q:s}$ the vector $(x_q, \ldots, x_s)$ for any $1 \leq q \leq s \leq p$. Throughout M is used to denote a constant whose meaning may change, depending upon the context; any (important) dependencies are written as $M(\cdot)$.

2 Sequential Monte Carlo

We wish to sample from a target distribution with density Π on $\mathbb{R}^d$ with respect to Lebesgue measure, known up to a normalizing constant. We introduce a sequence of 'bridging' densities which start from an easy to sample target and evolve toward Π. In particular, we will consider (e.g. [22]):

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

for $0 < \phi_0 < \cdots < \phi_{n-1} < \phi_n < \cdots < \phi_p = 1$. The effect of exponentiating with the small constant ϕ_0 is that $\Pi(x)^{\phi_0}$ is much 'flatter' than Π. Other choices of bridging densities are possible and are discussed in the sequel.

One can sample from the sequence of densities using an SMC sampler, which is, essentially, a Sequential Importance Resampling (SIR) algorithm or particle filter that can be designed to target the sequence of densities:

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

with domain $(\mathbb{R}^d)^n$ of dimension that increases with $n = 1, \ldots, p$, and $\{K_n\}$ a sequence of Markov kernels of invariant density $\{\Pi_n\}$. Assuming the weights appearing below are well-defined Radon Nikodym derivatives, the SMC algorithm we will ultimately explore is the one defined in Figure 1. With no resampling, the algorithm coincides with the annealed importance sampling in [41]. It is remarked that, due to the results of [7, 13, 48], it appears that

---

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

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

Calculate the Effective Sample Size (ESS):

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

If $ESS_{l:n}(N) < a$:

resample particles according to their normalised weights

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

set $l = n$ and re-initialise the weights by setting $w_{l:n}^i \equiv 1$, $1 \leq i \leq N$;

let $\check{x}_n^1, \ldots, \check{x}_n^N$ now denote the resampled particles and set $(x_n^1, \ldots, x_n^N) = (\check{x}_n^1, \ldots, \check{x}_n^N)$

Set $n = n + 1$.

Return to the start of Step 1.

Figure 1: The SMC algorithm.

the cost of the population Monte Carlo method of [17] would increase exponentially with the dimension; instead we will show that the 'bridging' SMC sampler framework above will be of smaller cost.

The ESS defined in (2) is typically used to quantify the quality of SMC approximations associated to systems of weighted particles. It is a number between 1 and N, and in general the larger the value, the better the approximation. Resampling is often performed when ESS falls below some pre-specified threshold such as $a = N/2$. The operation of resampling consists of sampling with replacement from the current set of particles via the normalized weights in (3) and resetting the (unnormalized) weights to 1. There is a wide variety of resampling techniques and we refer the reader to [28] for details; in this article we only consider the multinomial method just described above.

2.1 Framework

We will investigate the stability of the SMC algorithm in Figure 1 as $d \to \infty$. To obtain analytical results we will simplify the structure of the algorithm (similar to the results for MCMC in high dimensions in e.g. [5, 10, 43, 44]).

In particular, we will consider an i.i.d. target:

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

for some $g : \mathbb{R} \mapsto \mathbb{R}$. 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 (clearly, in an i.i.d. context one could use standard sampling schemes). Still, such a context allows for a rigorous mathematical treatment; at the same time (and similarly to corresponding extensions of results for MCMC algorithms in high dimensions) one would expect that the analysis we develop in this paper for i.i.d. targets will also be relevant in practice for more general scenarios; see Section 5 for some discussion. A further assumption that will facilitate the mathematical analysis is to apply independent kernels along the different co-ordinates. That is, we will assume:

$$K_n(x, dx') = \prod_{j=1}^{d} k_n(x_j, dx'_j) \ , \tag{5}$$

where each transition kernel $k_n(\cdot, \cdot)$ preserves $\pi_n(x)$; that is, $\pi_n k_n = \pi_n$. Clearly, this also implies that $\Pi_n K_n = \Pi_n$.

The stability of ESS will be investigated as $d \to \infty$: first without resampling and then with resampling. We study the case when one selects cooling constants ϕ_n and p as below:

$$p = d \ ; \quad \phi_n (= \phi_{n,d}) = \phi_0 + \frac{n(1-\phi_0)}{d} \ , \quad 0 \le n \le d \ , \tag{6}$$

with $0 < \phi_0 < 1$ given and fixed with respect to d. It will be shown that such a selection will indeed provide a 'stable' SMC algorithm as $d \to \infty$. Note that $\phi_0 > 0$ as we will be concerned with probability densities on non-compact spaces.

Remark 2.1. *Since $\{\phi_n\}$ will change with d, all elements of our SMC algorithm will also depend on d. We use the double-subscripted notation $k_{n,d}$, $\pi_{n,d}$ when needed to emphasize the dependence of k_n and π_n on d, which ultimately, depend on n, d through $\phi_{n,d}$. Similarly, we will sometimes write $X_n(d)$, or $x_n(d)$, for the Markov chain involved in the specification of the SMC algorithm.*

Remark 2.2. *Although the algorithm runs in discrete time, it will be convenient for the presentation of our results that we consider the successive steps of the algorithm as placed on the continuous time interval $[\phi_0, 1]$, incremented by the annealing discrepancy $(1-\phi_0)/d$. We will use the mapping*

$$l_d(t) = \lfloor d\big(\tfrac{t-\phi_0}{1-\phi_0}\big) \rfloor \tag{7}$$

to switch between continuous time and discrete time. Related to the above, 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:

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

That is, we will use the convention $\pi_n \equiv \pi_{\phi_n}$ with the subscript on the left running on the set $\{1, 2, \ldots, d\}$. Accordingly, $k_s(\cdot, \cdot)$, with $s \in (\phi_0, 1]$, will denote the transition kernel preserving π_s.

2.2 Conditions

We state the conditions under which we derive our results. Throughout, we set $k_{\phi_0} \equiv \pi_{\phi_0}$ and $(E, \mathscr{E}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$. We assume that $g(\cdot)$ is an upper bounded function. In addition, we make the following assumptions for the continuum of kernels/densities:

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

1. *(One-Step Minorization).* We assume that there exists a set $C \in \mathscr{E}$, a constant $\theta \in (0, 1)$ and some $\nu \in \mathscr{P}(E)$ such that for each $s \in (\phi_0, 1]$ the set C is $(1, \theta, \nu)$−small with respect to k_s.

2. (*One-step Drift Condition*). There exists $V : E \mapsto [1, \infty)$ with $\lim_{|x|\to\infty} V(x) = \infty$, constants $\lambda < 1$, $b < \infty$, and $C \in \mathscr{E}$ as specified in (i) such that for any $x \in E$ and $s \in (\phi_0, 1]$:

$$k_s\, V(x) \leq \lambda\, V(x) + b\,\mathbb{I}_C(x)\,.$$

   In addition $\pi_{\phi_0}(V) < \infty$.

3. (*Level Sets*). Define $C_c = \{x : V(x) \leq c\}$ with V as in (2). Then there exists a $c \in (1, \infty)$ such that for every $s \in (\phi_0, 1)$, C_c is a $(1, \theta, \nu)-$small set with respect to k_s. In addition, condition (ii) holds for $C = C_c$, and λ, b (possibly depending on c) such that $\lambda + b/(1 + c) < 1$.

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

There exists an $M < \infty$ such that for any $s, t \in (\phi_0, 1]$

$$|||k_s - k_t|||_V \leq M\,|s - t|\;.$$

The statement that C is $(1, \theta, \nu)-$small w.r.t. to k_s means that C is an one-step small set for the Markov kernel, with minorizing distribution ν and parameter $\theta \in (0, 1)$ (see e.g. [40]).

Assumptions like (A1) are fairly standard in the literature on adaptive MCMC (e.g. [1]). Note though that the context in this paper is different. For adaptive MCMC one typically has that the kernels will eventually converge to some limiting kernel. Conversely, in our set-up, the d bridges (resp. kernels) in between π_0 (resp. k_0) and π_d (resp. k_d) will effectively make up a continuum of densities π_s (resp. kernels k_s), with $s \in [\phi_0, 1]$, as d grows to infinity. The second assumption above differs from standard adaptive MCMC but will be verifiable in real contexts (see [8]). Note that one could maybe relax our assumptions to, e.g. sub-geometric drift conditions versus those for geometric ergodicity, at the cost of an increased level of complexity in the proofs. It is also remarked that the assumption that g is upper bounded is only used in Section 4, when controlling the resampling times. The assumptions adopted in this article are certainly not weak, but still are very close to the weakest assumptions adopted in state-of-the-art research on stability of SMC, see [49, 50, 51].

3 The Algorithm Without Resampling

We will now consider the case when we omit the resampling steps in the SMC algorithm in Figure 1. Critically, due to the i.i.d. structure of the bridging densities Π_n and the kernels K_n each particle will evolve according to a d-dimensional Markov chain X_n made up of d i.i.d. one-dimensional Markov chains $\{X_{n,j}\}_{n=0}^d$, with j the co-ordinate index, evolving under the kernel k_n. Also, all particles move independently.

We first consider the stability of the terminal ESS, i.e.,

$$\mathrm{ESS}_{(0,d)}(N) = \frac{\left(\sum_{i=1}^N w_d(x_{0:d-1}^i)\right)^2}{\sum_{i=1}^N w_d(x_{0:d-1}^i)^2} \tag{8}$$

where, due to the i.i.d. structure and our selection of ϕ_n's in (6), we can rewrite:

$$w_d(x_{0:d-1}) = \exp\left\{\frac{(1-\phi_0)}{d}\sum_{j=1}^d\sum_{n=1}^d g(x_{n-1,j})\right\}\,. \tag{9}$$

It will be shown that under our set-up $\mathrm{ESS}_{(0,d)}(N)$ converges in distribution to a non-trivial variable and analytically characterize the limit; in particular we will have $\lim_{d\to\infty}\mathbb{E}\,[\,\mathrm{ESS}_{(0,d)}(N)\,] \in (1, N)$.

3.1 Strategy of the Proof

To demonstrate that the selection of the cooling sequence ϕ_n in (6) will control the ESS we look at the behaviour of the sum:

$$\frac{1-\phi_0}{d}\sum_{j=1}^d\sum_{n=1}^d g(x_{n-1,j}) \tag{10}$$

appearing in the expression for the weights, $w_d(x_{0:d-1})$, in (9). Due to the nature of the expression for ESS one can re-center, so we can consider the limiting properties of:

$$\alpha(d) = \frac{1}{\sqrt{d}} \sum_{j=1}^{d} \overline{W}_j(d) \tag{11}$$

differing from (10) only in terms of a constant (the same for all particles), where we have defined:

$$\overline{W}_j(d) = W_j(d) - \mathbb{E}\,[\,W_j(d)\,] \tag{12}$$

and

$$W_j(d) = \frac{1-\phi_0}{\sqrt{d}} \sum_{n=1}^{d} \big\{ g(x_{n-1,j}) - \pi_{n-1}(g) \big\} \ . \tag{13}$$

As mentioned above, the dynamics of the involved random variables correspond to those of d independent scalar non-homogeneous Markov chains $\{X_{n,j}\}_{n=0}^{d} \equiv \{X_{n,j}(d)\}_{n=0}^{d}$ of initial position $X_{0,j} \sim \pi_0$ and evolution according to the transition kernels $\{k_n\}_{1\le n\le d}$. We will proceed as follows. For any fixed d and co-ordinate j, $\{X_{n,j}\}_{n=0}^{d}$ is a non-homogeneous Markov chain of total length $d+1$. Hence, for fixed j, $\{X_{n,j}\}_{d,n}$ constitutes an array of non-homogeneous Markov chains. We will thus be using the relevant theory to prove a central limit theorem (via a fCLT) for $\overline{W}_j(d)$ as $d \to \infty$. Then, the independency of the $\overline{W}_j(d)$'s over j will essentially provide a central limit theorem for $\alpha(d)$ as $d \to \infty$.

3.2 Results and Remarks for ESS

Let $t \in [\phi_0, 1]$ and recall the definition of $l_d(t)$ in (7). We define:

$$S_t = \frac{1-\phi_0}{\sqrt{d}} \sum_{n=1}^{l_d(t)} \{ g(X_{n-1,j}) - \pi_{n-1}(g) \} \ .$$

Note that $S_1 \equiv W_j(d)$. Our fCLT considers the continuous linear interpolation:

$$s_d(t) = S_t + \Big(d\,\frac{t-\phi_0}{1-\phi_0} - l_d(t) \Big) [S_{t^+} - S_t] \ ,$$

where we have denoted

$$S_{t^+} = \frac{1-\phi_0}{\sqrt{d}} \sum_{n=1}^{l_d(t)+1} \{ g(X_{n-1,j}) - \pi_{n-1}(g) \} \ .$$

Theorem 3.1 (fCLT)**.** *Assume (A1(i)(ii), A2) and that $g \in \mathscr{L}_{V^r}$ for some $r \in [0, \frac{1}{2})$. Then:*

$$s_d(t) \Rightarrow \mathcal{W}_{\sigma^2_{\phi_0:t}} \ ,$$

where $\{\mathcal{W}_t\}$ is a Brownian motion and

$$\sigma^2_{\phi_0:t} = (1-\phi_0) \int_{\phi_0}^{t} \pi_u \left(\widehat{g}_u^2 - k_u(\widehat{g}_u)^2 \right) du \ , \tag{14}$$

with $\widehat{g}_u(\cdot)$ the unique solution of the Poisson equation:

$$g(x) - \pi_u(g) = \widehat{g}_u(x) - k_u(\widehat{g}_u)(x) \ . \tag{15}$$

In particular, $W_j(d) \Rightarrow \mathcal{N}(0, \sigma^2_\star)$ with $\sigma^2_\star = \sigma^2_{\phi_0:1}$.

We will now need the following result on the growth of $W_j(d)$.

Lemma 3.1. *Assume (A1(i)(ii), A2) and that $g \in \mathscr{L}_{V^r}$ for some $r \in [0, \frac{1}{2})$. Then, there exists $\delta > 0$ such that:*

$$\sup_d \mathbb{E}\,[\,|W_j(d)\,|^{2+\delta}\,] < \infty \ .$$

Proof. This follows from the decomposition in Theorem A.1 and the following inequality:

$$\mathbb{E}\,[\,|W_j(d)|^{2+\delta}\,] \le \left(\frac{1}{\sqrt{d}}\right)^{2+\delta} M(\delta)\big(\,\mathbb{E}\,[\,|M_{0:d-1}|^{2+\delta}\,] + \mathbb{E}\,[\,|R_{0:d-1}|^{2+\delta}\,]\,\big)\ .$$

Applying the growth bounds in Theorem A.1 we get that the remainder term $\mathbb{E}\,[\,|R_{0:d-1}|^{2+\delta}\,]$ is controlled as $\pi_{\phi_0}(V^r) < \infty$ (due to $r \in [0, \frac{1}{2})$). The martingale array term $\mathbb{E}\,[\,|M_{0:d-1}|^{2+\delta}\,]$ is upper bounded by $Md^{(2+\delta)/2}$, which allows us to conclude. □

One can now obtain the general result.

Theorem 3.2. *Assume (A1(i)(ii), A2). Suppose also that $g \in \mathscr{L}_{V^r}$ for some $r \in [0, \frac{1}{2})$. Then, for any fixed $N > 1$, $\mathrm{ESS}_{(0,d)}(N)$ converges in distribution to*

$$\varepsilon_N := \frac{[\sum_{i=1}^N e^{X_i}]^2}{\sum_{i=1}^N e^{2X_i}}$$

where $X_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_\star^2)$ for $\sigma_\star^2$ specified in Theorem 3.1. In particular,

$$\lim_{d\to\infty} \mathbb{E}\left[\,\mathrm{ESS}_{(0,d)}(N)\,\right] = \mathbb{E}\left[\frac{[\sum_{i=1}^N e^{X_i}]^2}{\sum_{i=1}^N e^{2X_i}}\right]. \tag{16}$$

Proof. We will prove that $\alpha(d)$, as defined in (11), converges in distribution to $\mathcal{N}(0, \sigma_\star^2)$. The argument is standard: it suffices to check that the random variables $\overline{W}_j(d)$, $j = 1, ..., d$, satisfy the Lindeberg condition and that their second moments converge (see e.g. an adaptation of Theorem 2 of [47, pp.334]). To this end, note that $\{\overline{W}_j(d)\}_{d,j}$ form a triangular array of independent variables of zero expectation across each row. Let

$$S_d^2 = \frac{1}{d}\sum_{j=1}^d \mathbb{E}\,[\,\overline{W}_j(d)^2\,] \equiv \mathbb{E}\,[\,\overline{W}_1(d)^2\,]\ ,$$

the last equation following from $\overline{W}_j(d)$ being i.i.d. over j. Now, Theorem 3.1 gives that $W_1(d)$ converges in distribution to $\mathcal{N}(0, \sigma_\star^2)$ for $d \to \infty$. Lemma 3.1 implies that (e.g. Theorem 3.5 of [14]) also the first and second moments of $W_1(d)$ converge to 0 and $\sigma_\star^2$ respectively; we thus obtain:

$$\lim_{d\to\infty} S_d^2 = \sigma_\star^2\ . \tag{17}$$

We consider also the Lindeberg condition, and for each $\epsilon > 0$ we have:

$$\lim_{d\to\infty} \frac{1}{d}\sum_{j=1}^d \mathbb{E}\,[\,\overline{W}_j(d)^2\,\mathbb{I}_{|\overline{W}_j(d)|\ge\epsilon\sqrt{d}}\,] = 0 \tag{18}$$

a result directly implied again from Lemma 3.1. Therefore, by Theorem 2 of [47, pp.334], $\alpha(d)$ converges in distribution to $\mathcal{N}(0, \sigma_\star^2)$. In particular we have proved that:

$$(\alpha_1(d), \ldots, \alpha_N(d)) \Rightarrow \mathcal{N}_N(0, \sigma_\star^2 I_N)\ ,$$

where the subscripts denote the indices of the particles. The result now follows directly after noticing that

$$\mathrm{ESS}_{(0,d)} = \frac{[\sum_{i=1}^N e^{\alpha_i(d)}]^2}{\sum_{i=1}^N e^{2\alpha_i(d)}}$$

and the mapping $(\alpha_1, \alpha_2, \ldots, \alpha_N) \mapsto \frac{[\sum_{i=1}^N e^{\alpha_i}]^2}{\sum_{i=1}^N e^{2\alpha_i}}$ is bounded and continuous. □

3.3 Monte Carlo Error

We have shown that the choice of bridging steps in (6) leads to a stabilization of the ESS in high dimensions. The error in the estimation of expectations, which can be of even more practical interest than ESS, is now considered. In particular, we look at expectations associated with finite-dimensional marginals of the target distribution. Recall the definition of the weight of the i-th particle $w_d(x^i_{0:d-1})$ from (9), for $1 \leq i \leq N$. In order to consider the Monte Carlo error, we use the result below, which is of some interest in its own right.

