On the Stability of Sequential Monte Carlo Methods in High Dimensions

ALEXANDROS BESKOS[1], DAN CRISAN[2], AND AJAY JASRA[2]

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

E-Mail: alex@stats.ucl.ac.uk

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

E-Mail: d.crisan@ic.ac.uk, a.jasra@ic.ac.uk

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, e.g. [9, 11, 49], that, using a *single* importance sampling step, one produces an approximation for the target distribution 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' target density and moving to the one of interest and using an SMC method to sample from the sequence (see e.g. [15, 23, 34, 43]). 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 $d \to \infty$, such that $1 < \varepsilon_N < N$. The convergence is achieved with a computational cost proportional to Nd^2. If ε_N is reasonably close to N no additional work is needed. However, 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, the ESS of the system of samples converges to a random variable $\varepsilon_{N,m}$ as $d \to \infty$ and $\lim_{m\to\infty} \varepsilon_{N,m} = N$. In addition, we show that the corresponding 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.

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 used to approximate a sequence of distributions whose densities are known point-wise up to a normalizing constant and are 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 [29] and Chapter VIII in [20] for an overview.

SMC methods combine importance sampling and resampling to approximate distributions. The idea is to introduce a sequence of proposal densities and to 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 the proposal and the target distribution via importance weights. In almost all cases of practical interest, the variance of these importance weights increases as the algorithmic time increases (e.g. [37]); this can, to some extent, be dealt with via a resampling procedure. This consists of sampling with replacement from the current samples using the weights (normalized to sum to one) and resetting the weights to $1/N$. The variability of these weights is often measured by the effective sample size ([40]) and one often resamples when this number drops below a given threshold (dynamic-resampling).

There are a wide variety of convergence results for SMC methods. Most of the results are 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 for example [16, 22, 26, 27, 32, 38]) they differ from the scenario in this paper. At the same time, there is a vast literature on the performance of high-dimensional Markov chain Monte Carlo (MCMC) algorithms e.g. [10, 45]; our aim is to obtain a similar analytical understanding about the effect of dimension on SMC methods. The following publications have considered some problems in this direction: [6, 9, 11, 49]. In [9, 11, 49] 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 [49]); the algorithm considered is standard importance sampling. Empirically, [6] shows that when one introduces an accept/reject step (see [22]) then particle filtering methods can perform well. We discuss these results below.

1.1 Contribution of the Article

In this 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. [36]). 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 [15, 23, 34, 43]. The methodology investigated here is applied in many real contexts: financial modelling [35], regression [47] and approximate Bayesian inference [24]. The question we ask is whether such algorithms, as the dimension d of the distributions increases, are stable in any sense. More precisely, we seek to quantify the computational cost of the algorithm with respect to d and investigate the number of resampling steps required.

It is natural for high dimensional problems to be more cumbersome than their low dimensional counterparts. This can be illustrated via the following simple example: suppose we consider two multivariate normal distributions Π_1 and Π_2 in $\mathbb{R}^d$, where the first distribution Π_1 denotes $\mathcal{N}((0,\dots,0),I_d)$ (mean $(0,\dots,0)$ and covariance the $d-$dimensional identity matrix I_d) and Π_2 denotes $\mathcal{N}((1,\dots,1),I_d)$. Then the total variation distance between the two distributions is equal to $2\,\mathbb{P}\,[\,|X|\le d/2\,]$, where X is a standard normal (one-dimensional) random variable. Hence, as the dimension d increases, the two measures get further and further apart, becoming singular w.r.t. each other exponentially fast (the total variation distance between singular probability measures is 2). Therefore, as d increases, it becomes increasingly harder to use standard importance sampling, to construct a sample from Π_2 by using a proposal from Π_1. This fact has been empirically verified in [11].

Translated to a one-dimensional context the problem of 'moving' from Π_1 to Π_2 is equivalent to that of moving from a standard normal distribution $\mathcal{N}(0,1)$ to a normal distribution $\mathcal{N}(d,1)$ (the total variation distance between $\mathcal{N}(0,1)$ and $\mathcal{N}(d,1)$ is the same as that between Π_1 and Π_2). In this equivalent set-up, the solution is now 'obvious'. Rather than jumping from $\mathcal{N}(0,1)$ to $\mathcal{N}(d,1)$ in one step we get there in d steps: at each step moving from $\mathcal{N}(k-1,1)$ to $\mathcal{N}(k,1)$ for index $k=1,2,\dots,d$. This algorithm can be immediately transferred to the corresponding multidimensional set-up. This simple example contradicts, when using an SMC algorithm in the scenario above, the following statement made in [11]:

> '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 a certain class of target densities:

- The SMC algorithm analyzed, with computational cost linear in the number of particles and quadratic in the dimension d, is stable. More precisely we prove that the ESS converges weakly to a non-trivial random variable ε_N as d grows and the number of particles *is kept fixed*. The expected ESS also converges. 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 $\frac{1}{\sqrt{N}}$ *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.
- 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. This modification means only convergence of the ESS and error is relevant.

The first result is that of the weak convergence of the ESS as $d \to \infty$ (also the expected ESS converges to a constant in $(1, N)$). This is associated to the fact that the underlying, discrete-time, Feynman-Kac formula ([22]) will exist. It also allows one to show that a variety of quantities associated to SMC algorithms are well-behaved: for example the Monte Carlo error of fixed dimensional marginals. In particular, using the resampling-time construction of [25] we are able to establish the second point referring to algorithms actually implemented by practitioners. The results show that in high-dimensional problems, one is able to control the variability of the weights; this is a minimum requirement in order to apply 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 [11] quoted above. In the presence of a quadratic cost and increasingly sophisticated hardware (e.g. [39]) SMC methods are applicable, in the static context, in high-dimensions.