Proposition 3.1. *Assume (A1(i)(ii), A2). and let $\varphi \in \mathscr{L}_{V^r}$ for $r \in [0,1]$. Then we have:*

$$\lim_{d\to\infty} |\,\mathbb{E}\,[\,\varphi(X_{d,1})\,] - \pi(\varphi)\,| = 0 \ .$$

Proof. This follows from Proposition A.1 in the Appendix when choosing time sequences $s(d) \equiv \phi_0$ and $t(d) \equiv 1$. □

Remark 3.1. *The above result is interesting as it suggests one can run an alternative algorithm that just samples a single inhomogeneous Markov chain, with Markov kernels with invariant measures having annealing parameters on a grid of values and average the values of the function of interest. However, it is not clear how such an algorithm can be validated in practice (that is how many steps one should take for a finite time algorithm) and is of interest in the scenario where one fixes d and allows the time-steps to grow; see [49]. In our context, we are concerned with the performance of the estimator that one would use for fixed d (and hence a finite number of steps in practice) from the SMC sampler in high-dimensions; it is not at all clear a-priori that this will stabilize with a computational cost $\mathcal{O}(Nd^2)$ and if it does, how the error behaves.*

The Monte Carlo error result now follows; recall $\|\cdot\|_\varrho$ is defined in Section 1.3:

Theorem 3.3. *Assume (A1(i)(ii), A2) with $g \in \mathscr{L}_{V^r}$ for some $r \in [0, \frac{1}{2})$. Then for any $1 \leq \varrho < \infty$ there exists a constant $M = M(\varrho) < \infty$ such that for any $N \geq 1$, $\varphi \in \mathcal{C}_b(\mathbb{R})$*

$$\lim_{d\to\infty} \Big\| \sum_{i=1}^{N} \frac{w_d(X^i_{0:d-1})}{\sum_{l=1}^{N} w_d(X^l_{0:d-1})} \varphi(X^i_{d,1}) - \pi(\varphi) \Big\|_\varrho \leq \frac{M(\varrho)\|\varphi\|_\infty}{\sqrt{N}} \big[\, e^{\frac{\sigma_\star^2}{2}\varrho(\varrho-1)} + 1 \,\big]^{1/\varrho}.$$

Proof. Recall that the N particles remain independent. From the definition of the weights in (9), we can write $w_d(X_{0:d-1}) = e^{\frac{1}{\sqrt{d}}\sum_{j=1}^{d}\overline{W}_j(d)}$ for $\overline{W}_j(d)$ being i.i.d. and given in (12). Now, we have shown in the proof of Theorem 3.2 that $\frac{1}{\sqrt{d}}\sum_{j=1}^{d}\overline{W}_j(d) \Rightarrow \mathcal{N}(0,\sigma_\star^2)$, thus:

$$w_d(X_{0:d-1}) \Rightarrow e^X \ , \quad X \sim \mathcal{N}(0,\sigma_\star^2) \ . \tag{19}$$

Then, from Proposition 3.1, $X_{d,1}$ converges weakly to a random variable $Z \sim \pi$. A simple argument shows that the variables Z, X are independent as Z depends only on the first co-ordinate which will not affect (via $\overline{W}_1(d)$) the limit of $\frac{1}{\sqrt{d}}\sum_{j=1}^{d}\overline{W}_j(d)$. The above results allow us to conclude (due to the boundedness and continuity of the involved functions) that:

$$\lim_{d\to\infty} \Big\| \sum_{i=1}^{N} \frac{w_d(X^i_{0:d-1})}{\sum_{l=1}^{N} w_d(X^l_{0:d-1})} \varphi(X^i_{d,1}) - \pi(\varphi) \Big\|_\varrho = \Big\| \sum_{i=1}^{N} \frac{e^{X_i}}{\sum_{l=1}^{N} e^{X_l}} \varphi(Z_i) - \pi(\varphi) \Big\|_\varrho \ , \tag{20}$$

where the X_i are i.i.d. $\mathcal{N}(0,\sigma_\star^2)$ and independently Z_i are i.i.d. π. Now, the limiting random variable in the $\mathbb{L}_\varrho$-norm on the right-hand-side of (20) can be written as:

$$\frac{A_{N,\varphi}}{e^{\sigma_\star^2/2}A_N}\big[\, e^{\sigma_\star^2/2} - A_N \,\big] + e^{-\sigma_\star^2/2}\big[\, A_{N,\varphi} - e^{\sigma_\star^2/2}\pi(\varphi) \,\big] \tag{21}$$

for $A_{N,\varphi} = \frac{1}{N}\sum_{i=1}^{N} e^{X_i}\varphi(Z_i)$ and $A_N = \frac{1}{N}\sum_{l=1}^{N} e^{X_l}$. Now, using the Marcinkiewicz-Zygmund inequality (there is a version with $\varrho \in [1,2)$ see e.g. [21, Chapter 7]), the $\mathbb{L}_\varrho$-norm of the first summand in (21) is upper-bounded by:

$$\frac{\|\varphi\|_\infty}{e^{\sigma_\star^2/2}} \cdot \frac{M(\varrho)}{\sqrt{N}} \, \|e^{X_1} - e^{\sigma_\star^2/2}\|_\varrho$$

where $M(\varrho)$ is a constant that depends upon ϱ only. Then applying the C_p-inequality and doing standard calculation, this is upper-bounded by

$$\frac{M(\varrho)\|\varphi\|_\infty}{\sqrt{N}} \,\big[\, e^{\frac{\sigma_\star^2}{2}\varrho(\varrho-1)} + 1 \,\big]^{1/\varrho}$$

for some finite constant $M(\varrho)$ that only depends upon ϱ. For the $\mathbb{L}_\varrho$-norm of the second summand in (21), again after applying the Marcinkiewicz-Zygmund inequality we have the upper-bound:

$$e^{-\sigma_\star^2/2}\cdot\frac{M(\varrho)}{\sqrt{N}}\,\|e^{X_1}\varphi(Z_1)-e^{\sigma_\star^2/2}\pi(\varphi)\|_\varrho\ .$$

Using the C_p-inequality and standard calculations we have the upper bound:

$$\frac{M(\varrho)\|\varphi\|_\infty}{\sqrt{N}}[e^{\frac{\sigma_\star^2}{2}\varrho(\varrho-1)}+1]^{1/\varrho}$$

for some finite constant $M(\varrho)$ that only depends upon ϱ. Thus, we can easily conclude from here. □

4 Incorporating Resampling

We have already shown that, even without resampling, the expected ESS converges as $d\to\infty$ to a non-trivial limit. In practice, this limiting value could sometimes be prohibitively close to 1 depending on the value of $\sigma_\star^2$; related to this notice that the constant at the upper bound for the Monte Carlo error in Theorem 3.3 is an exponential function of $\sigma_\star^2$ and could be large if $\sigma_\star^2$ is big. As a result, it makes sense to consider the option of resampling in our analysis in high dimensions. We will see that this will result in smaller bounds for Monte Carlo estimates.

The algorithm carries out d steps as in the case of the algorithm without resampling considered in Section 3, but now resampling occurs at the instances when ESS goes below a specified threshold. For fixed d, the algorithm runs in discrete time. Recalling the analogue between discrete and continuous time we have introduced in Remark 2.2 a statement like 'resampling occurred at $t\in[\phi_0,1]$' will literally mean that resampling took place after $l_d(t)$ steps of the algorithm, for the mapping $l_d(t)$ between continuous and discrete instances defined in (7); in particular, the resampling times, when considered on the continuous domain, will lie on the grid G_d:

$$G_d=\{\phi_0+n\,(1-\phi_0)/d\,;\,n=1,\dots,d\}$$

for any fixed d.

Assume that $s\in[\phi_0,1]$ is a resampling time and $x'^{,1}_{l_d(s)},\dots x'^{,N}_{l_d(s)}$ are the (now equally weighted) resampled particles. Due to the i.i.d. assumptions in (4) and (5), after resampling each of these particles will evolve according to the Markov kernels $k_{l_d(s)+1}$, $k_{l_d(s)+2}$, ..., independently over the d co-ordinates and different particles. The empirical ESS will also evolve as:

$$\mathrm{ESS}_{(s,u)}(N)=\frac{\big(\sum_{i=1}^N\exp\{\frac{1}{\sqrt{d}}\sum_{j=1}^d S^i_{s:u,j}\}\big)^2}{\sum_{i=1}^N\exp\{\frac{2}{\sqrt{d}}\sum_{j=1}^d S^i_{s:u,j}\}}\tag{22}$$

for $u\in[s,1]$, where we have defined:

$$S^i_{s:u,j}=\frac{1-\phi_0}{\sqrt{d}}\sum_{n=l_d(s)+1}^{l_d(u)}\{g(x^i_{n-1,j})-\pi_{n-1}(g)\}\ ,\tag{23}$$

until the next resampling instance $t>s$, whence the N particles, $x^i_{l_d(t)}=(x^i_{l_d(t),1},\dots,x^i_{l_d(t),d})$ will be resampled according to their weights:

$$w_{l_d(t)}(x^i_{l_d(s):(l_d(t)-1)})=\exp\{\tfrac{1}{\sqrt{d}}\sum_{j=1}^d S^i_{s:t,j}\}\ .$$

Note that we have modified the subscripts of ESS in (22), compared to the original definition in (2), to now run in continuous time. It should be noted that the dynamics differ from the previous section due to the resampling steps. For instance $S^i_{s:u,j}$ are no longer independent over i or j, unless one conditions on the resampled particles $x'^{,i}_{l_d(s)}$, $1\le i\le N$.

4.1 Theoretical Resampling Times

We start by showing that the dynamically resampling SMC algorithm, using a deterministic version of ESS (namely, the expected ESS associated to the limiting ($N \to \infty$) algorithm) will resample a finite number of times (again as $d \to \infty$) and also exhibit convergence of ESS and of the Monte Carlo error. Subsequently, we show that a dynamically resampling SMC algorithm, using the empirical ESS (with some modification) will, with high probability, display the same convergence properties.

We use the resampling-times construction of [24]: this involves considering the expected value of the importance weight and its square, for a system with a *single* particle. In particular, the theoretical resampling times are defined as (we set $t_0(d) \equiv 0$):

$$t_1(d) = \inf\left\{ t \in [\phi_0, 1] : \frac{\mathbb{E}\left[\,\exp\left\{\frac{1}{\sqrt{d}}\sum_{j=1}^{d} S_{\phi_0:t,j}\right\}\,\right]^2}{\mathbb{E}\left[\,\exp\left\{\frac{2}{\sqrt{d}}\sum_{j=1}^{d} S_{\phi_0:t,j}\right\}\,\right]} < a \right\} ; \tag{24}$$

$$t_k(d) = \inf\left\{ t \in [t_{k-1}(d), 1] : \frac{\mathbb{E}\left[\,\exp\left\{\frac{1}{\sqrt{d}}\sum_{j=1}^{d} S_{t_{k-1}(d):t,j}\right\}\,\right]^2}{\mathbb{E}\left[\,\exp\left\{\frac{2}{\sqrt{d}}\sum_{j=1}^{d} S_{t_{k-1}(d):t,j}\right\}\,\right]} < a \right\} , \quad k \geq 2 , \tag{25}$$

for a constant $a \in (0,1)$, under the convention that $\inf \varnothing = \infty$, where the expected value is w.r.t. the dynamics of a single particle with each of it's co-ordinates moving in-between resampling instances according to our Markov kernels k_n, but drawn independently from $\pi_{t_k(d)}$, for $k = 0, 1, \ldots$, at the resampling instances. Note that, for most applications in practice, these times cannot be found analytically. We emphasize here that the dynamics of $S_{s:t}$ appearing above do not involve resampling but simply follow the evolution of a single particle with d i.i.d. co-ordinates, each of which starts at $\pi_{t_k(d)}$, $k \geq 0$, and then evolves according to the next kernel in the sequence. Intuitively, following the ideas in [24], one could think of the deterministic times in (24)-(25) as the limit of the resampling times of the practical SMC algorithm in Figure 1 as the number of particles N increases to infinity.

We will for the moment consider the behaviour of the above times in high dimensions. Consider the following instances:

$$t_1 = \inf\{ t \in [\phi_0, 1] : e^{-\sigma^2_{\phi_0:t}} < a \} ; \tag{26}$$

$$t_k = \inf\{ t \in [t_{k-1}, 1] : e^{-\sigma^2_{t_{k-1}:t}} < a \} , \quad k \geq 2 , \tag{27}$$

where for any $s < t$ in $[\phi_0, 1]$:

$$\sigma^2_{s:t} = \sigma^2_{\phi_0:t} - \sigma^2_{\phi_0:s} \equiv (1-\phi_0) \int_s^t \pi_u(\widehat{g}_u^2 - k_u(\widehat{g}_u)^2)du . \tag{28}$$

Under our standard assumptions (A1-2), and the requirement that $g \in \mathscr{L}_{V^r}$ for some $r \in [0, \frac{1}{2})$, we have that (using Lemma A.1 in the Appendix):

$$\pi_u(\widehat{g}_u^2 - k_s(\widehat{g}_u)^2) \leq M \pi_u(V^{2r}) \leq M' \pi_{\phi_0}(V) < \infty .$$

Thus, we can find a *finite* collection of times that dominate the t_k's (in the sense that there will be more than them), so also the number of the latter is finite and we can define:

$$m^* = \#\{\, t_k : k \geq 1 \,, t_k \in [\phi_0, 1] \,\} < \infty . \tag{29}$$

We have the following result.

Proposition 4.1. *As $d \to \infty$ we have that $t_k(d) \to t_k$ for any $k \geq 1$.*

Remark 4.1. *Note that the time instances $\{t_k\}$ are derived only through the asymptotic variance function $t \mapsto \sigma^2_{\phi_0:t}$; our main objective in the current resampling part of this paper will be to illustrate that investigation of these deterministic times provides essential information about the resampling times of the practical SMC algorithm in Figure 1. These latter stochastic times will coincide with the former (or, rather, a slightly modified version of it) as $d \to \infty$ with a probability that converges to 1 with a rate $\mathcal{O}(N^{-1/2})$.*

4.2 Stability under Theoretical Resampling Times

Consider an SMC algorithm similar to the one in Figure 1, but with the difference that resampling occurs at the times $\{t_k(d)\}$ in (24)-(25); it is assumed that $t_0(d) = \phi_0$. Note that due to Proposition 4.1, the number of these resampling times:

$$m_d^* = \#\{\, t_k(d) : n \geq 1 \,, t_k(d) \in [\phi_0, 1]\,\}$$

will eventually, for big enough d, coincide with m^* in (29). We will henceforth assume that d is big enough so that $m_d^* \equiv m^* < \infty$.

We state our result in Theorem 4.1 below, under the convention that $t_{m^*+1}(d) \equiv 1$. The proof can be found in Appendix C.2. It relies on a novel construction of a filtration, which starts with all the information of all particles and co-ordinates up-to and including the last resampling time. Subsequent $\sigma-$algebras are generated, for a given particle, by adding each dimension for a given trajectory. This allows one to a use a Martingale CLT approach by taking advantage of the independence of particles and co-ordinates once we condition on their positions at the resampling times.

Theorem 4.1. *Assume (A1-2) and $g \in \mathscr{L}_{V^r}$ with $r \in [0, \frac{1}{2})$. Then, for any fixed $N > 1$, any $k \in \{1, \ldots, m^* + 1\}$, times $t_{k-1} < t_k$, and $s_k(d) \in (t_{k-1}(d), t_k(d))$ any sequence converging to a point $s_k \in (t_{k-1}, t_k)$, we have that* $\mathrm{ESS}_{(t_{k-1}(d), s_k(d))}(N)$ *converges in distribution to a random variable*

$$\frac{[\sum_{i=1}^N e^{X_i^k}\,]^2}{\sum_{i=1}^N e^{2X_i^k}}$$

where $X_i^k \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2_{t_{k-1}:s_k})$ and $\sigma^2_{t_{k-1}:s_k}$ as in (28). In particular,

$$\lim_{d\to\infty} \mathbb{E}\left[\,\mathrm{ESS}_{(t_{k-1}(d), s_k(d))}(N)\,\right] = \mathbb{E}\Big[\frac{[\sum_{i=1}^N e^{X_i^k}\,]^2}{\sum_{i=1}^N e^{2X_i^k}}\Big].$$

Note that, had the $t_k(d)$'s been analytically available, resampling at these instances would deliver an algorithm of d bridging steps for which the expected ESS would be regularly regenerated. In addition, this latter quantity depends, asymptotically, on the 'incremental' variances $\sigma^2_{\phi_0:t_1}$, $\sigma^2_{t_1:t_2}$, ..., $\sigma^2_{t_{m^*}:1}$; in contrast, in the context of Theorem 3.2, the limiting expectation depends on $\sigma^2_{\phi_0:1} \equiv \sigma^2_{\star}$. We can also consider the Monte-Carlo error when estimating expectations w.r.t. a single marginal co-ordinate of our target. Again, the proof is in Appendix C.2.