To support this, [35] presents further empirical evidence of the results presented here. In particular, it is shown there that SMC techniques are algorithmically stable (in the sense of the variance of the importance weights) for models with state spaces with over 1000 dimensions, with computer simulations that run in just over one 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 is subject to further research.

When there is no resampling, the proof of our results rely on relatively well-known 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 the 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. The case of resampling is dealt with by using the construction of [25]. Like the proof in that article, our results depend, crucially, on the fact that the effective sample size cannot coincide with the threshold when one chooses to resample; this requires some minor modifications, which do not overly change the algorithm.

We conclude with the following two caveats: whilst the results presented here do not fully guarantee that the class of SMC algorithms we consider here will have a fast rate of convergence (certainly not for the estimation of functionals that grow with d; see also e.g. [15] for some discussion on ESS), they do establish that the algorithms will not collapse with regards to some criteria. Similarly, the convergence of the ESS without resampling does not imply that one should avoid resampling. Indeed, it can only be determined whether one should resample given a key quantity which is not explicit. This quantity can be numerically approximated, but its approximation is rather difficult (see e.g. [2] and the references therein). Nonetheless, we believe our results are an important first step in the analysis of high-dimensional SMC algorithms.

1.2 Structure of the Article

This article is structured as follows. In Section 2 we discuss the SMC algorithm of interest and the class of target distributions we consider. In Section 3 the first result of the paper is given showing that the expected ESS converges in distribution to a non-trivial random variable as $d \to \infty$ when our SMC 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 $\frac{M}{\sqrt{N}}$, where M is independent of d. We address the issue of resampling in Section 4, where it is shown that any dynamically resampling SMC algorithm, using the deterministic ESS (the expected ESS with one particle) will resample a finite number of times (again as $d \to \infty$) 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. In Section 5 we investigate our results, by verifying the assumptions for a particular example and numerically approximating some quantities of interest. These latter quantities show whether or not we will need to resample, and potentially the amount of resampling that is needed. Finally, in Section 6 the article is concluded with some remarks and ideas for future work. Proofs are collected in four appendices, the first two concerning the fCLT, the third on proofs for the case of resampling and the final on the verification of our assumptions.

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 $\mathbb{E}\,[\,|X|^\varrho\,] < \infty$. For $N \geq 1$, $\mathcal{N}_N(\mu, \Sigma)$ denotes the normal distribution of mean vector μ and covariance Σ; when $N = 1$ 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 consider the scenario when one wishes to sample from a target distribution with density Π on $\mathbb{R}^d$ with respect to Lebesgue measure, which is known point-wise up to a normalizing constant. In order to sample from Π, we introduce a sequence of 'bridging' densities which start from an easy to sample target and evolve toward Π. In particular, we will consider the densities (as in e.g. [23]):

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

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. Below, we use the short-hand Γ_n to denote un-normalized densities associated to Π_n. Other choices of sequences of bridging densities are possible and are discussed in Section 3.2.

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 targets the sequence of densities:

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

with domain $(\mathbb{R}^d)^n$ of dimension that increases with $n = 1, \ldots, p$; here, $\{L_n\}$ is a sequence of artificial backward Markov kernels that can, in principle, be arbitrarily selected. The work in [23] motivates the selection of $\{L_n\}$ and characterizes the optimal kernel, in terms of minimizing the variance of the importance weights for the SMC algorithm. Let $\{K_n\}$ be a sequence of Markov kernels of invariant density $\{\Pi_n\}$ and Υ a distribution; 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. This algorithm with no resampling is the annealed importance sampling described in [43] and arises when the backward Markov kernel L_n is chosen as follows:

$$L_n(x, x') = \frac{\Pi_{n+1}(x')K_{n+1}(x', x)}{\Pi_{n+1}(x)} \ .$$

For simplicity, we will henceforth assume that $\Upsilon \equiv \Pi_0$. It is remarked that, due to the results of [9, 11, 49], it appears that the cost of the population Monte Carlo method of [13] 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) (Figure 1) 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 the 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 are a wide variety of resampling techniques and we refer the reader to [29] for details; in this article we only consider the multinomial method just described above.

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

$$w_0(x_0^i) = \frac{\Gamma_0(x_0^i)}{\Upsilon(x_0^i)} \ .$$

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

1. If $n \leq p$, for each i sample $X_n^i \mid x_{n-1}^i$ from K_n and calculate the weights

$$w_n(x_{l:n-1}^i) = \frac{\Gamma_n(x_{n-1}^i)}{\Gamma_{n-1}(x_{n-1}^i)} \, w_{n-1}(x_{l:n-2}^i)$$

with the convention $x_{0:-1}^i \equiv x_0^i$. Calculate the Effective Sample Size (ESS):

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

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

resample $x_n^1, \dots x_n^N$ according to their normalised weights

$$w_n(x_{l:n-1}^i) / \sum_{j=1}^N w_n(x_{l:n-1}^j) \ ; \tag{3}$$

set $l = n$;

re-initialise the weights by setting $w_n(x_{l:n-1}^i) \equiv 1$, $1 \leq i \leq N$;

let $x_n^1, \dots x_n^N$ now denote the resampled particles.

Set $n = n + 1$.

Return to the start of Step 1.

Figure 1: The SMC algorithm analyzed in this article.