Theorem 4.2. *Assume (A1-2) with $g \in \mathscr{L}_{V^r}$ for some $r \in [0, \frac{1}{2})$. Then for any $1 \leq \varrho < \infty$ there exists a constant $M = M(\varrho) < \infty$ such that for any fixed $N \geq 1$, $\varphi \in \mathcal{C}_b(\mathbb{R})$*

$$\lim_{d\to\infty} \Big\| \sum_{i=1}^N \frac{w_d(X^i_{l_d(t_{m^*}(d)):(d-1)})}{\sum_{l=1}^N w_d(X^l_{l_d(t_{m^*}(d)):(d-1)})} \varphi(X^i_{d,1}) - \pi(\varphi) \Big\|_{\varrho} \leq \frac{M(\varrho)\|\varphi\|_\infty}{\sqrt{N}} \left[e^{\frac{\sigma^2_{t_{m^*}:1}}{2}\varrho(\varrho-1)} + 1\right]^{1/\varrho}.$$

Remark 4.2. *In comparison to the bound in Theorem 3.3, the bound is smaller with resampling: as $\phi_0 \leq t_{m^*}$ the bound in Theorem 4.2 is clearly less than in Theorem 3.3. Whilst these are both upper-bounds on the error they are based on the same calculations - that is a CLT and using the Marcinkiewicz-Zygmund inequality.*

Remark 4.3. *On inspection, the bound in the above result can be seen as counter-intuitive. Essentially, the bound gets* smaller *as t_{m^*} increases, i.e. the closer to the end one resamples. However, this can be explained as follows. As shown in Proposition 3.1, the terminal point, thanks to the ergodicity of the system, is asymptotically drawn from the correct distribution π. Thus, in the limit $d \to \infty$ the particles do not require weighting. Clearly, in finite dimensions, one needs to assign weights to compensate for the finite run time of the algorithm.*

We remark that our analysis, in the context of resampling, relies on the fact that N is fixed and $d \to \infty$. If N is allowed to grow as well our analysis must be modified when one resamples. Following closely the proofs in the Appendix, it should be possible by considering bounds (which do not increase with N and d) on quantities of the form

$$\mathbb{E}\Big[\sum_{i=1}^N \frac{w_{l_d(t_k(d))}(X^i_{l_d(t_{k-1}(d)):(l_d(t_k(d))-1)})}{\sum_{l=1}^N w_{l_d(t_k(d))}(X^l_{l_d(t_{k-1}(d)):(l_d(t_k(d))-1)})} V(X^i_{l_d(t_k(d))})\Big]$$

to establish results also for large N; we are currently investigating this. However, at least following our arguments, the asymptotics under resampling will only be apparent for N much smaller than d; we believe that is only due to mathematical complexity and does not need to be the case.

4.3 Practical Resampling Times

We now consider the scenario when one resamples at stochastic times. The approach we adopt is to first consider an algorithm which resamples at deterministic times, which the analysis of Section 4.2 applies to. Then, we will show that the algorithm that resamples at the empirical versions of these times will, with high probability that increases with N, resample precisely at those deterministic times (Theorem 4.3 below).

In the context of the above programme, we will follow closely the proof of [24]. For technical reasons (we give more details on this after the statement of Theorem 4.3 below) we will only consider an SMC algorithm which can only possibly resample at values on the following grid G_δ. Let $\delta \in \mathbb{Z}^+$; we define

$$G_\delta = \{\phi_0, \phi_0 + (1-\phi_0)/\delta, \phi_0 + 2(1-\phi_0)/\delta, \ldots, 1\} .$$

We note that, one would like to choose $\delta = d$, but the proof construction we adopt does not appear to be amenable to this scenario and we fix δ and allow d to grow. Thus, we consider the SMC algorithm that attempts to resample only when crossing the instances of the grid G_δ. We now define the following *theoretical* resampling times:

$$t_1^\delta(d) = \inf\left\{ t \in G_\delta \cap [\phi_0, 1] : \frac{\mathbb{E}\left[\, \exp\left\{\frac{1}{\sqrt{d}} \sum_{j=1}^d S_{\phi_0:t,j}\right\}\right]^2}{\mathbb{E}\left[\, \exp\left\{\frac{2}{\sqrt{d}} \sum_{j=1}^d S_{\phi_0:t,j}\right\}\right]} < a_1 \right\} ;$$

$$t_k^\delta(d) = \inf\left\{ t \in G_\delta \cap [t_{k-1}^\delta(d), 1] : \frac{\mathbb{E}\left[\, \exp\left\{\frac{1}{\sqrt{d}} \sum_{j=1}^d S_{t_{k-1}^\delta(d):t,j}\right\}\right]^2}{\mathbb{E}\left[\, \exp\left\{\frac{2}{\sqrt{d}} \sum_{j=1}^d S_{t_{k-1}^\delta(d):t,j}\right\}\right]} < a_k \right\} , \quad k \ge 2 .$$

Let $m_d^*(\delta)$ be the number of times in this sequence. The algorithm which uses these resampling times will resample at deterministic time instances. For example, the first resampling time is the first time $t \in [\phi_0, 1]$ for which:

$$\frac{\mathbb{E}\left[\, \exp\left\{\frac{1}{\sqrt{d}} \sum_{j=1}^d S_{\phi_0:t,j}\right\}\right]^2}{\mathbb{E}\left[\, \exp\left\{\frac{2}{\sqrt{d}} \sum_{j=1}^d S_{\phi_0:t,j}\right\}\right]}$$

drops below a_1 and *also* lies in G_δ (recall the expectation is w.r.t. the path of a single particle). We can, for a moment, obtain an understanding of the behavior of these times as $d \to \infty$. Define:

$$t_1^\delta = \inf\{t \in G_\delta \cap [\phi_0, 1] : e^{-\sigma^2_{\phi_0:t}} < a_1\} ;$$
$$t_k^\delta = \inf\{t \in G_\delta \cap [t_{k-1}^\delta, 1] : e^{-\sigma^2_{t_{k-1}^\delta:t}} < a_k\} , \quad k \ge 2 .$$

If $m^*(\delta)$ denotes the number of these times, we have that $m^*(\delta) \le m^*$ (with m^* now taking into account the choices of different thresholds a_k), but for δ large enough these values will be very close.

Proposition 4.2. *As $d \to \infty$ we have that $t_k^\delta(d) \to t_k^\delta$ for any $k \ge 1$; also $m_d^*(\delta) \to m^*(\delta)$.*

Proof. The proof of $t_1(d) \to t_1$ in Proposition 4.1 is based on showing uniform convergence of

$$t \mapsto \frac{\mathbb{E}\left[\, \exp\left\{\frac{1}{\sqrt{d}} \sum_{j=1}^d S_{\phi_0:t,j}\right\}\right]^2}{\mathbb{E}\left[\, \exp\left\{\frac{2}{\sqrt{d}} \sum_{j=1}^d S_{\phi_0:t,j}\right\}\right]}$$

to $t \mapsto e^{-\sigma^2_{\phi_0:t}}$. Repeating this argument also for subsequent time instances gave that $t_k(d) \to t_k$ for all relevant $k \ge 1$. This uniform convergence result can now be called upon to provide the proof of the current proposition. $\square$

Also, Theorems 4.1 and 4.2 hold under these modified times on G_δ. Again, we shall assume that d is big enough so that $m_d^*(\delta) \equiv m^*(\delta)$.

Main Result and Interpretation

Our objective will now be to consider the *stochastic* resampling times $\{T_k = T_k^N(d)\}$ used in practice when executing the algorithm which are now modified to:

$$T_1 = \inf\{t \in G_\delta \cap [\,\phi_0, 1\,] : \tfrac{1}{N}\,\mathrm{ESS}_{\phi_0:t}(N) < a_1\} ;$$
$$T_k = \inf\{t \in G_\delta \cap [\,T_{k-1}, 1\,] : \tfrac{1}{N}\,\mathrm{ESS}_{T_{k-1}:t}(N) < a_k\} , \quad k \ge 2 ,$$

for a collection of thresholds (a_k) in $(0,1)$. We will use the construction in [24]. The results therein determine the behavior of the SMC method for d *fixed* and increasing number of particles N, as described in the sequel. Define, for a given $\upsilon \in (0,1)$, the following event:

$$\Omega_d^N = \Omega_d^N(\upsilon, \{a_k\}_{1\le k\le m^*(\delta)}) := \Big\{ \text{for all } \ 1 \le k \le m^*(\delta),\, s \in G_d \cap (\, t_{k-1}^\delta(d), t_k^\delta(d)\,] : \\ \big|\, \tfrac{1}{N}\, \mathrm{ESS}_{(t_{k-1}^\delta(d),s)}(N) - \mathrm{ESS}_{(t_{k-1}^\delta(d),s)}\, \big| < \upsilon \,\big|\, \mathrm{ESS}_{(t_{k-1}^\delta(d),s)} - a_k \big| \Big\}$$

where

$$\mathrm{ESS}_{(t_{k-1}^\delta(d),s)} = \frac{\mathbb{E}\,\big[\, \exp\big\{ \frac{1}{\sqrt{d}} \sum_{j=1}^{d} S_{t_{k-1}^\delta(d):s,j} \big\}\,\big]^2}{\mathbb{E}\,\big[\, \exp\big\{ \frac{2}{\sqrt{d}} \sum_{j=1}^{d} S_{t_{k-1}^\delta(d):s,j} \big\}\,\big]}$$

corresponds to the expected ESS over a single particle involved in the definition of $\{t_k^\delta(d)\}$. Here $(a_k)_{1\le k\le m^*}$ are a collection of thresholds sampled from some absolutely continuous distribution; this is simply to avoid the degenerate situation when the a_k's coincide with $\mathrm{ESS}_{(t_{k-1}^\delta(d),s)}$ on the grid, that is for $s \in G_d \cap (\, t_{k-1}^\delta(d), t_k^\delta(d)\,]$ and $1 \le k \le m^*(\delta)$. The work in [24], for fixed d, establishes the following:

1. Within Ω_d^N, if the deterministic resampling criteria tell us to resample, so do the empirical ones. That is:

$$\mathrm{ESS}_{(t_{k-1}^\delta(d),s)} > a_k \;\;\Rightarrow\;\; \tfrac{1}{N}\, \mathrm{ESS}_{(t_{k-1}^\delta(d),s)}(N) > a_k \;, \quad s \in G_\delta \cap (\, t_{k-1}^\delta(d), t_k^\delta(d)\,] \;,$$

   and

$$\mathrm{ESS}_{(t_{k-1}^\delta(d),s)} < a_k \;\;\Rightarrow\;\; \tfrac{1}{N}\, \mathrm{ESS}_{(t_{k-1}^\delta(d),s)}(N) < a_k \;, \quad s \in G_\delta \cap (\, t_{k-1}^\delta(d), t_k^\delta(d)\,] \;.$$

2. A consequence of the above is that (this is Proposition 5.3 of [24]):

$$\bigcap_{1\le k\le m^*(\delta)} \{\, T_k = t_k^\delta(d)\,\} \;\supset\; \Omega_d^N \;.$$

3. Conditionally on $\{a_k\}_{1\le k\le m^*(\delta)}$, we have that $\mathbb{P}\,[\,\Omega \setminus \Omega_d^N\,] \to 0$ as N grows [24, Theorem 5.4] (d is fixed).

The above results provide the interpretation that, with a probability that increases to 1 with N, the theoretical resampling times $\{t_k^\delta(d)\}$ will coincide with the practical $\{T_k = T_k^{\delta,N}(d)\}$, for any fixed dimension d.

Our own contribution involves looking at the stability of these results as the dimension grows, $d \to \infty$.

Theorem 4.3. *Assume (A1-2) and that $g \in \mathscr{L}_{V^r}$, with $r \in [0, \frac{1}{2})$. Conditionally on almost every realization of the random threshold parameters $\{a_k\}$, there exists an $M = M(\delta, \phi_0) < \infty$ such that for any $1 \le N < \infty$, we have:*

$$\lim_{d\to\infty} \mathbb{P}\,[\,\Omega \setminus \Omega_d^N\,] \le \frac{M}{\sqrt{N}} \;.$$

Thus, investigation of the times $\{t_k^\delta\}$ involving only the asymptotic variance function $\sigma_{s:t}^2$ can provide an understanding for the number and location of resampling times of the practical algorithm that uses the empirical ESS. This is because, with high probability, that depends on the number of particles (uniformly in d), the practical resampling times will coincide with $\{t_k(d)\}$.

The proof in Appendix C.3 focuses on point 3. above. It is based on controlling the probability that all empirical ESS's at the instances of the grid are close enough to the corresponding expected ESS's (i.e. closer than the distance of the expected ESS from the relevant threshold). This should also explain the use of the grid G_δ: not doing that, and comparing empirical and expected ESS's on *all* d steps of the sampler would give bounds in the proof of Theorem 4.3 that would depend on d and would be difficult to control as $d \to \infty$. However, in practice our results so far establish that the expected and empirical ESS can only change by $\mathcal{O}(1/d)$ at each of the d steps; thus considering a finite grid already provides important insights for the empirical resampling times.

5 Discussion and Extensions

We now discuss the general context of our results, provide some extra results and look at potential generalizations.

5.1 On the Number of Bridging Steps

Our analysis has relied on using $\mathcal{O}(d)$ bridging steps. An important question is what happens when one has more or less time steps. We restrict our discussion to the case where one does not resample, but one can easily extend the results to the resampling scenario. Suppose one takes $\lfloor d^{1+\delta} \rfloor$ steps, for some real $\delta > -1$ and annealing sequence:

$$\phi_n = \phi_0 + \tfrac{n(1-\phi_0)}{\lfloor d^{1+\delta} \rfloor} \ , \quad n \in \{0, \dots, \lfloor d^{1+\delta} \rfloor\} \ .$$

We are to consider the weak convergence of the centered log-weights, which are now equal to:

$$\frac{\sqrt{d}}{\lfloor d^{1+\delta} \rfloor^{1/2}}\, \alpha_i(d)$$

where we have defined

$$\alpha_i(d) = \frac{1}{\sqrt{d}} \sum_{j=1}^{d} \overline{W}_j(d) \, ; \quad \overline{W}_j(d) = W_j(d) - \mathbb{E}\left[\, W_j(d) \,\right] \, ,$$

with $i \in \{1, \dots, N\}$ and

$$W_j(d) = \frac{1-\phi_0}{\lfloor d^{1+\delta} \rfloor^{1/2}} \sum_{n=1}^{\lfloor d^{1+\delta} \rfloor} \left\{\, g(x_{n-1,j}) - \pi_{n-1}(g) \,\right\} \ .$$

One can follow the arguments of Theorem 3.2 to deduce that, under our conditions:

$$\alpha_i(d) \Rightarrow \mathcal{N}(0, \sigma_\star^2) \ . \tag{30}$$

This observation can the be used to provide the following result.

Corollary 5.1. *Assume (A1(i)(ii), A2) and that $g \in \mathscr{L}_{V^r}$ for some $r \in [0, \frac{1}{2})$. Then, for any fixed $N > 1$:*

- *If $\delta > 0$ then $\mathrm{ESS}_{(0,\lfloor d^{1+\delta} \rfloor)}(N) \rightarrow_{\mathbb{P}} N$.*
- *If $-1 < \delta < 0$ then $\mathrm{ESS}_{(0,\lfloor d^{1+\delta} \rfloor)}(N) \rightarrow_{\mathbb{P}} 1$.*

Proof. Following (30), if $\delta > 0$ then we have that $\frac{\sqrt{d}}{\lfloor d^{1+\delta} \rfloor^{1/2}}\, \alpha_i(d) \rightarrow_{\mathbb{P}} 0$. All particles are independent, so the proof of the ESS convergence follows easily.

For the case when $-1 < \delta < 0$ we work as follows. We consider the maximum $M(d) = \max\{\alpha_i(d); 1 \le i \le d\}$. Let $\bar{\alpha}_{(1)}(d) \le \bar{\alpha}_{(2)}(d) \le \cdots \le \bar{\alpha}_{(N)}(d)$ denote the ordering of the variables $\alpha_1(d) - M(d)$, $\alpha_2(d) - M(d)$, $\dots$, $\alpha_N(d) - M(d)$. We have that (setting for notational convenience $f_d := \frac{\sqrt{d}}{\lfloor d^{1+\delta} \rfloor^{1/2}}$):

$$\mathrm{ESS}_{(0,\lfloor d^{1+\delta} \rfloor)}(N) = \frac{\left(\, \sum_{i=1}^N e^{\alpha_i(d) f_d} \,\right)^2}{\sum_{i=1}^N e^{2\,\alpha_i(d)\, f_d}} \equiv \frac{\left(\, 1 + \sum_{i=1}^{N-1} e^{\bar{\alpha}_{(i)}(d)\, f_d} \,\right)^2}{1 + \sum_{i=1}^{N-1} e^{2\,\bar{\alpha}_{(i)}(d)\, f_d}} \ . \tag{31}$$

Due to the continuity of the involved mappings, the fact that $(\alpha_1(d), \dots, \alpha_N(d)) \Rightarrow \mathcal{N}(0, \sigma_\star^2\, I_N)$ implies the weak limit $(\bar{\alpha}_{(1)}(d), \dots, \bar{\alpha}_{(N-1)}(d)) \Rightarrow (\bar{\alpha}_{(1)}, \dots, \bar{\alpha}_{(N-1)})$ as $d \to \infty$ with the latter variables denoting the ordering $\bar{\alpha}_{(1)} \le \bar{\alpha}_{(2)} \le \cdots \le \bar{\alpha}_{(N)} \equiv 0$ of $\alpha_1 - M, \alpha_2 - M, \dots, \alpha_N - M$ where the α_i's are i.i.d. from $\mathcal{N}(0, \sigma_\star^2)$ and M is their maximum. Since $(\bar{\alpha}_{(1)}(d), \dots, \bar{\alpha}_{(N-1)}(d))$ and their weak limit take a.s. negative values, we have that $(\bar{\alpha}_{(1)}(d) f_d, \dots, \bar{\alpha}_{(N-1)}(d) f_d) \Rightarrow (-\infty, \dots, -\infty)$ which (continuing from (31)) implies the stated result. □

For the stable scenario, with $\delta > 0$, we also have the following.

Corollary 5.2. *Assume (A1(i)(ii), A2) with $g \in \mathscr{L}_{V^r}$ for some $r \in [0, \frac{1}{2})$. Then for any $1 \le \varrho < \infty$, $N \ge 1$, $\varphi \in \mathcal{C}_b(\mathbb{R})$, $\delta > 0$:*

$$\lim_{d \to \infty} \left\| \sum_{i=1}^{N} \frac{w_d(X^i_{0:\lfloor d^{1+\delta} \rfloor - 1})}{\sum_{l=1}^N w_d(X^l_{0:\lfloor d^{1+\delta} \rfloor - 1})} \varphi(X^i_{\lfloor d^{1+\delta} \rfloor, 1}) - \pi(\varphi) \right\|_\varrho = \left\| \frac{1}{N} \sum_{i=1}^{N} \varphi(Z_i) - \pi(\varphi) \right\|_\varrho$$

where $Z_i \overset{i.i.d.}{\sim} \pi$.

Proof. This follows from the proof of Theorem 3.3 and Corollary 5.1. □

Thus, a number of steps of $\mathcal{O}(d)$ is a critical regime: less than this, will lead to the algorithm collapsing w.r.t. the ESS and more steps is 'too-much' effort as one obtains very favourable results.

5.2 Full-Dimensional Kernels

An important open problem is investigation of the stability properties of SMC, as $d \to \infty$, when one uses full-dimensional kernels $K_n(x, dx')$, instead of product of univariate kernels. We still consider here an i.i.d. target and no resampling. Consider the Markov kernel $P_n(x, dx')$ with invariant density Π_n corresponding to RWM with proposal $X_{pr} = x + \sqrt{h}\, Z$ with step-size $h = l^2/d$, $l > 0$, and $Z \sim \mathcal{N}_d(0, I_d)$, so that $X' = x_{pr}$ with probability $a(x, x_{pr}) = 1 \wedge \{\Pi_n(x_{pr})/\Pi_n(x)\}$; otherwise $X' = x$. The particular scaling of h was found [43, 44, 5] to provide algorithms that do not degenerate with d. We consider the SMC method in Figure 1 with $K_n = (P_n)^d$ for RWM so that at each instance n we synthesize d steps from $P_n(x, dx')$. We conjecture that this choice for $K_n(x, dx')$ will be stable as $d \to \infty$. Some of the steps at the investigation of the asymptotic properties of the ESS when using product kernels are: (i) the independency over the d co-ordinates; (ii) each co-ordinate is making $\mathcal{O}(1)$-steps in it's state space with some ergodicity properties. As explained in [43, 44, 5], convolution of d steps for RWM provides, *asymptotically*, independency between the co-ordinates, with each co-ordinate making d steps of size $1/d$ along the path (over the time period $[0, 1]$) of the following limiting scalar SDE:

$$dY_n(t) = \tfrac{a_n(l)l}{2}(\log \pi_n)'(Y_n(t))dt + \sqrt{a_n(l)l}\, d\mathcal{W}_t \tag{32}$$

with $a_n(l) = \lim_{d\to\infty} \mathbb{E}\,[\, a(X, X_{pr})\,] \in (0, 1)$; the expectation is in stationarity, $X \sim \Pi_n$. We conjecture here that the weak limit of the centered log-weights:

$$\tfrac{1}{\sqrt{d}} \sum_{j=1}^{d} \sum_{n=1}^{d} \{\, g(x_{n-1,j}) - \pi_n(g)\,\}/\sqrt{d} \tag{33}$$

would remain unchanged if the dynamics of the Markov chain with kernels $K_n = (P_n)^d$ are replaced with those of a Markov chain with:

$$K_n^*(x, dx') = \prod_{j=1}^{d} k_n^*(x_j, dx_j')\ ; \quad k_n^*(x_j, dx_j') = \mathbb{P}\,[\, Y_n(1) \in dx_j' \,|\, Y_n(0) = x_j\,]\ .$$

With these dynamics, we are in the context of Section 3 and, under the assumptions stated there, we can prove weak convergence of (33) to $\mathcal{N}(0, \sigma_\star^2)$ for $\sigma_\star^2$ now involving the continuum $k_s^*(x_j, dx_j')$ of the SDE transition densities.

Thus, the technical challenge is proving that: $\frac{1}{d}\sum_{n,j=1}^{d} \{\, g(x_{n-1,j}) - g(y_{n-1,j}(1))\,\} \Rightarrow 0$. This requires *coupling* the probability measures $\Pi_0\, K_1 \cdots K_n$ and $\Pi_0\, K_1^* \cdots K_n^*$ determining the dynamics of the time-inhomogeneous d-dimensional Markov chains $\{x_0, x_1, \ldots, x_d\}$ and $\{y_0(1), y_1(1), \ldots, y_d(1)\}$ respectively. This is a non-trivial task that goes beyond the aforementioned MCMC literature where limiting results are based on convergence of generators and do not require strong path-wise convergence. Under our conjecture, the SMC method based on full-dimensional RWM kernels, with stabilize at a total cost of $\mathcal{O}(Nd^3)$. A similar conjecture for MALA (Metropolis-adjusted Langevin algorithm) will involve stability of the SMC method at a reduced cost of $\mathcal{O}(Nd^{7/3})$ as for MALA one has to synthesize $\mathcal{O}(d^{1/3})$ steps of size $\mathcal{O}(d^{-1/3})$ to obtain the diffusion limit (see [44]).

5.3 Beyond I.I.D. Targets

In the MCMC literature first attempts to move beyond i.i.d. targets looked at restricted cases, e.g. [16, 15, 5]. The most recent contributions look at targets defined as changes of measure from Gaussian laws ([11, 38, 42]) containing a large family of practically relevant models (see e.g. [12]). We discuss an extension of our results in this direction. As in [38, 42] we consider a target distribution on an infinite-dimensional separable Hilbert space $\mathcal{H}$ determined via the change of measure:

$$\{d\Pi/d\Pi_0\}(x) \propto \exp\{-\Psi(x)\}\ , \quad x \in \mathcal{H}\ ,$$

for $\Psi : \mathcal{H} \mapsto \mathbb{R}$ and $\Pi_0 = \mathcal{N}(0, \mathcal{C})$ a Gaussian law on $\mathcal{H}$. Let $\{e_j\}_{j\in\mathbb{N}}$ be the orthonormal base of $\mathcal{H}$ made of eigenvectors of $\mathcal{C}$ with eigenvalues $\{\lambda_j^2\}_{n\in\mathbb{N}}$. Π_0 can be expressed in terms of it's *Karhunen-Loève* expansion:

$$\Pi_0 \stackrel{law}{=} \sum_{i=1}^{\infty} \lambda_j \xi_j\, e_j\ , \quad \xi_j \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)\ .$$

In practice, one must consider some d-dimensional approximation, and a standard generic approach for this is to truncate the basis expansion; that is, to work with the d-dimensional target:

$$\Pi(x) \propto \exp\{-\Psi_d(x) - \tfrac{1}{2}\langle x, C_d^{-1}x\rangle\}\ ,\ x \in \mathbb{R}^d\ ; \quad C_d = \mathrm{diag}\{\lambda_1^2, \ldots, \lambda_d^2\}\ ,\ \ \Psi_d(x) = \Psi(\sum_{j=1}^{d} x_j\, e_j)\ .$$

One can use the algorithm in Figure 1 with bridging densities $\Pi_n(x) \propto \{\Pi(x)\}^{\phi_n}$, where $\phi_n = \phi_0 + n(1-\phi_0)/d$, and kernels $K_n = (P_n)^d$, with P_n corresponding to a RWM with target Π_n and proposal $X_{pr} = X + \sqrt{hC_d}\,Z$, with $h = l^2/d$ and $Z \sim \mathcal{N}_d(0, I_d)$. Again, we do not consider resampling. Our conjecture is that the SMC method will be stable as $d \to \infty$, for fixed number of particles N, at a cost of $\mathcal{O}(Nd^3)$. [38] shows that the above choice of step-size h provides a non-degenerate MCMC algorithm as $d \to \infty$. The centered log-weights will now be:

$$\tfrac{1-\phi_0}{d}\sum_{n=1}^{d}\big(-\Psi_d(x_{n-1}) + \mathbb{E}\,[\,\Psi_d(x_{n-1})\,] - \tfrac{1}{2}\langle x_{n-1}, C_d^{-1}\,x_{n-1}\rangle + \tfrac{1}{2}\,\mathbb{E}\,[\,\langle x_{n-1}, C_d^{-1}\,x_{n-1}\rangle\,]\,\big)$$

with $X_n \mid X_{n-1} = x_{n-1} \sim K_n(x_{n-1}, \cdot)$. We conjecture that starting from a d-variate version of the Poisson equation (a generalization of the univariate version for the results proven in this paper) one should aim at showing:

$$\tfrac{1}{d}\sum_{n=1}^{d}\big\{\Psi_d(x_{n-1}) - \mathbb{E}\,[\,\Psi_d(x_{n-1})\,]\,\big\} \Rightarrow 0\ , \quad \tfrac{1}{d}\sum_{n=1}^{d}\big\{-\tfrac{1}{2}\,\langle x_{n-1}, C_d^{-1}\,x_{n-1}\rangle + \tfrac{1}{2}\,\mathbb{E}\,[\,\langle x_{n-1}, C_d^{-1}\,x_{n-1}\rangle\,]\,\big\} \Rightarrow \mathcal{N}(0, \sigma_\star^2)\ ,$$

for some $\sigma_\star^2$. For the first limit, one should consider a Poisson equation associated to the functional $x \mapsto \Psi_d(x)$, for the Markov chain with dynamics K_n. For the second limit, the d-variate Poisson equation should apply upon the functional $x \mapsto \langle x, C_d^{-1}\,x\rangle/\sqrt{d}$. Both these functionals seem to stabilize as $d \to \infty$. The asymptotic variance $\sigma_\star^2$ is expected to involve an integral over the transition density of the limiting $\mathcal{H}$-valued SDEs.

5.4 Further Remarks

An important application of SMC samplers is the estimation of the normalizing constant of Π. This is a non-trivial extension of the work in this article, but we have obtained the stability in high-dimensions of the relative $\mathbb{L}_2$-error of the SMC estimate at a $\mathcal{O}(Nd^2)$ cost with stronger assumptions than in this article; we refer the reader to [9].

Recall we have used the annealing sequence (6). However, one can 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 $\phi_{n,d} = \phi(n/d)$; this is also considered in [9]. The asymptotic results generalized to the choice of $\phi_{n,d}$ here would involve the variances:

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

So, e.g., the bound in Theorem 3.3 becomes $M(\varrho)\|\varphi\|_\infty/\sqrt{N}\,[\,\exp\{\sigma_{\phi_0:1}^{2,\phi}\varrho(\varrho-1)/2\} + 1\,]^{1/\varrho}$. In theory, one could use this quantity to choose between SMC algorithms with different annealing schemes; see [9] for some discussion.

Using the analysis in this article one could consider a comparison between MCMC and SMC. It is certainly the case that for the i.i.d. setup in this paper one could just use the terminal kernel K_d preserving Π to construct an MCMC algorithm which, since it will correspond to d separate MCMC methods along each co-ordinate, would be more efficient (i.e. with regards to computational cost) than the SMC sampler studied here. However, there are a wide variety of practical contexts where one would prefer to use SMC, for instance: multimodal situations (e.g. [22]), some sequential inference problems (e.g. [18]), Approximate Bayesian Computation [23] or when one wants to use consistent adaptive algorithms with little user input [32]. In many of these scenarios the aspect of dimensionality can be a driving reason why the simulation problem is challenging. We mention, that our analysis has been extended to sequential inference under some strong assumptions [9], but in the main, the approach in this article focuses on a very simple model which is potentially not relevant for comparisons between MCMC and SMC. In the context of the above scenarios, such a comparison would be very interesting and it is an important direction of future research.

Acknowledgements

We thank Pierre Del Moral, Christophe Andrieu, Nicolas Chopin and Adam Johansen for valuable conversations associated to this work. The work of Dan Crisan was partially supported by the EPSRC Grant No: EP/H0005500/1 and of Alexandros Beskos by EPSRC Grant No: EP/J01365X/1. The work of Ajay Jasra was supported by a MOE grant and was employed by Imperial College during part of this project. We thank Nikolas Kantas for some useful feedback. We thank the referees and associate editor for extremely useful comments that have greatly improved the paper and the editor (Prof. A. Barbour) for his diligence and patience in assisting us with the submission.

A Technical Results

In this appendix we provide some technical results that will be used in the proofs that follow. The results in Lemma A.1 are fairly standard within the context of the analysis of non-homogeneous Markov chains with drift conditions (e.g. [27]). The decomposition in Theorem A.1 will be used repeatedly in the proofs.

For a starting index $n_0 = n_0(d)$ we denote here by $\{X_n(d)\,;\, n_0 \le n \le d\}$ the non-homogeneous scalar Markov chain evolving via:

$$\mathbb{P}\,[\,X_n(d) \in dy \mid X_{n-1}(d) = x\,] = k_{n,d}(x, dy)\ , \quad n_0 < n \le d\ ,$$

with the kernels $k_{n,d}$ preserving $\pi_{n,d}$. All variables $X_n(d)$ take values in the homogeneous measurable space $(E, \mathscr{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. For simplicity, we will often omit indexing the above quantities with d.

Given the Markov kernel k_s with invariant distribution π_s (here, $s \in [\phi_0, 1]$), and some function φ, we consider the Poisson equation

$$\varphi(x) - \pi_s(\varphi) = f(x) - k_s(f)(x)\ ;$$

under (A1) there is a unique solution $f(\cdot)$ (see e.g. [40]), which can be expressed via the infinite series $f(x) = \sum_{l\ge 0}[k_s^l - \pi_s](\varphi)(x)$. We use the notation $f = \mathcal{P}(\varphi, k_s, \pi_s)$ to define the solution of such an equation.

We will sometimes use the notation $\mathbb{E}_{X_{n_0}}[\,\cdot\,] \equiv \mathbb{E}\,[\,\cdot \mid X_{n_0}\,]$.

Lemma A.1. *Assume (A1-2). Then, the following results hold.*

i) Let $\varphi \in \mathscr{L}_{V^r}$ for some $r \in [0,1]$ and set $\widehat{\varphi} = \mathcal{P}(\varphi, k_s, \pi_s)$. Then, there exists $M = M(r)$ such that

$$|\widehat{\varphi}(x)| \le M |\varphi|_{V^r}\, V(x)^r\ .$$

ii) Let $\varphi_s, \varphi_t \in \mathscr{L}_{V^r}$ for some $r \in [0,1]$ and consider $\widehat{\varphi}_s = \mathcal{P}(\varphi_s, k_s, \pi_s)$ and $\widehat{\varphi}_t = \mathcal{P}(\varphi_t, k_t, \pi_t)$. Then, there exists $M = M(r)$ such that:

$$|\widehat{\varphi}_t(x) - \widehat{\varphi}_s(x)| \le M\,(\,|\varphi_t - \varphi_s|_{V^r} + |\varphi_t|_{V^r}\, |||k_s - k_t|||_{V^r}\,)\, V(x)^r\ .$$

iii) For any $r \in (0,1]$ and $0 \le n_0 \le n$:

$$\mathbb{E}\,[\,V(X_n)^r \,|\, X_{n_0}\,] \le \lambda^{(n-n_0)r} V^r(X_{n_0}) + \tfrac{1-\lambda^{r(n-n_0)}}{1-\lambda^r}\, b^r \le M\, V^r(X_{n_0})\ .$$

Proof. **i):** We proceed using the geometric ergodicity of k_s:

$$|\widehat{\varphi}(x)| = |\,\sum_{l\ge 0}[k_s^l - \pi_s](\varphi)(x)\,| \le |\varphi|_{V^r} \sum_{l\ge 0} \|[k_s^l - \pi_s](x)\|_{V^r} \le M\, |\varphi|_{V^r} [\,\sum_{l\ge 0}\rho^l]\, V(x)^r$$

for some $\rho \in (0,1)$ and $M > 0$ not depending on s via (A1); it is now straightforward to conclude.

ii) Via the Poisson equation we have $\widehat{\varphi}_t(x) - \widehat{\varphi}_s(x) = A(x) + B(x)$ where

$$\begin{aligned} A(x) &= \sum_{l\ge 0}[k_t^l - \pi_t](\varphi_t)(x) - \sum_{l\ge 0}[k_s^l - \pi_s](\varphi_t)(x)\ ; \\ B(x) &= \sum_{l\ge 0}[k_s^l - \pi_s](\varphi_t - \varphi_s)(x)\ . \end{aligned} \tag{34}$$

We start with $B(x)$. For each summand we have:

$$\begin{aligned} |\,[k_s^l - \pi_s](\varphi_t - \varphi_s)(x)\,| &= |\varphi_t - \varphi_s|_{V^r}\, |\,[k_s^l - \pi_s]\big(\tfrac{\varphi_t - \varphi_s}{|\varphi_t - \varphi_s|_{V^r}}\big)(x)\,| \\ &\le |\varphi_t - \varphi_s|_{V^r}\, \|k_s^l - \pi_s\|_{V^r} \le M\, |\varphi_t - \varphi_s|_{V^r}\, \rho^l\, V(x)^r\ , \end{aligned}$$

where $M > 0$ and $\rho \in (0,1)$ depending only on r due to (A1). Hence, summing over l, there exist a $M > 0$ such that for any $x \in E$:

$$B(x) \le M\, |\varphi_t - \varphi_s|_{V^r}\, V(x)^r\ .$$

Returning to $A(x)$ in (34), one can use Lemma C2 of [2] to show that this is equal to:

$$\sum_{l\ge 0}\Big[\sum_{i=0}^{l-1}[k_t^i - \pi_t][k_t - k_s][k_s^{l-i-1} - \pi_s](\varphi_t)(x) - [\pi_t - \pi_s]\big([k_s^l - \pi_s](\varphi_t)\big)\Big]\ .$$

Using identical manipulations to [2], it follows that:

$$\sum_{l\geq 0}\left|\sum_{i=0}^{l-1}[k_t^i-\pi_t][k_t-k_s][k_s^{l-i-1}-\pi_s](\varphi_t)(x)\right|\leq M\,|\varphi_t|_{V^r}\,|||k_s-k_t|||_{V^r}\,V(x)^r$$

and, for some constant $M=M(r)>0$:

$$|\sum_{n\geq 0}[\pi_t-\pi_s]\big([k_s^n-\pi_s](\varphi_t)\big)\,|\leq M\,|\varphi_t|_{V^r}\,|||k_s-k_t|||_{V^r}\,V(x)^r\ .$$

iii) We will use the drift condition in (A1). Using Jensen's inequality (since $r\leq 1$) we obtain $k_n(V^r)(X_{n-1})\leq \lambda^r V^r(X_{n-1})+b^r$ for the constants b, λ appearing in the drift condition. Using this inequality and conditional expectations:

$$\mathbb{E}\,[\,V^r(X_n)\,|\,X_{n_0}\,]=\mathbb{E}\,[\,k_n(V^r(X_{n-1}))\,|\,X_{n_0}\,]\leq\lambda^r\,\mathbb{E}\,[\,V^r(X_{n-1})\,|\,X_{n_0}\,]+b^r\ .$$

Applying this iteratively gives the required result. □

Theorem A.1 (Decomposition)**.** *Assume (A1(i)(ii),A2). Consider the collection of functions $\{\varphi_s\}_{s\in[\phi_0,1]}$ with $\varphi_s\in\mathscr{L}_{V^r}$ for some $r\in[0,1)$ and such that:*

i) $\sup_s|\varphi_s|_{V^r}<\infty$,

ii) $|\varphi_t-\varphi_s|_{V^r}\leq M\,|t-s|$.