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 need to simplify the structure of the algorithm (similarly to the MCMC results in high dimensions in e.g. [7, 10, 45]). 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 more 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 much more general scenarios. 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 $\phi_n = \phi_n(d)$ and $p = p(d)$ 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 the sequence $\{\phi_n\}$ will change with d, all elements of our SMC algorithm will also depend on d. We will 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*

$$l_d(t) = \left\lfloor d\left(\frac{t-\phi_0}{1-\phi_0}\right)\right\rfloor \tag{7}$$

to switch between continuous time and discrete time. Related to the above, it will be convenient to also 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*. This is an analogue to the connection between bridge sampling and path sampling in [30].*

2.2 Conditions

We now state the conditions under which we will 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* $\{k_s\}$.

(i) *(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.

(ii) (*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 $s \in (\phi_0, 1]$:

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

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

(iii) (*Level Sets*). Define $C_c = \{x : V(x) \le c\}$ with V as in (ii). 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 $s, t \in (\phi_0, 1]$

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

Note that the statement that C is $(1, \theta, \nu)-$small w.r.t. to k_s means that C is a one-step small set for the Markov kernel, with minorizing distribution ν and parameter $\theta \in (0, 1)$ (see e.g. [42]).

Assumptions like (A1) are fairly standard in the literature for 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 (where the same invariant measure is assumed for each time-step) by virtue of having changing invariant measures, but will be verifiable in real contexts. Note that one can relax our assumptions to, e.g. sub-geometric ergodicity versus geometric ergodicity, at the cost of an increased level of complexity in our proofs.

It is also remarked that the assumption that g is upper bounded is only used in Section 4, when controlling the resampling times.

3 The Algorithm Without Resampling

We will now consider the case when we omit the resampling steps in the specification of our 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 consider first 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 characterise 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 will 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 the 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 $d\to\infty$. It is remarked that one can treat the double-sum in (10) in one go, as in [22, pp. 295-297], but the end result will be the same.

3.2 Results and Remarks for the 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:

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

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 $\sigma^2_\star = \sigma^2_{\phi_0:1}$.

Remark 3.1. *The quantity $\sigma^2_\star$ above when divided by $(1-\phi_0)^2$ is a continuous average of the asymptotic variances of the $\{k_s\}$ in the Markov chain CLT (e.g. [42]). It will be key in understanding the dependence of the algorithm on the underlying kernels, as well as understanding when to resample (if at all).*

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:*

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

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

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

Applying the growth bounds in Theorem A.1 we get that the remainder term $\mathbb{E}\left[\,|R_{0:d-1}|^{2+\delta}\,\right]$ is controlled as $\pi_{\phi_0}(V^r) < \infty$ (due to $r \in [0, \frac{1}{2})$). The martingale array term $\mathbb{E}\left[\,|M_{0:d-1}|^{2+\delta}\,\right]$ 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 [48, 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 [12]) 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:

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

a result directly implied again from Lemma 3.1. Therefore, by Theorem 2 of [48, 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. ∎

Remark 3.2. *Note that the asymptotic ESS,* $\varepsilon_N = \frac{[\sum_{i=1}^N e^{X_i}]^2}{\sum_{i=1}^N e^{2X_i}}$*, takes values in the interval* $(1, N)$*. Moreover, one can use standard manipulations to yield*

$$\begin{aligned}
\mathbb{E}\Big[\frac{[\sum_{i=1}^N e^{X_i}]^2}{\sum_{i=1}^N e^{2X_i}}\Big] &= 1 + \mathbb{E}\Big[\frac{\sum_{l\neq q} e^{[X_l+X_q]}}{\sum_{i=1}^N e^{2X_i}}\Big] \\
&= 1 + \sum_{l\neq q}\mathbb{E}\Big[\Big(\sum_{j:j\neq l, j\neq q} e^{[2X_j - X_l - X_q]} + e^{[X_l - X_q]} + e^{[X_q - X_l]}\Big)^{-1}\Big] \\
&\geq 1 + \sum_{l\neq q}\mathbb{E}\Big[\sum_{j:j\neq l, j\neq q} e^{[2X_j - X_l - X_q]} + e^{[X_l - X_q]} + e^{[X_q - X_l]}\Big]^{-1} \\
&= 1 + \frac{N(N-1)}{(N-2)e^{3\sigma_\star^2} + 2e^{\sigma_\star^2}} \\
&\geq 1 + (N-1)e^{-3\sigma_\star^2}
\end{aligned} \tag{19}$$

where we have used Jensen's inequality in the third line. The result suggests that one can compensate for the high dimensionality by reducing the gap in the bridges with $\mathcal{O}(d^2)$ *cost; the stability of the expected ESS will depend upon* $\sigma_\star^2$ *(i.e. the closer* $\sigma_\star^2$ *to 0, the closer to* N *the asymptotic expected ESS is).*

Remark 3.3. *An interesting point is the computational cost relative to MCMC. As proved in [45], random walk Metropolis requires a computational cost proportional to* d^2*. However, the criteria used here and in [45] are minimal requirements for stability as* $d \to \infty$*. Clearly, one cannot say, using the work on high-dimensional MCMC and the work here that the SMC algorithm leads to faster convergence with a similar computational cost than MCMC, but, this latter observation is apparent in empirical results (e.g. [17]).*

Remark 3.4. *Our result is associated to the stability of the underlying (d-time marginal) Feynman-Kac formula [22] (see the proof of Proposition 4.1). That is, the underlying dynamical system which one is trying to approximate via SMC algorithms, converges to a limit as d grows. If there are less than d steps, the formula does not exist and so in a sense one cannot trade off the particle count with the number of densities, to reduce the computational cost.*