Set $\varphi_n(=\varphi_{n,d}):=\varphi_{\{s=\phi_n(d)\}}$ and consider the solution to the Poisson equation $\widehat{\varphi}_n=\mathcal{P}(\varphi_n,k_n,\pi_n)$. Then, for $n_0\leq n_1\leq n_2$ we can write:

$$\sum_{n=n_1}^{n_2}\{\,\varphi_n(X_n)-\pi_n(\varphi_n)\}=M_{n_1:n_2}+R_{n_1:n_2}$$

for the martingale term:

$$M_{n_1:n_2}=\sum_{n=n_1+1}^{n_2}\{\,\widehat{\varphi}_n(X_n)-k_n(\widehat{\varphi}_n)(X_{n-1})\,\}$$

such that for any $p>1$ with $r\,p\leq 1$:

$$\mathbb{E}\,[\,|M_{n_1:n_2}|^p\,|\,X_{n_0}\,]\leq M\,d^{\frac{p}{2}\vee 1}\,V^{rp}(X_{n_0})\ ,$$

and a residual term $R_{n_1:n_2}$ such that for any $p>0$ with $r\,p\leq 1$:

$$\mathbb{E}\,[\,|R_{n_1:n_2}|^p\,|\,X_{n_0}\,]\leq M\,V^{rp}(X_{n_0})\ .$$

Proof. Using the Poisson equation $\varphi_n(\cdot)-\pi_n(\varphi_n)=\widehat{\varphi}_n(\cdot)-k_n(\widehat{\varphi}_n)(\cdot)$, simple addition and subtraction of the appropriate terms gives that:

$$\sum_{n=n_1}^{n_2}\{\,\varphi_n(X_n)-\pi_n(\varphi_n)\,\}=M_{n_1:n_2}+D_{n_1:n_2}-E_{n_1:n_2}+T_{n_1:n_2}\ ; \tag{35}$$

$$\begin{aligned}
D_{n_1:n_2}&=\sum_{n=n_1+1}^{n_2}[\widehat{\varphi}_n(X_{n-1})-\widehat{\varphi}_{n-1}(X_{n-1})]\ ,\\
E_{n_1:n_2}&=\sum_{n=n_1+1}^{n_2}[\varphi_n(X_{n-1})-\varphi_{n-1}(X_{n-1})]\ ,\\
T_{n_1:n_2}&=\widehat{\varphi}_{n_1}(X_{n_1})-\widehat{\varphi}_{n_2}(X_{n_2})-\pi_{n_1}(\varphi_{n_1})+\varphi_{n_2}(X_{n_2})\ .
\end{aligned}$$

Now, using Lemma A.1(i),(iii) and the uniform bound in assumption (i) we get directly that:

$$\mathbb{E}\,[\,|T_{n_1:n_2}|^p\,\mid\,X_{n_0}\,]\leq M\ V^{rp}(X_{n_0})\ . \tag{36}$$

Also, Lemma A.1(i) together with assumption (i) imply that:

$$|\,(\varphi_n-\varphi_{n-1})(X_{n-1})\,| \leq |\varphi_n-\varphi_{n-1}|_{V^r}\, V^r(X_{n-1}) \leq M\,\tfrac{1}{d}\,V^r(X_{n-1})\ ,$$

thus, calling again upon Lemma A.1(iii), one obtains that:

$$\mathbb{E}\,[\,|E_{n_1:n_2}|^p \mid X_{n_0}\,] \leq M\,V^{rp}(X_{n_0})\ . \tag{37}$$

Consider now $D_{n_1:n_2}$. Using first Lemma A.1(ii), then conditions (i)-(ii) and (A2) one yields:

$$|\widehat{\varphi}_n(X_{n-1})-\widehat{\varphi}_{n-1}(X_{n-1})| \leq M\,\tfrac{1}{d}\,V(X_{n-1})^r\ .$$

Thus, using also Lemma A.1(iii) we obtain directly that:

$$\mathbb{E}\,[\,|D_{n_1:n_2}|^p \mid X_0\,] \leq M\,V(X_{n_0})^{rp}\ . \tag{38}$$

The bounds (36), (37) and (38) prove the stated result for the growth of $\mathbb{E}\,[\,|R_{n_1:n_2}|^p\,]$.

Now consider the martingale term $M_{n_1:n_2}$. One can use a modification of the Burkholder-Davis-Gundy inequality (e.g. [47, pp. 499-500]) which states that for any $p>1$:

$$\mathbb{E}\,[\,|M_{n_1:n_2}|^p\,|\,X_{n_0}\,] \leq M(p)\,d^{\frac{p}{2}\vee 1-1}\sum_{n=n_1+1}^{n_2}\mathbb{E}\,[\,|\,\widehat{\varphi}_n(X_n)-k_n(\widehat{\varphi}_n)(X_{n-1})\,|^p\,|\,X_{n_0}\,]\ , \tag{39}$$

see [3] for the proof. Using Lemma A.1(i) we obtain that:

$$|\,\widehat{\varphi}_n(X_n)-k_n(\widehat{\varphi}_n)(X_{n-1})\,| \leq M\,|\varphi_n|_{V^r}\,(\,V^r(X_n)+k_n(V^r)(X_{n-1})\,)\ .$$

Using this bound, Jensen inequality giving $(k_n(V^r)(X_{n-1}))^p \leq k_n(V^{rp})(X_{n-1})$, the fact that $r\,p\leq 1$ and Lemma A.1(iii), we continue from (39) to obtain the stated bound for $M_{n_1:n_2}$. □

Proposition A.1. *Let $\varphi\in\mathscr{L}_{V^r}$ with $r\in[0,1]$. Consider two sequences of times $\{s(d)\}_d$, $\{t(d)\}_d$ in $[\phi_0,1]$ such that $s(d)<t(d)$ and $s(d)\to s$, $t(d)\to t$ with $s<t$. If we also have that* $\sup_d\,\mathbb{E}\,[\,V^r(X_{l_d(s(d))})\,]<\infty$*, then:*

$$\mathbb{E}_{X_{l_d(s(d))}}\,[\,\varphi(X_{l_d(t(d))})\,] \to \pi_t(\varphi)\ ,\quad \textit{in }\ \mathbb{L}_1\ .$$

Proof. Recall that $\pi_u(x)\propto\exp\{u\,g(x)\}$ for $u\in[\phi_0,1]$. We define, for $c\in(0,\frac{1}{2})$:

$$n_d = l_d(t(d))-l_d(s(d))\ ;\quad m_d=\lfloor\{l_d(t_d)-l_d(s_d)\}^c\rfloor\ ;\quad u_d=l_d(s(d))+n_d-m_d\ .$$

Note that from the definition of $l_d(\cdot)$ we have $n_d=\mathcal{O}(d)$, whereas $m_d=\mathcal{O}(d^c)$. We have that:

$$\begin{aligned}|\,\mathbb{E}_{X_{l_d(s(d))}}\,[\,\varphi(X_{l_d(t(d))})\,]-\pi_t(\varphi)\,| &\leq |\,\mathbb{E}_{X_{l_d(s(d))}}\,[\,\varphi(X_{l_d(t(d))})-k_{u_d}^{m_d}(\varphi)(X_{u_d})\,]\,| \\ &\quad +|\,\mathbb{E}_{X_{l_d(s(d))}}\,[\,k_{u_d}^{m_d}(\varphi)(X_{u_d})\,]-\pi_{u_d}(\varphi)\,|+|\,\pi_{u_d}(\varphi)-\pi_t(\varphi)\,|\ .\end{aligned} \tag{40}$$

Now, the last term on the R.H.S. of (40) goes to zero as $d\to\infty$: this is via dominated convergence after noticing that

$$\pi_{u_d}(\varphi)=\frac{\int\varphi(x)e^{(\phi_0+\frac{u_d}{d}(1-\phi_0))g(x)}dx}{\int e^{(\phi_0+\frac{u_d}{d}(1-\phi_0))g(x)}dx}$$

with the integrand of the term, for instance, in the numerator converging almost everywhere (w.r.t. Lebesque) to $\varphi(x)e^{t\,g(x)}$ (simply notice that $\lim u_d/d=\lim\{l_d(t(s))/d\}=(t-\phi_0)/(1-\phi_0)$) and being bounded in absolute value (due to the assumption of g being upper bounded) by the integrable function $M\,V^r(x)e^{\phi_0 g(x)}$. Also, the second term on the R.H.S. of (40) goes to zero in $\mathbb{L}_1$, due the uniform in drift condition in (A1); to see this, note that (working as in the proof of Lemma A.1(i)) condition A1 gives $\|k_s^l-\pi_s\|_{V^r}\leq M\,\rho^l\,V(x)^r$ for any $s\in(\phi_0,1]$, so we also have that $|k_{u_d}^{m_d}(\varphi)(X_{u_d})-\pi_{u_d}(\varphi)|\leq M\,\rho^{m_d}\,V(X_{u_d})^r$. Taking expectations and using Lemma A.1(iii) we obtain that:

$$|\,\mathbb{E}_{X_{l_d(s(d))}}\,[\,k_{u_d}^{m_d}(\varphi)(X_{u_d})\,]-\pi_{u_d}(\varphi)\,| \leq M\,\rho^{m_d}\,V(X_{l_d(s(d))})^r\ .$$

which vanishes in $\mathbb{L}_1$ as $d\to\infty$ due to the assumption $\sup_d\,\mathbb{E}\,[\,V^r(X_{l_d(s(d))})\,]<\infty$.

We now focus on the first term on the R.H.S. of (40). The following decomposition holds, as intermediate terms in the sum below cancel out, for $u_d \geq 1$:

$$\mathbb{E}_{X_{l_d(s(d))}}\left[\,\varphi(X_{l_d(t(d))}) - k_{u_d}^{m_d}(\varphi)(X_{u_d})\,\right] =$$
$$\mathbb{E}_{X_{l_d(s(d))}}\Big[\,\sum_{j=0}^{m_d-1}\{k_{(u_d+1):(l_d(t(d))-j)}\,k_{u_d}^{j}(\varphi)(X_{u_d}) - k_{(u_d+1):(l_d(t(d))-(j+1))}k_{u_d}^{j+1}(\varphi)(X_{u_d})\}\,\Big]$$

where we use the notation $k_{i:j}(\varphi)(x) = \int k_i(x,dx_1)\times\cdots\times k_j(\varphi)(x_{j-i+1})$, $i\leq j$. Each of the summands is equal to

$$k_{u_d+1:l_d(t(d))-(j+1)}[k_{l_d(t(d))-j} - k_{u_d}](k_{u_d}^{j}(\varphi))(X_{u_d})$$

which is bounded in absolute value by

$$M\,|\varphi|_{V^r}\,k_{u_d+1:l_d(t(d))-(j+1)}(V^r)(X_{u_d})\,|||k_{l_d(t(d))-j} - k_{u_d}|||_{V^r}\ .$$

Now, from Lemma A.1(iii):

$$k_{u_d+1:l_d(t(d))-(j+1)}(V^r)(X_{u_d}) \leq MV^r(X_{u_d})\ .$$

Also, from condition (A2), there exists an $M>0$ such that

$$|||k_{l_d(t(d))-j} - k_{u_d}|||_{V^r} \leq M\,\tfrac{(1-\phi_0)}{d}\,(l_d(t(d)) - j - u_d) \equiv M\,\tfrac{(1-\phi_0)}{d}\,(m_d - j)\ .$$

Thus, using again Lemma A.1(iii) we are left with

$$|\,\mathbb{E}_{X_{l_d(s(d))}}\left[\,\varphi(X_{l_d(t(d))}) - k_{u_d}^{m_d}(\varphi)(X_{u_d})\,\right]| \leq M\,V^r(X_{l_d(s(d))})\sum_{j=0}^{m_d-1}\frac{m_d-j}{d}\ .$$

As $\sup_d \mathbb{E}\,[\,V^r(X_{l_d(s(d))})\,] < \infty$, since $m_d = \mathcal{O}(d^c)$ with $c\in(0,\frac{1}{2})$ we can easily conclude. □

B Proofs for Section 3

There are related results to Theorem 3.1 (see e.g. [39, 52]), however in our case, the proofs will be based on assumptions commonly made in the MCMC and SMC literature, which will be easily verifiable. The general framework will involve constructing a Martingale difference array (an approach also followed in the above mentioned papers).

Proposition B.1. *Assume (A1(i)(ii), A2) and $g\in\mathscr{L}_{V^r}$ with $r\in[0,\frac{1}{2})$. The family of functions $\{\varphi_s\}_{s\in[\phi_0,1]}$ specified as:*

$$\varphi_s(x) = k_s(\widehat{g}_s^2)(x) - \{k_s(\widehat{g}_s)(x)\}^2\ ,\quad \widehat{g}_s = \mathcal{P}(g,k_s,\pi_s)\ ,$$

satisfies conditions (i) and (ii) of Theorem A.1 for $\bar{r} = 2r \in [0,1)$.

Proof. Lemma A.1(i) gives that $|\widehat{g}_s(x)| \leq M\,|g|_{V^r}\,V^r(x)$. Thus, due to the presence of quadratic functions in the definition of $\varphi_s(\cdot)$ we get directly that $|\varphi_s(x)| \leq M\,V^{\bar{r}}(x)$ so condition (i) in Theorem A.1 is satisfied. We move on to condition (ii) of the theorem. Let us first deal with:

$$\{k_t(\widehat{g}_t)(x)\}^2 - \{k_s(\widehat{g}_s)(x)\}^2$$

which is equal to

$$\{k_t(\widehat{g}_t)(x) - k_s(\widehat{g}_t)(x)\}\{k_t(\widehat{g}_t)(x) + k_s(\widehat{g}_t)(x)\} + \{k_s(\widehat{g}_t - \widehat{g}_s)(x)\}\{k_s(\widehat{g}_t + \widehat{g}_s)(x)\}\ .$$

The terms with the additions are bounded in absolute value by $M\,V^r(x)$, whereas:

$$|\,k_t(\widehat{g}_t)(x) - k_s(\widehat{g}_t)(x)\,| \leq M\,|t-s|\,V(x)^r\ ,\quad |\,k_s(\widehat{g}_t - \widehat{g}_s)(x)\,| \leq M\,|t-s|\,V(x)^r\ ,$$

the first inequality following from assumption (A2) and the second from Lemma A.1(ii). Thus, we have proved:

$$|\,\{k_t(\widehat{g}_t)(x)\}^2 - \{k_s(\widehat{g}_s)(x)\}^2\,| \leq M\,|t-s|\,V(x)^{\bar{r}}$$

for $\bar{r} = 2r \in (0,1)$. We move on to the second term at the expression for φ_s and work as follows:

$$k_t(\widehat{g}_t^2)(x) - k_s(\widehat{g}_s^2)(x) = k_t(\widehat{g}_t^2)(x) - k_s(\widehat{g}_t^2)(x) + k_s(\widehat{g}_t^2)(x) - k_s(\widehat{g}_s^2)(x) \ .$$

The first difference is controlled, from assumption (A2), by $M\,|t-s|\,V(x)^{\bar{r}}$, whereas for the second difference we use Cauchy-Schwarz to obtain:

$$\begin{aligned}|k_s(\widehat{g}_t^2)(x) - k_s(\widehat{g}_s^2)(x)| &\leq \{k_s(\widehat{g}_t - \widehat{g}_s)^2(x)\}^{1/2}\{k_s(\widehat{g}_t + \widehat{g}_s)^2(x)\}^{1/2} \\ &\leq M\,|t-s|\,V(x)^{\bar{r}}\end{aligned}$$

where, for the second inequality, we have used Lemma A.1(ii). The proof is now complete. □

Proof of Theorem 3.1. We adopt the decomposition as in Theorem A.1. Set $\widehat{g}_s$ to be a solution to the Poisson equation (with π_s, k_s) and $\widehat{g}_{n-1,d} = \widehat{g}_{\{s=\phi_{n-1}\}}$. The decomposition is then:

$$\sum_{n=1}^{l_d(t)}\{g(X_{n-1}(d)) - \pi_{n-1,d}(g)\} = M_{0:l_d(t)-1} + R_{0:l_d(t)-1}$$

where

$$M_{0:l_d(t)-1} = \sum_{n=1}^{l_d(t)-1}\{\widehat{g}_{n,d}(X_n(d)) - k_{n,d}(\widehat{g}_{n,d})(X_{n-1}(d))\} \ .$$

It is clear, via Theorem A.1, that $R_{0:l_d(t)-1}/\sqrt{d}$ goes to zero in $\mathbb{L}_1$ and hence we need consider the Martingale array term only.

Writing

$$\xi_{n,d} = \widehat{g}_{n,d}(X_n(d)) - k_{n,d}(\widehat{g}_{n,d})(X_{n-1}(d))$$

one observes that $\{\xi_{n,d}, \mathscr{F}_{n,d}\}_{n=1}^{d-1}$, with $\mathscr{F}_{n,d}$ denoting the filtration generated by $\{X_n(d)\}$, is a square-integrable Martingale difference array with zero mean. In order to prove the fCLT, one can use Theorem 5.1 of [6] which gives the following sufficient conditions for proving Theorem 3.1:

a) For every $\epsilon > 0$, $I_{\epsilon,d} := \frac{1}{d}\sum_{n=1}^{d}\mathbb{E}\,[\,\xi_{n,d}^2\,\mathbb{I}_{|\xi_{n,d}|\geq\epsilon\sqrt{d}} \mid \mathscr{F}_{n-1,d}\,] \to 0$ in probability.

b) For any $t \in [\phi_0, 1]$, $I_d(t) := \frac{1}{d}\sum_{n=1}^{l_d(t)}\mathbb{E}\,[\,\xi_{n,d}^2 \mid \mathscr{F}_{n-1,d}\,]$ converges in probability to the quantity $\sigma_{\phi_0:t}^2/(1-\phi_0)^2$.

We proceed by proving these two statements.

We prove a) first. Recall that $r \in [0, \frac{1}{2})$, so we can choose $\delta > 0$ so that $r(2+\delta) \leq 1$. In the first line below, one can use simple calculations and in the second line Lemma A.1(i) and the drift condition with $r(2+\delta) \leq 1$, to obtain:

$$\begin{aligned}|\xi_{n,d}|^{2+\delta} &\leq M(\delta)\big(\,|\widehat{g}_{n,d}(X_n(d))|^{2+\delta} + |k_{n,d}(\widehat{g}_{n,d})(X_{n-1}(d))|^{2+\delta}\,\big) \\ &\leq M(\delta)\big(\,V(X_n(d)) + V(X_{n-1}(d))\,\big) \ ,\end{aligned}$$

Thus, using Lemma A.1(iii) we get: $\sup_{n,d}\mathbb{E}\,[\,|\xi_{n,d}|^{2+\delta}\,] < \infty$. A straightforward application of Hölder's inequality, then followed by Markov's inequality, now gives that:

$$\mathbb{E}\,[\,I_{\epsilon,d}\,] \leq \frac{1}{d}\sum_{n=1}^{d}\big(\,\mathbb{E}\,[\,|\xi_{n,d}|^{2+\delta}\,]\,\big)^{\frac{2}{2+\delta}}\,\big(\,\mathbb{P}\,[\,|\xi_{n,d}| \geq \epsilon\sqrt{d}\,]\,\big)^{\frac{\delta}{2+\delta}} \leq M\,d^{-\frac{1}{2}\frac{\delta}{2+\delta}} \ .$$

Thus, we have proved a).