Remark 3.5. *Using similar arguments, one can also investigate SMC stability when employing the sequence of densities:*

$$\Pi_n(x) \propto \exp\Big\{\phi_n \sum_{i=1}^d g(x_i) + (1-\phi_n)\sum_{i=1}^d h(x_i)\Big\}\ ,$$

for some 'simple' $h : E \mapsto \mathbb{R}$*, see e.g. [30]. When adopting the sequence* $\phi_n = \phi_0 + (1-\phi_0)\frac{n}{d}$*, one can establish that the expected ESS converges to a similar quantity as in Theorem 3.2, except* *that the variance term will now be:*

$$\sigma_\star^2 = (1-\phi_0)\int_{\phi_0}^1 \pi_s\left(\widehat{[g-h]}_s^{\,2} - k_s(\widehat{[g-h]}_s)^2\right)ds\ ,$$

with $\widehat{[g-h]}_s$ *defined via the Poisson equation in the obvious way, see (15). In general, it is not clear whether this will lead to a higher ESS, as* π_s *and* k_s *differ here from their earlier definitions.*

Another interesting tempering scheme is 'data-point tempering' [15]. In this scenario, one associates the function g *to observed data and constructs* $\{\Pi_n\}$ *so that the incremental weights at time* n *of the algorithm are* $\exp\{\sum_{j=1}^d g(y_n; x_{n-1,j})\}$ *where we have assumed an i.i.d. structure for the target and data,* $y_n \in \mathbb{R}^{d_y}$*. Under no resampling and data* $y_1,\ldots,y_p$ *one obtains the un-normalized weight* $\exp\{\sum_{n=1}^p\sum_{j=1}^d g(y_n; x_{n-1,j})\}$*. This algorithm will lead to weights that explode with the dimension, but the cost is only* pNd*. The algorithm is easily stabilized, in high dimensions, by introducing the potential for a new data-point via a linear annealing scheme with* $\lfloor d/p\rfloor$ *steps, yielding a computational cost of* Nd^2*.*

Remark 3.6. *The SMC scheme that we have described can be, to an extent, automated (see [35]). That is, there are procedures where the cooling sequence* ϕ_n *can be determined with no user input. Thus, one might argue that the number of densities and spacings are less of a concern to practitioners. However, our result helps to provide guidelines for such adaptive methods and is still of use in this context. For example, if an adaptive procedure leads to too many densities, it can be made more efficient.*

3.3 Monte Carlo Error

We have showed that the choice of bridging steps as in (6) stabilises the ESS in high dimensions. The error in the estimation of expectations, which can be of even more practical interest than the 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 $1\le i\le N$. In order to consider the Monte Carlo error, we use the below result, 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 directly from Proposition A.1 in the Appendix when choosing time sequences $s(d)\equiv\phi_0$ and $t(d)\equiv 1$. ∎

The Monte Carlo error result now follows:

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

$$\lim_{d\to\infty} \Big\| \sum_{i=1}^{N} \frac{w_d(X_{0:d-1}^i)}{\sum_{l=1}^{N} w_d(X_{0:d-1}^l)} \varphi(X_{d,1}^i) - \pi(\varphi) \Big\|_{\varrho} \leq \frac{M}{\sqrt{N}} \ .$$

Proof. Recall that the N particles are 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 defined 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:

$$w_d(X_{0:d-1}) \Rightarrow e^X \ , \quad X \sim \mathcal{N}(0, \sigma_\star^2) \ . \tag{20}$$

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:

$$\lim_{d\to\infty} \Big\| \sum_{i=1}^{N} \frac{w_d(X_{0:d-1}^i)}{\sum_{l=1}^{N} w_d(X_{0:d-1}^l)} \varphi(X_{d,1}^i) - \pi(\varphi) \Big\|_{\varrho} = \Big\| \sum_{i=1}^{N} \frac{e^{X_i}}{\sum_{l=1}^{N} e^{X_l}} \varphi(Z_i) - \pi(\varphi) \Big\|_{\varrho} \ , \tag{21}$$

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 (21) can be written as:

$$\frac{\frac{1}{N}\sum_{i=1}^{N} e^{X_i}\varphi(Z_i)}{e^{\sigma_\star^2/2}\frac{1}{N}\sum_{l=1}^{N} e^{X_l}} \Big[e^{\sigma_\star^2/2} - \frac{1}{N}\sum_{l=1}^{N} e^{X_l} \Big] + e^{-\sigma_\star^2/2} [\frac{1}{N}\sum_{i=1}^{N} e^{X_i}\varphi(Z_i) - e^{\sigma_\star^2/2}\pi(\varphi)] \ . \tag{22}$$

Finally, we have $\big\| \frac{1}{N}\sum_{l=1}^{N} e^{X_l} - e^{\sigma_\star^2/2} \big\|_{\varrho}$ and $\big\| \frac{1}{N}\sum_{i=1}^{N} e^{X_i}\varphi(Z_i) - e^{\sigma_\star^2/2}\pi(\varphi) \big\|_{\varrho}$ are both bounded by $\frac{M_1(\varrho,\varphi)}{\sqrt{N}}$ as $\mathbb{L}_\varrho$-norms of averages of i.i.d. variables of zero expectation; the fact that the coefficients in front of these averages in (22) are bounded by some $M_2(\varphi)$ completes the proof. □

Remark 3.7. *The result establishes that the error in the estimation of a fixed dimensional marginal is stable as the overall dimension grows to infinity. It is to be expected that for a fixed number of particles, the error in calculating the expectation of high-dimensional marginals (i.e. whose dimension grows with d), will increase with d. In line with the results of [14] for normalizing constants, we conjecture the cost to be $\mathcal{O}(d^3)$, when $\varphi : \mathbb{R}^d \to \mathbb{R}$.*

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 $\sigma_\star^2$ (see (19)). As a result, it makes sense to consider the option of resampling in our analysis in high dimensions.

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 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:

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

for any fixed d.

Assume that $s \in [\phi_0, 1]$ is a resampling time and $x_{l_d(s)}^{\prime,1}, \ldots x_{l_d(s)}^{\prime,N}$ 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:

$$\mathrm{ESS}_{(s,u)}(N) = \frac{\big(\sum_{i=1}^N \exp\{\frac{1}{\sqrt{d}}\sum_{j=1}^d S_{s:u,j}^i\}\big)^2}{\sum_{i=1}^N \exp\{\frac{2}{\sqrt{d}}\sum_{j=1}^d S_{s:u,j}^i\}} \tag{23}$$

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

$$S_{s:u,j}^i = \frac{1-\phi_0}{\sqrt{d}} \sum_{n=l_d(s)+1}^{l_d(u)} \{g(x_{n-1,j}^i) - \pi_{n-1}(g)\}\ ,$$

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

$$w_{l_d(t)}(x_{l_d(s):(l_d(t)-1)}^i) = \exp\{\tfrac{1}{\sqrt{d}}\sum_{j=1}^d S_{s:t,j}^i\}\ .$$

Note that we have modified the subscripts of ESS in (23), 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 $S_{s:u,j}^i$ are no longer independent over i or j, unless one conditions on the resampled particles $x_{l_d(s)}^{\prime,i}$, $1 \le i \le N$.

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 with one particle) will resample a finite number of times (again as $d \to \infty$) and also exhibit convergence of the ESS and 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 of the ESS and Monte Carlo error.

We use the resampling-times construction of [25]: this involves considering the expected value of the importance weight, and its square, over a system with a *single* particle. The theoretical resampling times are defined as:

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

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

under the convention that $\inf \varnothing = \infty$. 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. coordinates, each of which starts at $x_{0,j} \sim \pi_0$ and then evolves according to the kernels k_n.

Intuitively, following the ideas in [25], 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 modified version of it) as $d \to \infty$ with a probability that converges to 1 with a rate $\mathcal{O}(N^{-1/2})$. In particular, the question of whether one has to resample will depend, asymptotically, just upon $\sigma^2_{s:t}$. If this is not sufficiently small, then it will be necessary to resample. We discuss how this quantity may be approximated in Section 5.1.*

4.2 Stability of ESS and Monte-Carlo Error under the 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. The proof 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 that $g \in \mathscr{L}_{V^r}$, with $r \in [0, \frac{1}{2})$. Then, for any fixed $N > 1$, any $k \in \{1, \ldots, m^* + 1\}$, $t_{k-1} < t_k$ 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 $X_i^k \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2_{t_{k-1}:s_k})$ *and* $\sigma^2_{t_{k-1}:s_k}$ *as in* (28)*. In particular,*

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

Remark 4.2. *Observe that the random variables* X_i^k *converge to* 0 *as* $a \to 1$ *and, therefore, the asymptotic* ESS *converges to the upper limit* N *as the number of resampling steps increases.*

Note that if the $\{t_k(d)\}$ were analytically available then resampling at these instances would deliver an algorithm of d-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* $r \in [0, \frac{1}{2})$ *and let* $\varphi \in \mathcal{C}_b(\mathbb{R})$*. Then for any* $1 \le \varrho < \infty$ *there exists a constant* $M = M(\varrho, \varphi) < \infty$ *such that for any fixed* $N \ge 1$

$$\lim_{d\to\infty} \Bigg\| \sum_{i=1}^N \frac{w_d(X^i_{l_d(t_{m^*}(d)):(d-1)})}{\sum_{l=1}^N w_d(X^l_{l_d(t_{m^*}(d)):(d-1)})}\varphi(X^i_{d,1}) - \pi(\varphi)\Bigg\|_{\varrho} \le \frac{M}{\sqrt{N}}\ .$$

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 weight, 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. It should be possible by considering bounds (which do not increase with N and d) on quantities of the form

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

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 the empirical versions of the times (24)-(25). To this end, we will follow closely the proof of [25] and this will require the consideration of a mesh at the definition of the resampling times.

Consider some positive integer δ, and times:

$$\{s,t\} \subseteq \{\phi_0, \phi_0 + (1-\phi_0)/\delta, \phi_0 + 2(1-\phi_0)/\delta, \dots, 1\} \ .$$

Instead of looking at the times $\{t_k(d)\}_{k\geq 1}$ and $\{t_k\}_{k\geq 1}$, we consider times defined on a discrete-mesh

$$\Delta^{\delta}_{[s,t]} := \{s, s + (1-\phi_0)/\delta, s + 2(1-\phi_0)/\delta, \dots, t\} \ ,$$

determined as follows:

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

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

for a collection of thresholds (a_k). Clearly, via Proposition 4.1, these times converge to

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

Let $m^*(\delta)$ denote the number of these times, we have that $m^*(\delta) \leq m^*$ (with m^* now taking into account the choices of different thresholds a_k), but for δ large enough these values will be very close. Also, Theorems 4.1 and 4.2 hold under these modified times.