For b), we can rewrite:

$$I_d(t) = \frac{1}{d}\sum_{n=1}^{l_d(t)}\Big[k_{n,d}(\widehat{g}_{n,d}^2)(X_{n-1}(d)) - \big\{k_{n,d}(\widehat{g}_{n,d})(X_{n-1}(d))\big\}^2\Big] \ . \tag{41}$$

We will be calling upon Theorem A.1 to prove convergence of the above quantity to an asymptotic variance. Note that, via Proposition B.1, the mappings

$$\varphi_s := k_s(\widehat{g}_s^2) - \big\{k_s(\widehat{g}_s)\big\}^2$$

satisfy conditions (i)-(ii) of Theorem A.1. We define $\varphi_{n,d} = \varphi_{\{s=\phi_n(d)\}}$ and rewrite $I_d(t)$ as:

$$I_d(t) = \frac{1}{d} \sum_{n=0}^{l_d(t)-1} \varphi_{n+1,d}(X_n(d)) \ .$$

We also define:

$$J_d(t) = \frac{1}{d} \sum_{n=0}^{l_d(t)-1} \varphi_{n,d}(X_n(d)) \ .$$

Due to condition (ii) of Theorem A.1, we have that $I_d(t) - J_d(t) \to 0$ in $\mathbb{L}_1$. Applying Theorem A.1 one can deduce that:

$$\lim_{d\to\infty} \{ \, J_d(t) - \frac{1}{d} \sum_{n=0}^{l_d(t)-1} \pi_{n,d}(\varphi_{n,d}) \, \} = 0 \ , \quad \text{in } \mathbb{L}_1 \ .$$

Now, $s \mapsto \pi_s(\varphi_s)$ is continuous as a mapping on $[\phi_0, 1]$, so from standard calculus we get that $\frac{1-\phi_0}{d} \sum_{n=0}^{l_d(t)-1} \pi_{n,d}(\varphi_{n,d}) \to \int_{\phi_0}^{t} \pi_s(\varphi_s) ds$. Combining the results, we have proven that:

$$I_d(t) \to (1-\phi_0)^{-1} \int_{\phi_0}^{t} \pi_s(\varphi_s) ds \equiv \sigma^2_{\phi_0:t}/(1-\phi_0)^2 \ , \quad \text{in } \mathbb{L}_1 \ .$$

Note that by Corollary 3.1 of Theorem 3.2 of [29] we also have an CLT for S_1.

□

C Proofs for Section 4

C.1 Results for Proposition 4.1

We will first require a proposition summarising convergence results, with emphasis on uniform convergence w.r.t. the time index.

Proposition C.1. *Assume (A1-2). Let $s(d)$ be a sequence on $[\phi_0, 1]$ such that $s(d) \to s$. Consider the random variable $S_{s(d):t,j}$ as defined in (23) having it's distribution determined by a single particle, with it's j-th co-ordinate sampled from $\pi_{l_d(s(d))}$ at step $l_d(s(d))$ and then propagated according to the appropriate Markov kernels $\{k_n\}$; all d co-ordinates are i.i.d.. We then have:*

i) $\sup_{t\in[s(d),1]} \mathbb{E}\,[\,|S_{s(d):t,j}\,|\,]/\sqrt{d} \to 0$.

ii) $\sup_{t\in[s(d),1]} |\,\mathbb{E}\,[\,S^2_{s(d):t,j}\,] - \sigma^2_{s:t}\,| \to 0$.

iii) $\sup_{t\in[s(d),1]} |\,\mathbb{E}\,[\,S_{s(d):t,j}\,]\,| \to 0$.

iv) $\sup_{d\geq 1, s\in[s(d),t]} \mathbb{E}\,[\,S^{2+\epsilon}_{s(d):t}\,] < \infty$, *for some* $\epsilon > 0$.

Proof. For simplicity, we will omit reference to the co-ordinate index j. Applying the decomposition of Theorem A.1 for $\varphi_s \equiv g$ and $n_0 = l_d(s(d))$ gives that:

$$S_{s(d):t} = \tfrac{(1-\phi_0)}{\sqrt{d}} \left(M_{l_d(s(d)):(l_d(t)-1)} + R_{l_d(s(d)):(l_d(t)-1)} \right)$$

with (choosing $p = 2+\epsilon$ for $\epsilon > 0$ so that $r\,p < 1$):

$$\mathbb{E}[\,|M_{l_d(s(d)):(l_d(t)-1)}|^{2+\epsilon}\,] \leq M\, d^{1+\frac{\epsilon}{2}}\, \mathbb{E}\,[\,V(X_{l_d(s(d))})\,] \leq M'\, d^{1+\frac{\epsilon}{2}}\, \pi_0(V)$$

and (choosing $p = 2+\epsilon$ for $\epsilon > 0$ so that $r\,p < 1$):

$$\mathbb{E}[\,|R_{l_d(s(d)):(l_d(t)-1)}|^{2+\epsilon}\,] \leq M\, \mathbb{E}\,[\,V(X_{l_d(s(d))})\,] \leq M'\, \pi_0(V) \ .$$

Notice that in both cases, the right-most inequality is due to the fact that for any $s \in [\phi_0, 1]$ we have that:

$$\pi_s(V) \leq \pi_0(V) \cdot \tfrac{\int \exp\{\phi_0(g(x)-g_{\max})\}dx}{\int \exp\{g(x)-g_{\max}\}dx}$$

where $g_{\max} = \sup g < \infty$. Returning to the two main inequalities, one now needs to notice that the bounds are *uniform* in s, t, d, thus statements (i) and (iv) of the proposition follow directly from the above estimates; statement (iii) also follows directly after taking under consideration that $\mathbb{E}\,[\, M_{l_d(s(d)):(l_d(t)-1)}\,] = 0$. It remains to prove (ii). The residual term $R_{l_d(s(d)):(l_d(t)-1)}/\sqrt{d}$ vanishes in the limit in $\mathbb{L}_{2+\epsilon}$-norm, thus it will not affect the final result, that is:

$$\sup_{t\in[s(d),1]} |\,\mathbb{E}\,[\, S^2_{s(d):t}\,] - \tfrac{(1-\phi_0)^2}{d}\,\mathbb{E}\,[\, M^2_{d,l_d(s(d)):(l_d(t)-1)}\,]\,| \to 0\ .$$

Now, straightforward analytical calculations yield:

$$\begin{aligned}\tfrac{1}{d}\,\mathbb{E}\,[\, M^2_{d,l_d(s(d)):(l_d(t)-1)}\,] &= \frac{1}{d}\sum_{n=l_d(s(d))}^{l_d(t)-1} \mathbb{E}\,[\,\{\widehat{g}_n(X_n) - k_n(\widehat{g}_n)(X_{n-1})\}^2\,] \\ &= \mathbb{E}\,\Big[\,\frac{1}{d}\sum_{n=l_d(s(d))-1}^{l_d(t)-2} \varphi_{n+1}(X_n)\Big]\ ,\end{aligned}$$

where we have set:

$$\varphi_s = k_s(\widehat{g}_s^2) - \{k_s(\widehat{g}_s)\}^2\ ;\quad \varphi_n = \varphi_{\{s=\phi_n\}}\ .$$

Since $|\varphi_{n+1} - \varphi_n|_{V^{2r}} \le M\frac{1}{d}$ from Proposition B.1, we also have:

$$\sup_{t\in[s(d),1]}\Big|\,\mathbb{E}\,\Big[\,\frac{1}{d}\sum_{n=l_d(s(d))-1}^{l_d(t)-2}\varphi_{n+1}(X_n)\Big] - \mathbb{E}\,\Big[\,\frac{1}{d}\sum_{n=l_d(s(d))-1}^{l_d(t)-2}\varphi_n(X_n)\Big]\,\Big| \to 0\ .$$

Now, Theorem A.1 and Proposition B.1 imply that:

$$\sup_{t\in[s(d),1]}\mathbb{E}\,\Big|\,\frac{1}{d}\sum_{n=l_d(s(d))-1}^{l_d(t)-2}\{\varphi_n(X_n) - \pi_n(\varphi_n)\}\,\Big| \to 0\ .$$

Finally, due to the continuity of $s \mapsto \pi_s(\varphi_s)$, it is a standard result from Riemann integration (see e.g. Theorem 6.8 of [45]) that:

$$\sup_{t\in[s(d),1]}\Big|\,\tfrac{1-\phi_0}{d}\sum_{n=l_d(s(d))-1}^{l_d(t)-2}\pi_n(\varphi_n) - \int_s^t \pi_u(\varphi_u)du\,\Big| \to 0$$

and we conclude. □

Proof of Proposition 4.1. For some sequence $s(d)$ in $[\phi_0, 1]$ such that $s(d) \to s$, we will consider the function in $t \in [s(d), 1]$:

$$f_d(s(d),t) := \frac{\mathbb{E}^2\,\big[\,\exp\big\{\frac{1}{\sqrt{d}}\sum_{j=1}^d S_{s(d):t,j}\big\}\,\big]}{\mathbb{E}\,\big[\,\exp\big\{\frac{2}{\sqrt{d}}\sum_{j=1}^d S_{s(d):t,j}\big\}\,\big]} \equiv \left(\frac{\mathbb{E}^2\,\big[\,\exp\big\{\frac{1}{\sqrt{d}}S_{s(d):t,1}\big\}\,\big]}{\mathbb{E}\,\big[\,\exp\big\{\frac{2}{\sqrt{d}}S_{s(d):t,1}\big\}\,\big]}\right)^d$$

the second result following due to the independence over j. In the rest of the proof we will omit reference to the co-ordinate index 1. Due to the ratio in the definition of $f_d(s(d),t)$, we can clearly re-write:

$$f_d(s(d),t) = \left(\frac{\mathbb{E}^2\,\big[\,\exp\big\{\frac{1}{\sqrt{d}}\overline{S}_{s(d):t}\big\}\,\big]}{\mathbb{E}\,\big[\,\exp\big\{\frac{2}{\sqrt{d}}\overline{S}_{s(d):t}\big\}\,\big]}\right)^d$$

for $\overline{S}_{s(d):t} = S_{s(d):t} - \mathbb{E}\,[\,S_{s(d):t}\,]$. We will use the notation '$h_d(t) \to_t h(t)$' to denote convergence, as $d \to \infty$, uniformly for all t in $[s(d), 1]$, that is $\sup_{t\in[s(d),t]}|h_d(t) - h_t| \to 0$. We will aim at proving, using the results in Proposition C.1, that:

$$f_d(s(d),t) \to_t e^{-\sigma^2_{s:t}}\ , \tag{42}$$

or, equivalently, that $\sup_{t\in[s(d),1]}|f_d(s(d),t) - e^{-\sigma^2_{s:t}}| \to 0$, under the convention that $\sigma^2_{s:t} \equiv 0$ for $t \le s$. Once we have obtained this, the required result will follow directly by induction. To see that, note that for proving that $t_1(d) \to t_1$ we will use the established result for $s(d) \equiv \phi_0$: uniform convergence of $f_d(\phi_0, t)$ to $e^{-\sigma^2_{\phi_0:t}}$ together with the fact that $e^{-\sigma^2_{\phi_0:t}}$ is decreasing in t will give directly that the hitting time of the threshold a for $f_d(\phi_0, t)$

will converge to that of $e^{-\sigma^2_{\phi_{0:t}}}$. Now, assuming we have proved that $t_n(d) \to t_n$, we will then use the established uniform convergence result for $s(d) = t_n(d)$ to obtain directly that $t_{n+1}(d) \to t_{n+1}$.

We will now establish (42). Note that we have, by construction: $\mathbb{E}\,[\,\overline{S}_{s(d):t}\,] = 0$. We use directly Taylor expansions to obtain for any fixed $t \in [s(d), 1]$:

$$e^{\frac{2}{\sqrt{d}}\overline{S}_{s(d):t}} = 1 + \tfrac{2}{\sqrt{d}}\,\overline{S}_{s(d):t} + \tfrac{2}{d}\,\overline{S}^2_{s(d):t} e^{2\zeta_{d,t}} \ ; \tag{43}$$

$$e^{\frac{1}{\sqrt{d}}\overline{S}_{s(d):t}} = 1 + \tfrac{1}{\sqrt{d}}\,\overline{S}_{s(d):t} + \tfrac{1}{2d}\,\overline{S}^2_{s(d):t}\, e^{\zeta'_{d,t}} \ , \tag{44}$$

where $\zeta_{d,t}, \zeta'_{d,t} \in \big[\,\frac{1}{\sqrt{d}}\,\overline{S}_{s(d):t}\wedge 0\,,\ \frac{1}{\sqrt{d}}\overline{S}_{s(d):t}\vee 0\,\big]$. Note here that since g is upper bounded and $\sup_{n,d}\mathbb{E}\,[\,|g(X_{n,1}(d))|\,] < \infty$, we have that $\frac{1}{\sqrt{d}}\overline{S}_{s(d):t}$ is upper bounded. Thus, we obtain directly that:

$$\xi_{d,t} \le M\,,\ \zeta'_{d,t} \le M\ ; \quad |\zeta_{d,t}| + |\zeta'_{d,t}| \le M\,|\tfrac{1}{\sqrt{d}}\,\overline{S}_{s(d):t}|\ .$$

Taking expectations in (43):

$$\mathbb{E}\,[\,e^{\frac{2}{\sqrt{d}}\overline{S}_{s(d):t}}\,] = 1 + \tfrac{2}{d}\,\mathbb{E}\,[\,\overline{S}^2_{s(d):t}\, e^{2\zeta_{d,t}}\,]\ .$$

Now consider the term:

$$a_d(t) := \mathbb{E}\,[\,\overline{S}^2_{s(d):t}\, e^{2\zeta_{d,t}}\,] = \mathbb{E}\,[\,\overline{S}^2_{s(d):t}\,] + \mathbb{E}\,[\,\overline{S}^2_{s(d):t}\,(e^{2\zeta_{d,t}} - 1)\,]\ .$$

Using Holder's inequality and the fact that $\mathbb{E}\,[\,|e^{2\zeta_{d,t}} - 1|^q\,] \le M(q)\,\mathbb{E}\,[\,|\zeta_{d,t}|\,]$ for any $q \ge 1$, via the Lipschitz continuity of $x \mapsto |e^{2x} - 1|^q$ on $(-\infty, M]$, we obtain that for $\epsilon > 0$ as in Proposition C.1(iii):

$$\begin{aligned} |\,\mathbb{E}\,[\,\overline{S}^2_{s(d):t}\,(e^{2\zeta_{d,t}} - 1)\,]\,| &\le \mathbb{E}^{\frac{2}{2+\epsilon}}\,[\,\overline{S}^{2+\epsilon}_{s(d):t}\,]\ \mathbb{E}^{\frac{\epsilon}{2+\epsilon}}\,[\,|e^{2\zeta_{d,t}} - 1|^{\frac{2+\epsilon}{\epsilon}}\,] \\ &\le M\,\mathbb{E}^{\frac{\epsilon}{2+\epsilon}}\,[\,|\zeta_{d,t}|\,] \to_t 0 \end{aligned}$$

the last limit following from Proposition C.1(i). Thus, using also Proposition C.1(ii)-(iii), we have proven that $a_d(t) \to_t \sigma^2_{s:t}$. Note now that:

$$|\big(1 + \tfrac{2}{d}\,a_d(t)\big)^d - \big(1 + \tfrac{2\sigma^2_{s:t}}{d}\big)^d| \le M\,|a_d(t) - \sigma^2_{s:t}|\ ; \quad \big(1 + \tfrac{2\sigma^2_{s:t}}{d}\big)^d \to_t e^{2\sigma^2_{s:t}}\ ,$$

the first result following from the derivative of $x \mapsto \big(1 + \frac{2x}{d}\big)^d$ being bounded for $x \in [0, M]$. Thus we have proven that: $\big(\,\mathbb{E}\,[\,e^{\frac{2}{\sqrt{d}}\,\overline{S}_{s(d):t}}\,]\,\big)^d \to_t e^{2\sigma^2_{s:t}}$. Using similar manipulations and the Taylor expansion (44) we obtain that:

$$\big(\,\mathbb{E}^2\,[\,e^{\frac{1}{\sqrt{d}}\,\overline{S}_{s(d):t}}\,]\,\big)^d \to_t e^{\sigma^2_{s:t}}\ .$$

Taking the ratio, the uniform convergence result in (42) is proved. □

C.2 Results for Theorems 4.1 and 4.2

To prove Theorems 4.1 and 4.2, we will first require some technical lemmas. Here the equally weighted d-dimensional resampled (at the deterministic time instances $t_k(d)$) particles are written with a prime notation; so $X'^{,i}_{l_d(t_k(d)),j}$ will denote the j-th co-ordinate of the i-th particle, immediately *after* the resampling procedure at $t_k(d)$.

Proposition C.2. *Assume (A1(i)(ii)) and let $k \in \{1, \ldots, m^*\}$. Then, there exists an $M(k) < \infty$ such that for any $N \ge 1$, $d \ge 1$, $i \in \{1, \ldots, N\}$, $j \in \{1, \ldots, d\}$:*

$$\mathbb{E}\,[\,V(X'^{,i}_{l_d(t_k(d)),j})\,] \le M(k)N^k.$$

Proof. We will use an inductive proof on the resampling times (assumed to be deterministic). It is first remarked (using Lemma A.1)(iii)) that for every $k \in \{1, \ldots, m^*\}$:

$$\mathbb{E}\,[\,V(X^i_{l_d(t_k(d)),j})\,|\,\mathscr{F}'^{,N}_{t_{k-1}(d)}\,] \le M\,V(x'^{,i}_{l_d(t_{k-1}(d)),j}) \tag{45}$$

where $\mathscr{F}'^{,N}_{t_{k-1}(d)}$ is the filtration generated by the particle system up-to and including the $(k-1)^{th}$ resampling time and $M < \infty$ does not depend upon $t_k(d)$, $t_{k-1}(d)$ or indeed d.