4.3.1 Result and Interpretation

Define, for $\upsilon \in (0,1)$, the following event:

$$\Omega^N_{m^*(\delta),d}(\upsilon, (a_k)_{1\leq k\leq m^*(\delta)}) := \Big\{ \forall\, 0 \leq k \leq m^*(\delta)-1,\ s \in \Delta^{\delta}_{[t^{\delta}_k(d), t^{\delta}_{k+1}(d)]},$$
$$\big|\, \tfrac{1}{N}\, \mathrm{ESS}_{(t^{\delta}_k(d),s)}(N) - \mathrm{ESS}_{(t^{\delta}_k(d),s)} \,\big| < \upsilon \,\big|\, \mathrm{ESS}_{(t^{\delta}_k(d),s)} - a_{k+1} \big| \Big\}$$

where

$$\mathrm{ESS}_{(t^{\delta}_k(d),s)} = \frac{\mathbb{E}\left[\ \exp\left\{\frac{1}{\sqrt{d}}\sum_{j=1}^{d} S_{t^{\delta}_k(d):s,j}\right\}\right]^2}{\mathbb{E}\left[\ \exp\left\{\frac{2}{\sqrt{d}}\sum_{j=1}^{d} S_{t^{\delta}_k(d):s,j}\right\}\right]}$$

correspond to the expected ESS over a single particle involved in the definition of $\{t^{\delta}_k(d)\}$. Here $(a_k)_{1\leq k\leq m^*}$ are a collection of thresholds which are sampled from some absolutely continuous distribution; they are required to avoid the degenerate situation when the thresholds coincide with the criteria; see [25] for details. Now, the work in [25], for fixed d, establishes the following:

1. It shows that on the event $\Omega^N_{m^*(\delta),d}(\upsilon,(a_k)_{1\le k\le m^*(\delta)})$, if the deterministic resampling criteria tell us to reample, so do the empirical ones [25, Lemma 5.3]. That is:

$$\mathrm{ESS}_{(t_k^\delta(d),s)} > a_{k+1} \Rightarrow \frac{1}{N}\,\mathrm{ESS}_{(t_k^\delta(d),s)}(N) > a_{k+1}\ , \quad s\in\Delta^\delta_{[t_k^\delta(d),t_{k+1}^\delta(d)]}\ ,$$

   and

$$\mathrm{ESS}_{(t_k^\delta(d),s)} < a_{k+1} \Rightarrow \frac{1}{N}\,\mathrm{ESS}_{(t_k^\delta(d),s)}(N) < a_{k+1}\ , \quad s\in\Delta^\delta_{[t_k^\delta(d),t_{k+1}^\delta(d)]}\ .$$

2. Then, for any $\upsilon\in(0,1)$:

$$\bigcap_{1\le k\le m^*(\delta)}\{\, t_k^{\delta,N}(d)=t_k^\delta(d)\,\} \supset \Omega^N_{m^*(\delta),d}(\upsilon,(a_k)_{1\le k\le m^*(\delta)})\ ,$$

   [25, Proposition 5.3] with $t_k^{\delta,N}(d)$ denoting the practical empirical resampling times on the δ-grid.

3. Conditionally on $(a_k)_{1\le k\le m^*(\delta)}$, we have that $\mathbb{P}\,[\,\Omega\setminus\Omega^N_{m^*(\delta),d}(\upsilon,(a_k)_{1\le k\le m^*(\delta)})\,]\to 0$ as N grows, [25, Theorem 5.4].

A combination of 2 and 3 provides the interpretation that, with a probability that increases to 1 with N, the deterministic resampling times $\{t_k(d)\}$ will coincide with the practical resampling times $\{t_k^{\delta,N}(d)\}$. Our result is as follows.

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), and for any $\upsilon\in(0,1)$, $\delta\in\Delta$ there exists an $M=M(m^*(\delta))<\infty$ such that for any $1\le N<\infty$, we have*

$$\lim_{d\to\infty}\mathbb{P}\,[\,\Omega\setminus\Omega^N_{m^*(\delta),d}(\upsilon,(a_k)_{1\le k\le m^*(\delta)})\,]\le\frac{M}{\sqrt{N}}\ .$$

Our proof of Theorem 4.3, in Appendix C.3, will focus on point 3. of the results in [25] above with the aim of establishing the asymptotic bound stated in the theorem.

5 Example on Symmetric Random Walk

We will now verify assumptions (A1-2) when the π_s-invariant transition kernel is a Random-Walk Metropolis algorithm, with proposed increments $\mathcal{N}(0,s^{-1})$. That is:

$$q_s(x,dy)=\frac{\sqrt{s}}{\sqrt{2\pi}}\,e^{-s\,\frac{(y-x)^2}{2}}\,dy$$

with acceptance probability:

$$a_s(x,y) = 1 \wedge \frac{\pi_s(y)}{\pi_s(x)} \ .$$

For simplicity we set $q_s(dy) \equiv q_s(0,dy)$. That is, we will look at the Markov kernel:

$$k_s(x,dy) = a_s(x,y)\, q_s(x,dy) + \delta_x(dy) \int_E (1 - a_s(x,y))q_s(x,dy) \ . \tag{30}$$

Notice that we assume that the variance of the proposal is $1/s$, $s \in [\phi_0, 1]$. One can use $f(s)^{-1}$ for the proposal variance, where f is a bounded positive continuous function that is monotonically increasing with a bounded derivative. This is omitted only for notational clarity and using f in the proofs will only complicate the subsequent notations.

We will assume that for every $s \in [\phi_0, 1]$ one has

- π_s is bounded away from zero on compact sets and is upper-bounded.
- π_s is super-exponential with asymptotically regular contours; see [33] for details.

We will add the condition

$$C^* := \sup_{x\in\mathbb{R}\,, s\in[\phi_0,1]} \left\{ \int_{A(x)^c} G(x,z) q_s(z) dz \right\} < +\infty \tag{31}$$

with $G(x,z) = g(x) - g(x+z) > 0$ on $A(x)^c$ (see (56) for details on $A(x)$). This assumption is used to simplify some calculations in the proof and is verifiable (see Remark 5.2). The above assumptions will be termed E in the following proposition. The proof can be found in Appendix D.

Proposition 5.1. *Assume (E). Then the symmetric random walk kernel* (30) *satisfies (A1-2).*

Remark 5.1. *It is straightforward to verify (A1) using standard results in the literature. However, (A2) is non-standard, due to the difference of invariant measures present in* (30).

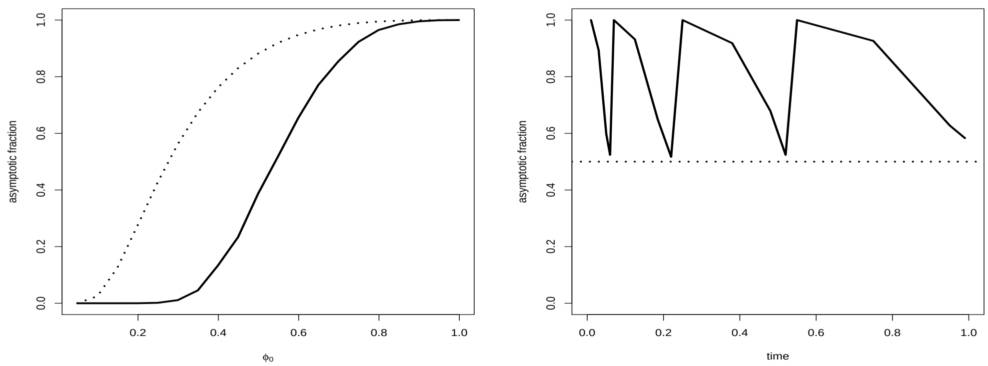
Remark 5.2. *Note, for* (31)*, that if* $g(x) = -x^2/2$ *then* $G(x,z) = \frac{1}{2}[z^2 + 2xz]$ *. Hence we have*

$$\int_{A(x)^c} G(x,z) q_s(z) dz \le \frac{1}{2s} \le \frac{1}{2\phi_0} \ .$$

Thus, assumption (31) will hold in the Gaussian case.

5.1 Numerical Calculations

Whilst the expression we have given on the RHS of (16) has an intuitive lower-bound (19), in general one has to numerically approximate the asymptotic variances and then perform a Monte Carlo simulation to deduce the value of the limiting expected ESS. To reduce this work-load we look at the quantities $e^{-\sigma^2_{s:t}}$ for $s < t$ with $s, t \in [\phi_0, 1]$.



(a) No Resampling (b) Resampling

Figure 2: A Plot of $e^{-\sigma^2_{\phi_0:s}}$ against ϕ_0 (a) and $e^{-\sigma^2_{\phi_0:s}}$ against the time parameter when resampling is performed (b). In (a), the dotted line is when $\sigma^2_\star$ is replaced with the integral over the i.i.d. variance. In (b), when a resampling operation occurs, the fraction re-initializes to 1. The dotted line is the resampling threshold.

We consider the scenario where $\pi_s(x) \propto e^{-\frac{s}{2}x^2}$ and the Markov kernels are symmetric random walks with proposals that are normal with variance $1/(cs)$. Here, we choose $c = 1/25$ to ensure that the kernels have average acceptance rates around 0.23 (see [45]). Note, that (A1-2) are satisfied and $g \in \mathscr{L}_{V^r}$, $r \in (0, \frac{1}{2})$. We use standard Monte Carlo with the trapezoidal rule, to approximate $\sigma^2_\star \equiv \sigma_{\phi_0:1}$.

In Figure 2 (a) we can see the numerically approximated fraction and the scenario (in dots) where we replace $\sigma^2_\star$ with $(1-\phi_0)\int_{\phi_0}^1 \mathbb{E}_{\pi_s}\left[\,(g(X)-\pi_s(g))^2\,\right] ds$ which reflects the best possible mixing of the Markov kernels (and corresponds to $k_s(x,dy) \equiv \pi_s(dy)$). The plot shows that the algorithm can perform quite well, w.r.t. the ESS when $\phi_0 \approx 0.4$ and improves dramatically after this. This is unsurprising; since π_{ϕ_0} is sampled from, we would like it to be close to π. In the perfectly mixing case, the algorithm appears to be reasonable even when $\phi_0 \approx 0.2$. In Figure 2 (b) we plot the asymptotic fraction when one resamples (hence re-initializing to 1) at the limiting times (26)-(27), when $a = 0.5$ and $\phi_0 = 0.01$. The plot shows, for this example, the ESS is controlled and one needs only to resample three times. The plot is also typical of those found in empirical results in high dimensions (e.g. [35]).

The results indicate, for this example, that one would have to incorporate resampling for this problem. In addition, it is likely that one needs to resample more often at the start of the algorithm, with the rate of resampling tending to fall as one approaches the target of interest. This latter observation is consistent with our empirical experience of applying the algorithm.

6 Summary

In this paper we have considered the stability, in high dimensions, for a class of SMC methods. It has been established that the expected ESS converges to a constant bigger than 1 and the Monte Carlo error is $\mathcal{O}(N^{-1/2})$ for the estimation of fixed dimensional marginals. In addition, that if one performs dynamic resampling at deterministic times then, regardless of the number of dimensions, one resamples only a finite number of times. The convergence of the ESS and Monte Carlo error when resampling at stochastic times is also established, with high probability.

There are a variety of extensions to our work. A first question is the analysis of the case when the target is not i.i.d.. When one does not assume any particular dependence structure, without resampling, the problem is to prove a fCLT/CLT for a double array of non-homogeneous chains. At present, we are not aware of the appropriate martingale theory, which would allow us to follow the approach here. It may be more fruitful to consider dependence structures that are studied in the MCMC literature (e.g. [7]).

Secondly, we have only considered some specific bridging and kernel densities. However, it would be of interest to look at alternative sequences of densities, for example those introduced in [18] or when considering more powerful MCMC kernels as in [19].