At the first resampling time, we have (averaging over the resampling index) that

$$\mathbb{E}\,[\,V(X'^{,i}_{l_d(t_1(d)),j})\,|\,\mathscr{F}^N_{t_1(d)}\,] = \sum_{i=1}^{N} \overline{w}_{l_d(t_1(d))}(x^i_{l_d(t_0(d)):l_d(t_1(d))-1})\,V(x^i_{l_d(t_1(d)),j})$$

where $\mathscr{F}^N_{t_1(d)}$ is the filtration generated by the particle system up-to the 1st resampling time (but excluding re-sampling) and $\overline{w}_{l_d(t_1(d))}(x^i_{l_d(t_0(d)):l_d(t_1(d))-1})$ is the normalized importance weight. Now, clearly (due to normalised weights be bounded by 1):

$$\mathbb{E}\,[\,V(X'^{,i}_{l_d(t_1(d)),j})\,|\,\mathscr{F}^N_{t_1(d)}\,] \le \sum_{i=1}^{N} V(x^i_{l_d(t_1(d)),j})$$

and, via (45), $\mathbb{E}\,[\,V(X'^{,i}_{l_d(t_1(d),j})\,] \le NM$ which gives the result for the first resampling time.

Using induction, if we assume that the result holds at the $(k-1)^{th}$ time we resample ($k \ge 2$), it follows that (for $\mathscr{F}^N_{t_k(d)}$ being the filtration generated by the particle system up-to the k-th resampling time, but excluding resampling):

$$\begin{aligned}\mathbb{E}\,[\,V(X'^{,i}_{l_d(t_k(d)),j})\,|\,\mathscr{F}^N_{t_k(d)}\,] &= \sum_{i=1}^{N} \overline{w}_{l_d(t_k(d))}(x^i_{l_d(t_{k-1}(d)):l_d(t_k(d))-1})\,V(x^i_{l_d(t_k(d)),j}) \\ &\le \sum_{i=1}^{N} V(x^i_{l_d(t_k(d)),j})\ .\end{aligned}$$

Thus, via (45) and the exchangeability of the particle and dimension index, we obtain that

$$\mathbb{E}\,[\,V(X'^{,i}_{l_d(t_k(d),j})\,] \le NM\,\mathbb{E}\,[\,V(X'^{,i}_{l_d(t_{k-1}(d),j})\,]\ .$$

The proof now follows directly. □

Proposition C.3. *Assume (A1(i)(ii), A2). Let $\varphi \in \mathscr{L}_{V^r}$, $r \in [0, \frac{1}{2})$. Then for any fixed N, any $k \in \{1, \ldots, m^*\}$ and any $i \in \{1, \ldots, N\}$ we have*

$$\frac{1}{d}\sum_{j=1}^{d} \varphi(X'^{,i}_{l_d(t_k(d)),j}) \to \pi_{t_k}(\varphi)\ , \quad \textit{in } \mathbb{L}_1\ .$$

Proof. We distinct between two cases: $k = 1$ and $k > 1$. When $k = 1$, due to the boundedness of the normalised weights and the exchangeability of the particle indices we have that:

$$\mathbb{E}\,\Big|\,\frac{1}{d}\sum_{j=1}^{d} \varphi(X'^{,i}_{l_d(t_1(d)),j}) - \pi_{t_1}(\varphi)\,\Big| \le N\,\mathbb{E}\,\Big|\,\frac{1}{d}\sum_{j=1}^{d} \varphi(X^{i}_{l_d(t_1(d)),j}) - \pi_{t_1}(\varphi)\,\Big| \tag{46}$$

Adding and subtracting the term $\mathbb{E}\,[\,\varphi(X^i_{l_d(t_1(d)),j})\,]$ we obtain that the expectation on the R.H.S. of the above equation is bounded by:

$$\mathbb{E}\,\Big|\,\frac{1}{d}\sum_{j=1}^{d} \varphi(X^{i}_{l_d(t_1(d)),j}) - \mathbb{E}\,[\,\varphi(X^i_{l_d(t_1(d)),j})\,]\,\Big| + |\,\mathbb{E}\,[\,\varphi(X^i_{l_d(t_1(d)),j})] - \pi_{t_1}(\varphi)\,|\ . \tag{47}$$

For the first term, due to the independency across dimension, considering second moments we get the upper bound:

$$\frac{1}{\sqrt{d}}\,\mathbb{E}^{1/2}\,\Big[\,\big(\,\varphi(X^i_{l_d(t_1(d)),j}) - \mathbb{E}\,[\,\varphi(X^i_{l_d(t_1(d)),j})\,]\,\big)^2\,\Big]\ .$$

As $\varphi \in \mathscr{L}_{V^r}$ with $r \le 1/2$ the argument of the expectation is upper-bounded by $MV(X^i_{l_d(t_1(d)),j})$ whose expectation is controlled via Lemma A.1(iii). Thus the above quantity is $\mathcal{O}(d^{-1/2})$. For the second term in (47) we can use directly Proposition A.1 (for time sequences required there selected as $s(d) \equiv \phi_0$ and $t(d) \equiv t_1(d)$) to show also that this term will vanish in the limit $d \to \infty$.

The general case with $k > 1$ is similar, but requires some additional arguments as resampling eliminates the i.i.d. property. Again, integrating out the resampling index as in (46) we are left with the quantity:

$$\mathbb{E}\,\Big|\,\frac{1}{d}\sum_{j=1}^{d}\varphi(X^i_{l_d(t_k(d)),j}) - \pi_{t_k}(\varphi)\,\Big|\ .$$

Adding and subtracting $\frac{1}{d}\sum_{j=1}^{d}\mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}[\,\varphi(X^i_{l_d(t_k(d)),j})\,]$ within the expectation, the above quantity is upper bounded by:

$$\mathbb{E}\,\Big|\,\frac{1}{d}\sum_{j=1}^{d}\varphi(X^i_{l_d(t_k(d)),j}) - \frac{1}{d}\sum_{j=1}^{d}\mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}\,[\,\varphi(X^i_{l_d(t_k(d)),j})\,]\,\Big| \quad + \\ \mathbb{E}\,\Big|\,\frac{1}{d}\sum_{j=1}^{d}\mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}\,[\,\varphi(X^i_{l_d(t_k(d)),j})\,] - \pi_{t_k}(\varphi)\,\Big|\ . \tag{48}$$

For the first of these two terms, due to *conditional* independency across dimension and exchangeability in the dimensionality index j, looking at the second moment we obtain the upper bound:

$$\frac{1}{\sqrt{d}}\mathbb{E}^{1/2}\big[\,\big(\varphi(X^i_{l_d(t_k(d)),j}) - \mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}\,[\,\varphi(X^i_{l_d(t_k(d)),j})\,]\,\big)^2\,\big]\ .$$

Since $|\varphi(x)| \le M\,V^r(x)$ with $r \le \frac{1}{2}$, the variable in the expectation above is upper bounded by $M(V(X'^{,i}_{l_d(t_k(d)),j}) + V(X'^{,i}_{l_d(t_{k-1}(d)),j}))$ which due to Proposition C.2 is bounded in expectation by some $M(N,k)$. Thus, the first term in (48) is $\mathcal{O}(d^{-1/2})$. The second term in (48) now, due to exchangeability over j, is upper bounded by $\mathbb{E}\,\big|\,\mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}\,[\,\varphi(X^i_{l_d(t_k(d)),j})\,] - \pi_{t_k}(\varphi)\,\big|$, which again due to Proposition A.1 vanishes in the limit $d \to \infty$. ☐

For the Markov chain $X^i_{n,j}$ considered on the instances $n_1 \le n \le n_2$ we will henceforth use the notation $\mathbb{E}_{\pi_s}[\,g(X^i_{n,j})\,]$ to specify that we impose the initial distribution $X^i_{n_1,j} \sim \pi_s$.

Proposition C.4. *Assume (A1-2) and that $g \in \mathscr{L}_{V^r}$ with $r \in [0, \frac{1}{2})$. For $k \in \{1,\dots,m^*\}$, $i \in \{1,\dots,N\}$ and a sequence $s_k(d)$ with $s_k(d) > t_{k-1}(d)$ and $s_k(d) \to s_k > t_{k-1}$ we define:*

$$E_{i,j} = \sum_n \big\{\,\mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}\,[\,g(X^i_{n,j})\,] - \mathbb{E}_{\pi_{t_{k-1}}}\big[\,g(X^i_{n,j})\,\big]\,\big\}\ , \quad 1 \le j \le d\ ,$$

for subscript n in the range $l_d(t_{k-1}(d)) \le n \le l_d(s_k(d)) - 1$. Then, we have that:

$$\frac{1}{d}\sum_{j=1}^{d} E_{i,j} \to 0\ , \quad \textit{in } \mathbb{L}_1\ .$$

Proof. We will make use of the Poisson equation and employ the decomposition (35) used in the proof of Theorem A.1. In particular, a straight-forward calculation gives that:

$$R_{i,j} = \sum_{n=n_1+1}^{n_2}\big\{\,(\,\mathbb{E}_{X_{n_1,j}} - \mathbb{E}_{\pi_{t_{k-1}}}\,)[\,\widehat{g}_n(X^i_{n-1,j}) - \widehat{g}_{n-1}(X^i_{n-1,j})\,]\,\big\} \\ + (\,\mathbb{E}_{X_{n_1,j}} - \mathbb{E}_{\pi_{t_{k-1}}}\,)[\,g(X_{n_2,j}) - \widehat{g}_{n_2}(X_{n_2,j})\,] + \widehat{g}_{n_1}(X_{n_1,j}) - \pi_{t_{k-1}}(\widehat{g}_{n_1})\ , \tag{49}$$

where $\widehat{g}_n = \mathcal{P}(g, k_n, \pi_n)$, and we have set:

$$n_1 = l_d(t_{k-1}(d))\ ; \quad n_2 = l_d(s_k(d)) - 1\ ;\ X_{n_1,j} \equiv X'^{,i}_{l_d(t_{k-1}(d)),j}\ .$$

It is remarked that the martingale term in the original expansion (35) has expectation 0, so is not involved in our manipulations. We will first deal with the sum in the first line of (49), that is (when taking into account the averaging over j) with:

$$A_d := \frac{1}{d}\sum_{j=1}^{d}\sum_{n=n_1+1}^{n_2}[\,\delta_{X_{n_1,j}} - \pi_{t_{k-1}}\,]\big(\,(k_{n_1+1:n})[\widehat{g}_n - \widehat{g}_{n-1}]\,\big)\ .$$

Now each summand in the above double sum is upper bounded by

$$\frac{M}{d}\,\|\,[\delta_{X_{n_1,j}} - \pi_{t_{k-1}}](k_{n_1+1:n})\,\|_{V^r}\ .$$

To bound this V^r-norm one can apply Theorem 8 of [27]; here, under (A1-2) we have that either:

$$\|\,[\delta_{X_{n_1,j}} - \pi_{t_{k-1}}](k_{n_1+1:n})\,\|_{V^r} \le M\rho^{n-n_1}V(X_{n_1,j})^r + M'\zeta^{n-n_1} \tag{50}$$

for some $\rho,\zeta\in(0,1)$, $0<M,M'<\infty$, when $B_{j-1,n}$ (of that paper) is 1. Or, if $B_{j-1,n}>1$, one has the bound

$$\|\,[\delta_{X_{n_1,j}} - \pi_{t_{k-1}}](k_{n_1+1:n})\,\|_{V^r} \le M\rho^{\lfloor j^*(n-n_1)\rfloor}V(X_{n_1,j})^r + M'\zeta^{\lfloor j^*(n-n_1)\rfloor}$$

with j^* as the final equation of [27, pp. 1650]. (Note that this follows from a uniform in time drift condition which follows from Proposition 4 of [27] (via (A1))). By summing up first over n and then over j (and dividing with d), using also Proposition C.3 along the way to control $\sum_j V(X_{n_1,j})^r/d$, we have that:

$$A_d \to 0\ ,\quad \text{in } \mathbb{L}_1\ .$$

A similar use of the bound in (50) and Proposition C.3 can give directly that the second term in (49) will vanish in the limit when summing up over j and dividing with d. Finally, for the last term in (49): Proposition C.3 is not directly applicable here as one has to address the fact that the function $\widehat{g}_{n_1}$ depends on d. Using Lemma A.1 (ii), one can replace $\widehat{g}_{n_1}\equiv\widehat{g}_{l_d(t_{k-1}(d))}$ by $\widehat{g}_{t_{k-1}}$ and then apply Proposition C.3 and the fact that $t_{k-1}(d)\to t_{k-1}$ to show that the remainder term goes to zero in $\mathbb{L}_1$ (when averaging over j). The proof is now complete.

□

Proof of Theorem 4.1. Recall the definition of the ESS:

$$\mathrm{ESS}_{(t_{k-1}(d),s_k(d))}(N) = \frac{\big(\sum_{i=1}^N e^{(1-\phi_0)a^i(d)}\big)^2}{\sum_{i=1}^N e^{2(1-\phi_0)a^i(d)}}\ .$$

where we have defined:

$$a^i(d) = \frac{1}{d}\sum_{j=1}^{d}\{\overline{G}_{i,j} + E_{i,j}\}$$

with:

$$\begin{aligned}
\overline{G}_{i,j} &= \sum_n \big\{\, g(X^i_{n,j}) - \mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}[\,g(X^i_{n,j})\,]\,\big\}\ ;\\
E_{i,j} &= \sum_n \big\{\, \mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}[\,g(X^i_{n,j})\,] - \mathbb{E}_{\pi_{t_{k-1}}}[\,g(X^i_{n,j})\,]\,\big\}\ ,
\end{aligned}$$

for subscript n in the range $l_d(t_{k-1}(d))\le n\le l_d(s_k(d))-1$. From Proposition C.4 we get directly that $\sum_{j=1}^d E_{i,j}/d\to 0$ (in $\mathbb{L}_1$). Thus, we are left with $\overline{G}_{i,j}$ which corresponds to a martingale under the filtration we define below. In the below proof, we consider the weak convergence for a single particle. However, it possible to prove a multivariate CLT for all the particles using the Cramer-Wold device. This calculation is very similar to that given below and is hence omitted.

Consider some chosen particle i, with $1\le i\le N$. For any $d\ge 1$ we define the filtration $\mathcal{G}_{0,d}\subseteq\mathcal{G}_{1,d}\subseteq\cdots\subseteq\mathcal{G}_{d,d}$ as follows:

$$\begin{aligned}
\mathcal{G}_{0,d} &= \sigma(X'^{,l}_{l_d(t_{k-1}(d)),j}, 1\le j\le d,\, 1\le l\le N)\ ;\\
\mathcal{G}_{j,d} &= \mathcal{G}_{j-1,d}\bigvee\sigma(X^i_{n,j},\, l_d(t_{k-1}(d))\le n\le l_d(s_k(d))-1)\ ,\quad j\ge 1\ .
\end{aligned}\tag{51}$$

That is, σ-algebra $\mathcal{G}_{0,d}$ contains the information about *all* particles, along *all* d co-ordinates until (and including) the resampling step; then the rest of the filtration is build up by adding information for the subsequent trajectory of the various co-ordinates. Critically, conditionally on $\mathcal{G}_{0,d}$ these trajectories are independent. One can now easily check that

$$\beta^i_j(d) = \frac{1}{d}\sum_{k=1}^{j}\overline{G}_{i,k}\ ,\quad 1\le j\le d\ ,$$

is a martingale w.r.t. the filtration in (51). Now, to apply the CLT for triangular martingale arrays, we will show that for every $i\in\{1,\ldots,N\}$:

a) That in $\mathbb{L}_1$:

$$\lim_{d\to\infty} \frac{1}{d^2}\sum_{j=1}^{d} \mathbb{E}\,[\,\overline{G}_{i,j}^2\,|\,\mathcal{G}_{j-1,d}\,] = \sigma^2_{t_{k-1}:s_k}$$

b) For any $\epsilon>0$, that in $\mathbb{L}_1$:

$$\lim_{d\to\infty} \frac{1}{d^2}\sum_{j=1}^{d} \mathbb{E}\,[\,\overline{G}_{i,j}^2\mathbb{I}_{|\overline{G}_{i,j}|\geq \epsilon d}\,|\,\mathcal{G}_{j-1,d}\,] = 0\ .$$

This will allow us to show that $(1-\phi_0)a^i(d)$ will converge weakly to the appropriate normal random variable. Notice, that due to the conditional independency mentioned above and the definition of the filtration in (51) we in fact have that:

$$\mathbb{E}\,[\,\overline{G}_{i,j}^2\,|\,\mathcal{G}_{j-1,d}\,] \equiv \mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}\,[\,\overline{G}_{i,j}^2\,]\ ;$$
$$\mathbb{E}\,[\,\overline{G}_{i,j}^2\mathbb{I}_{|\overline{G}_{i,j}|\geq \epsilon d}\,|\,\mathcal{G}_{j-1,d}\,] \equiv \mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}\,[\,\overline{G}_{i,j}^2\mathbb{I}_{|\overline{G}_{i,j}|\geq \epsilon d}\,]\ .$$

We make the following definition:

$$G_{i,j} = \sum_{n}\big\{g(X^i_{n,j})-\pi_n(g)\big\} \equiv M_{n_1:n_2,i,j}+R_{n_1:n_2,i,j}\ ,$$

(for convenience we have set $n_1 = l_d(t_{k-1}(d))$ and $n_2 = l_d(s_k(d))-1$) with the terms $M_{n_1:n_2,i,j}$ and $R_{n_1:n_2,i,j}$ defined as in Theorem A.1 with the extra subscripts indicating the number of particle and the co-ordinate. Notice that $\overline{G}_{i,j} = G_{i,j}-\mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}[\,G_{i,j}\,]$.

We start with a). We first use the fact that:

$$\frac{1}{d^2}\sum_{j=1}^{d}\mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}\,[\,\overline{G}_{i,j}^2\,] - \frac{1}{d^2}\sum_{j=1}^{d}\mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}\,[\,G_{i,j}^2\,] \to 0\ ,\quad \text{in } \mathbb{L}_1\ .$$

To see that, simply note that the above difference is equal to:

$$\frac{1}{d^2}\sum_{j=1}^{d}\mathbb{E}^2_{X'^{,i}_{l_d(t_{k-1}(d)),j}}\,[\,G_{i,j}\,] \equiv \frac{1}{d^2}\sum_{j=1}^{d}\mathbb{E}^2_{X'^{,i}_{l_d(t_{k-1}(d)),j}}\,[\,R_{i,j}\,] \leq \frac{1}{d^2}\sum_{j=1}^{d} V(X'^{,i}_{l_d(t_{k-1}(d)),j})^{2r}$$

where we first used the fact that $M_{n_1:n_2,i,j}$ is a martingale (thus, of zero expectation) and then Theorem A.1 to obtain the bound; the bounding term vanishes due to Proposition C.2. We then have that:

$$\begin{aligned}\frac{1}{d^2}\sum_{j=1}^{d}\mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}\,[\,G_{i,j}^2\,] &= \frac{1}{d^2}\sum_{j=1}^{d}\mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}\,[\,M_{i,j}^2+R_{i,j}^2+2\,M_{i,j}R_{i,j}\,] \\ &= \frac{1}{d^2}\sum_{j=1}^{d}\mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}\,[\,M_{i,j}^2\,] + \mathcal{O}(d^{-1/2})\ . \end{aligned}\tag{52}$$

To yield the $\mathcal{O}(d^{-1/2})$ one can use the bound

$$\mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}\,[\,R_{i,j}^2\,] \leq M\,V(X'^{,i}_{l_d(t_{k-1}(d)),j})^{2r}$$

from Theorem A.1, and then (using Cauchy-Schwärtz and Theorem A.1):

$$\begin{aligned}|\,\mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}\,[\,M_{i,j}R_{i,j}\,]\,| &\leq \mathbb{E}^{1/2}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}\,[\,M_{i,j}^2\,]\cdot\mathbb{E}^{1/2}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}\,[\,R_{i,j}^2\,] \\ &\leq M\,\sqrt{d}\,V(X'^{,i}_{l_d(t_{k-1}(d)),j})^{2r}\ .\end{aligned}$$

One then only needs to make use of Proposition C.2 to get (52). Now, using the analytical definition of $M_{i,j}$ from Theorem A.1 we have:

$$
\begin{aligned}
\frac{1}{d^2}\sum_{j=1}^{d}\mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}[\,M_{i,j}^2\,] &= \frac{1}{d^2}\sum_{j=1}^{d}\sum_{n=n_1+1}^{n_2}\big\{\,\mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}[\,\widehat{g}_n^2(X^i_{n,j}) - k_n^2(\widehat{g}_n)(X^i_{n-1,j})\,]\,\big\} \\
&= \frac{1}{d^2}\sum_{j=1}^{d}\sum_{n=n_1}^{n_2-1}\mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}[\,\varphi_{n+1}(X^i_{n,j})\,] =: A_d
\end{aligned}
\tag{53}
$$

where:

$$
\varphi_n = k_n(\widehat{g}_n^2) - [k_n(\widehat{g}_n)]^2\ ;\quad \widehat{g}_n = \mathcal{P}(g,k_n,\pi_n)\ .
$$

Using again the decomposition in Theorem A.1, but now for φ_n as above (which due to Proposition B.1 satisfies the requirements of Theorem A.1), we get that:

$$
\begin{aligned}
\big|\,\mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}\big[\sum_{n=n_1}^{n_2-1}\varphi_{n+1}(X^i_{n,j}) - \pi_n(\varphi_{n+1})\big]\,\big| &= \big|\,\mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}[\,R'_{n_1:(n_2-1),i,j}\,]\,\big| \\
&\le M\,V^{2r}(X'^{,i}_{l_d(t_{k-1}(d)),j})\ .
\end{aligned}
$$

Thus, continuing from (53), and using the above bound and Proposition C.2, we have:

$$
\big|\,A_d - \frac{1}{d}\sum_{n=n_1}^{n_2-1}\pi_n(\varphi_{n+1})\,\big| = \mathcal{O}(d^{-1})\ .
\tag{54}
$$

The proof for a) is completed using to the deterministic limit:

$$
\frac{1-\phi_0}{d}\sum_{n=n_1}^{n_2-1}\pi_n(\varphi_{n+1}) \to \int_{t_{k-1}}^{s_k}\pi_u(\widehat{g}_u^2 - k_u(\widehat{g}_u)^2)du\ .
$$

For b), we choose some δ so that $r(2+\delta)\le 1$, and obtain the following bound:

$$
\begin{aligned}
\mathbb{E}_{\,X'^{,i}_{l_d(t_{k-1}(d)),j}}[\,\overline{G}_{i,j}^{2+\delta}\,] &\le M\,\mathbb{E}_{\,X'^{,i}_{l_d(t_{k-1}(d)),j}}[\,G_{i,j}^{2+\delta}\,] \\
&\le M\,\mathbb{E}_{\,X'^{,i}_{l_d(t_{k-1}(d)),j}}[\,M_{i,j}^{2+\delta} + R_{i,j}^{2+\delta}\,] \\
&\le MV(X'^{,i}_{l_d(t_{k-1}(d)),j})^{r(2+\delta)}\,d^{1+\frac{\delta}{2}}\ ,
\end{aligned}
$$

where for the last inequality we used the growth bounds in Theorem A.1. Also using, first, Holder's inequality, then, Markov inequality and, finally, the above bound we find that:

$$
\begin{aligned}
\mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}[\,\overline{G}_{i,j}^2\mathbb{I}_{|\overline{G}_{i,j}|\ge\epsilon d}\,] &\le \big(\mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}[\,\overline{G}_{i,j}^{2+\delta}\,]\,\big)^{\frac{2}{2+\delta}}\cdot\big(\mathbb{P}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}[\,|\overline{G}_{i,j}|^{2+\delta}\ge(\epsilon d)^{2+\delta}\,]\,\big)^{\frac{\delta}{2+\delta}} \\
&\le M\,V(X'^{,i}_{l_d(t_{k-1}(d)),j})^{2r}\,d\,\cdot\,\frac{V(X'^{,i}_{l_d(t_{k-1}(d)),j})^{r\delta}d^{\delta/2}}{(\epsilon\,d)^{\delta}}\ .
\end{aligned}
$$

Thus, we also have:

$$
\frac{1}{d^2}\sum_{j=1}^{d}\mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}[\,\overline{G}_{i,j}^2\mathbb{I}_{|\overline{G}_{i,j}|\ge\epsilon d}\,] \le M\,d^{-\delta/2}\,\frac{1}{d}\sum_{j=1}^{d}V(X'^{,i}_{l_d(t_{k-1}(d)),j})^{r(2+\delta)}\ .
$$

Due to Proposition C.2, this bound proves part b).

□

Proof of Theorem 4.2. The proof is similar to that of Theorem 4.1 (as the final resampling time is strictly less than 1) and Theorem 3.3; it is omitted for brevity. □

C.3 Stochastic Times

Proof of Theorem 4.3. Our proof will keep d fixed until the point at which we can apply Theorem 4.1. Conditionally on the chosen $\{a_k\}$ we have (we use the convention $t^{\delta}_{m^*(\delta)+1}(d)=1$):

$$\mathbb{P}[\,\Omega\setminus\Omega_d^N\,] \le \sum_{k=1}^{m^*(\delta)+1}\;\sum_{s\in G_\delta\cap(\,t^\delta_{k-1}(d),t^\delta_k(d)\,]} \mathbb{P}\big[\,\big|\tfrac{1}{N}\,\mathrm{ESS}_{(t^\delta_{k-1}(d),s)}(N)-\mathrm{ESS}_{(t^\delta_{k-1}(d),s)}\big|\ge \upsilon\big|\mathrm{ESS}_{(t^\delta_{k-1}(d),s)}-a_k\big|\,\big]\ .$$

We define:

$$\epsilon(d) := \inf_{k\in\{1,2,\dots,m^*(\delta)+1\}}\;\inf_{s\in G_\delta\cap(\,t^\delta_{k-1}(d),t^\delta_k(d)\,]}|\,\mathrm{ESS}_{(t^\delta_{k-1}(d),s)}-a_k\,|\ .$$

Thus, we have:

$$\mathbb{P}[\,\Omega\setminus\Omega_d^N\,] \le \sum_{k=1}^{m^*(\delta)}\;\sum_{s\in G_\delta\cap(\,t^\delta_{k-1}(d),t^\delta_k(d)\,]} \mathbb{P}\big[\,\big|\tfrac{1}{N}\,\mathrm{ESS}_{(t^\delta_{k-1}(d),s)}(N)-\mathrm{ESS}_{(t^\delta_{k-1}(d),s)}\big|\ge \upsilon\epsilon(d)\big|\,\big]\ .$$

Application of the Markov inequality yields that:

$$\mathbb{P}[\,\Omega\setminus\Omega_d^N\,] \le \frac{\delta}{\upsilon\epsilon(d)}\max_{k,s}\mathbb{E}\big[\,\big|\tfrac{1}{N}\,\mathrm{ESS}_{(t^\delta_{k-1}(d),s)}(N)-\mathrm{ESS}_{(t^\delta_{k-1}(d),s)}\big|\,\big]\ . \tag{55}$$

Note that $\lim_{d\to\infty}\epsilon(d)=\epsilon>0$ (with probability one); this is due to the uniform convergence of $\mathrm{ESS}_{(t^\delta_{k-1}(d),s)}$ to $\exp\{-\sigma^2_{t^\delta_{k-1}:s}\}$ as $d\to\infty$, see the proof of Proposition 4.1. Thus, it remains to bound the expectation on the R.H.S. of (55) (and it's maximum over k,s).

Application of Theorem 4.1 now yields:

$$\lim_{d\to\infty}\mathbb{E}\big[\,\big|\tfrac{1}{N}\mathrm{ESS}_{(t^\delta_{k-1}(d)),s)}(N)-\mathrm{ESS}_{(t^\delta_{k-1}(d),s)}\,\big|\,\big] = \mathbb{E}\big[\,\big|\tfrac{1}{N}\,\mathrm{ESS}_{(t^\delta_{k-1},s)}(N)-\mathrm{ESS}_{(t^\delta_{k-1},s)}\big|\,\big]$$

where:

$$\mathrm{ESS}_{(t^\delta_{k-1},s)}(N) = \frac{(\sum_{j=1}^N\exp\{X_j^k\})^2}{\sum_{j=1}^N\exp\{2X_j^k\}}\ ;\quad \mathrm{ESS}_{(t^\delta_{k-1},s)} = \exp\big\{-\sigma^2_{t^\delta_{k-1}:s}\big\}\ ,$$

with $X_j^k \overset{\text{i.i.d.}}{\sim}\mathcal{N}(0,\sigma^2_{t^\delta_{k-1}:s})$. We set:

$$\alpha_j^k=\exp\{X_j^k\}\ ;\quad \beta_j^k=\exp\{2X_j^k\}\ ;\quad \alpha^k=\exp\{\tfrac{1}{2}\,\sigma^2_{t^\delta_{k-1}:s}\}\ ;\quad \beta^k=\exp\{2\sigma^2_{t^\delta_{k-1}:s}\}\ .$$

Then, we are to bound:

$$\mathbb{E}\Big[\Big|\frac{(\frac{1}{N}\sum_{j=1}^N\alpha_j^k)^2}{\frac{1}{N}\sum_{j=1}^N\beta_j^k}-\frac{(\alpha^k)^2}{\beta^k}\Big|\Big]\ .$$

We have the decomposition

$$\frac{(\frac{1}{N}\sum_{j=1}^N\alpha_j^k)^2}{\frac{1}{N}\sum_{j=1}^N\beta_j^k}-\frac{(\alpha^k)^2}{\beta^k} = \Big(\frac{(\frac{1}{N}\sum_{j=1}^N\alpha_j^k)^2}{\beta^k\frac{1}{N}\sum_{j=1}^N\beta_j^k}\Big)\Big[\beta^k-\frac{1}{N}\sum_{j=1}^N\beta_j^k\Big]+\frac{1}{\beta^k}\Big[(\frac{1}{N}\sum_{j=1}^N\alpha_j^k)^2-(\alpha^k)^2\Big]\ .$$

For the first term of the R.H.S. in the above equation, as ESS divided by N is upper-bounded by 1, we can use Jensen and the Marcinkiewicz-Zygmund inequality. For the second term, via the relation $x^2-y^2=(x+y)(x-y)$ and Cauchy-Schwärtz, one can use the same inequality to conclude that for some finite $M(k,\delta,s)$:

$$\mathbb{E}\big[\,\big|\tfrac{1}{N}\,\mathrm{ESS}_{(t^\delta_{k-1},s)}(N)-\mathrm{ESS}_{(t^\delta_{k-1},s)}\big|\,\big]\le\frac{M(k,\delta,s)}{\sqrt{N}}\ .$$

Returning to (55) we have thus proven that: $\lim_{d\to\infty}\mathbb{P}[\,\Omega\setminus\Omega_d^N\,]\le\frac{M(\delta)}{\sqrt{N}}$ as required. ☐

References

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

[2] Andrieu, C., Jasra, A., Doucet, A. & Del Moral P. (2011). On non-linear Markov chain Monte Carlo. *Bernoulli*, **17**, 987–1014.

[3] Atchadé, Y. (2009). A strong law of large numbers for martingale arrays. Technical Report, University of Michigan.

[4] Baehr, C, & Pannekoucke, O. (2010). Some issues and results on the EnKF and particle filters for meteorological models. In *Chaotic Systems: Theory and Applications*, C. H. Skiadas & I. Dimotikalis Eds, 27–24, World Scientific: London.

[5] Bédard, M. (2007). Weak convergence of Metropolis algorithms for non-i.i.d. target distributions. *Ann. Appl. Prob.*, **17**, 1222–1244.

[6] Berger, E. (1986). Asymptotic behaviour of a class of stochastic approximation procedures. *Probab. Theory. Rel. Fields*, **71**, 517–552.

[7] 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.

[8] Beskos, A., Crisan, D., & Jasra, A. (2011). On the stability of sequential Monte Carlo methods in high-dimensions. Technical Report, Imperial College London.

[9] Beskos, A., Crisan, D., Jasra, A. & Whiteley, N. (2011). Error bounds and normalizing constants for sequential Monte carlo. Technical Report, Imperial College London.

[10] Beskos, A., Pillai, N., Roberts, G. O., Sanz-Serna, J. M. & Stuart A. (2013). Optimal tuning of the hybrid Monte Carlo algorithm. *Bernoulli*, (to appear).

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

[12] Beskos, A. & Stuart, A. (2009). MCMC methods for sampling function space. In *Proceedings of the International Congress of Industrial and Applied Mathematics*, Zurich, 2007.

[13] 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.

[14] Billingsley, P. (1999).*Convergence of Probability Measures.* 2nd edition. Wiley:New York.

[15] Breyer, L., Piccioni, M. & Scarlatti, S. (2004). Optimal scaling of MALA for non-linear regression. *Ann. Appl. Probab.*, **14**, 1479–1505.

[16] Breyer, L. & Roberts, G. (2000). From Metropolis to diffusions: Gibbs states and optimal scaling. *Stochastic Process. Appl.*, **90**, 181–206.

[17] Cappé, O, Gullin, A., Marin, J. M. & Robert, C. P. (2004). Population Monte Carlo. *J. Comp. Graph. Stat.*, **13**, 907–925.

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

[19] Chopin, N. (2004). Central limit theorem and its application to Bayesian inference. *Ann Statist.*, **32**, 2385–2411.

[20] Crisan, D. & Rozovsky, B. (2011). *The Oxford Handbook of Nonlinear Filtering* OUP:Oxford.

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

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

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

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

[25] Del Moral, P., Patras, F. & Rubenthaler, S. (2009). Coalescent tree based functional representations for some Feynman-Kac particle models. *Ann. Appl. Probab.*, **19**, 778–825.

[26] Douc, R. & Moulines, É. (2008). Limit theorems for weighted samples with applications to sequential Monte Carlo methods. *Ann. Statist.*, **36**, 2344–2376.

[27] Douc, R., Moulines, É. & Rosenthal, J. S. (2004). Quantitative bounds on convergence of time-inhomogeneous Markov Chains *Ann. Appl. Probab.*, **14**, 1643–1665.

[28] Doucet, A., De Freitas, J. F. G. & Gordon, N. J. (2001). *Sequential Monte Carlo Methods in Practice.* Springer: New York.

[29] Hall, P. & Heyde, C. (1980). *Martingale Limit Theory and its Application*. Academic Press: New York.

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

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

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

[33] Jasra, A., Stephens, D. A., & Holmes, C. C. (2007). On population-based simulation for static inference. *Statist. Comp.*, **17**, 263–269 .

[34] Kong, A., Liu, J. S. & Wong, W. H. (1994). Sequential imputations and Bayesian missing data problems. *J. Amer. Statist. Assoc.*, **89**, 278–288.

[35] Künsch, H. (2005). Recursive Monte Carlo filters: Algorithms and theoretical analysis. *Ann. Statist.*, **33**, 1983–2021.

[36] Lee, A., Yau, C., Giles, M., Doucet, A. & Holmes, C. C. (2010). On the utility of graphics cards to perform massively parallel implementation of advanced Monte Carlo methods. *J. Comp. Graph. Statist.*, **19**, 769–789.

[37] Liu, J. S. (2001). *Monte Carlo Strategies in Scientific Computing*. Springer: New York.

[38] Mattingly, J., Pillai, N. & Stuart, A. (2012). Diffusion limits of the Random walk Metropolis algorithm in High Dimensions. *Ann. Appl. Probab.*, **22**, 881–930.

[39] McLeish, D. L. (1974). Dependent central limit theorems and invariance principles. *Ann. Probab*, **2**, 620–628.

[40] Meyn, S. & Tweedie, R. L. (2009). *Markov Chains and Stochastic Stability*. 2nd edition, Cambridge: CUP.

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

[42] Pillai, N., Stuart, A. & Thiery, A. (2012). Optimal scaling and diffusion limits for the Langevin algorithm in high dimensions. *Ann. Appl. Probab.*, **22**, 2320–2356.

[43] Roberts, G. O., Gelman, A. & Gilks W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.*, **7**, 110–120.

[44] Roberts, G.O. & Rosenthal, J.S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Statist. Soc. Ser. B*, **60**, 255-268.

[45] Rudin, W. (1976). *Principles of Mathematical Analysis*. International Series in Pure and Applied Mathematics, McGraw-Hill Book Co.: New York.

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

[47] Shiryaev, A. (1996). *Probability*, Springer: New York.

[48] Snyder, C., Bengtsson, T., Bickel, P., & Anderson, J. (2008). Obstacles to high-dimensional particle filtering. *Month. Weather Rev.*, **136**, 4629–4640.

[49] Whiteley, N. (2012). Sequential Monte Carlo samplers: Error bounds and insensitivity to initial conditions. *Stoch. Anal. Appl.*, **30**, 774–798.

[50] Whiteley, N. (2013). Stability properties of some particle filters. *Ann. Appl. Probab.* (to appear).

[51] Whiteley, N., Kantas, N, & Jasra, A. (2012). Linear variance bounds for particle approximations of time homogeneous Feynman-Kac formulae. *Stoch. Proc. Appl.*, **122**, 1840–1865.

[52] Withers, C. (1981). Central limit theorems for dependent variables I. *Z. Wahrsch. Verw. Gebiete*, **57**, 509–534.