Thirdly, an interesting avenue to pursue is the stability of the SMC approximation of multi-level Feynman-Kac formulae [22]. This is particularly important for problems in rare-events analysis. In this case one introduces a sequence of sets which converge to the rare region of interest. The question is how to parameterize the sets such that, as one makes the set of interest rarer, the algorithm is stable (e.g. w.r.t. logarithmic efficiency). We suggest [14] and [21] from the splitting literature as useful starting points.

Fourthly, the analysis in Section 5.1 suggests that whilst one may still require a computational cost of Nd^2, it may be advantageous to have an annealing scheme with d steps, but which changes in a non-homogeneous fashion. For example, one would expect, on the basis of empirical experience, that an annealing sequence which moves more slowly at the beginning may lead to algorithms with higher, on average, ESS; an investigation of this would be of interest.

Acknowledgements

We thank Pierre Del Moral, Christophe Andrieu and Nicolas Chopin for valuable conversations associated to this work. The work of Dan Crisan was partially supported by the EPSRC Grant No: EP/H0005500/1. We also thank Nikolas Kantas for some useful feedback.

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. [28]). 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. [42]), 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 $0 \le n_0 \le n$:

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

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

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

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

$$\begin{aligned} A(x) &= \sum_{l\geq 0}[k_t^l - \pi_t](\varphi_t)(x) - \sum_{l\geq 0}[k_s^l - \pi_s](\varphi_t)(x)\ ; \\ B(x) &= \sum_{l\geq 0}[k_s^l - \pi_s](\varphi_t - \varphi_s)(x)\ . \end{aligned} \tag{32}$$

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

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

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 $x \in E$:

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

Returning to $A(x)$ in (32), one can use Lemma C2 of [4] to show that this is equal to:

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

Using identical manipulations to [4], it follows that:

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

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* $r\,p \leq 1$*:*

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

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

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

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:

$$\sum_{n=n_1}^{n_2} \{\, \varphi_n(X_n) - \pi_n(\varphi_n)\,\} = M_{n_1:n_2} + D_{n_1:n_2} - E_{n_1:n_2} + T_{n_1:n_2}\ ; \tag{33}$$

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

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

$$\mathbb{E}\,[\,|T_{n_1:n_2}|^p \mid X_{n_0}\,] \leq M\ V^{rp}(X_{n_0})\ . \tag{34}$$

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:

$$\mathbb{E}\,[\,|E_{n_1:n_2}|^p \mid X_{n_0}\,] \leq M\, V^{rp}(X_{n_0})\ . \tag{35}$$

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:

$$\mathbb{E}\,[\,|D_{n_1:n_2}|^p \mid X_0\,] \leq M\,V(X_{n_0})^{rp}\ . \tag{36}$$

The bounds (34), (35) and (36) 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. [48, pp. 499-500]) which states that for any $p > 1$:

$$\mathbb{E}\,[\,|M_{n_1:n_2}|^p \mid X_{n_0}\,] \leq M(p)\,d^{\frac{p}{2}\vee 1-1}\sum_{n=n_1+1}^{n_2}\mathbb{E}\,[\,|\,\widehat{\varphi}_n(X_n) - k_n(\widehat{\varphi}_n)(X_{n-1})\,|^p \mid X_{n_0}\,]\ , \tag{37}$$

see [5] 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 (37) 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:

$$\begin{aligned}|\,\mathbb{E}_{X_{l_d(s(d))}}\,[\,\varphi(X_{l_d(t(d))})\,] - \pi_t(\varphi)\,| &\leq |\,\mathbb{E}_{X_{l_d(s(d))}}\,[\,\varphi(X_{l_d(t(d))}) - k_{u_d}^{m_d}(\varphi)(X_{u_d})\,]\,| \\ &\quad + |\,\mathbb{E}_{X_{l_d(s(d))}}\,[\,k_{u_d}^{m_d}(\varphi)(X_{u_d})\,] - \pi_{u_d}(\varphi)\,| + |\,\pi_{u_d}(\varphi) - \pi_t(\varphi)\,|\ .\end{aligned} \tag{38}$$

Now, the last term on the R.H.S. of (38) 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. Lebesgue) 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 (38) 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 $\|k_s^l - \pi_s\|_{V^r} \le M\,\rho^l\,V(x)^r$ for any $s \in (\phi_0, 1]$, so we also have that $|k_{u_d}^{m_d}(\varphi)(X_{u_d}) - \pi_{u_d}(\varphi)| \le M\,\rho^{m_d}\,V(X_{u_d})^r$. Taking expectations and using Lemma A.1(iii) we obtain that:

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

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 (38). The following decomposition holds, as intermediate terms in the sum below cancel out, for $u_d \ge 1$:

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

where we use the notation $k_{i:j}(\varphi)(x) = \int k_i(x,dx_1)\times\cdots\times k_j(\varphi)(x_{j-i+1})$, $i \le j$. 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):

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

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

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

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

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

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. $\square$

B Proofs for Section 3

There are related results to Theorem 3.1 (see e.g. [41, 50]), 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 $|\widehat{g}_s(x)| \le M\,|g|_{V^r}\,V^r(x)$. Thus, due to the presence of quadratic functions in the definition of $\varphi_s(\cdot)$ we get directly that $|\varphi_s(x)| \le M\,V^{\bar{r}}(x)$ 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:

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

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

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

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:

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

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 [8] 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 $\sigma^2_{\phi_0:t}/(1-\phi_0)^2$.

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:

$$I_d(t) = \frac{1}{d}\sum_{n=1}^{l_d(t)} \Big[k_{n,d}(\widehat{g}^2_{n,d})(X_{n-1}(d)) - \big\{k_{n,d}(\widehat{g}_{n,d})(X_{n-1}(d))\big\}^2\Big] \ . \tag{39}$$

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 [31] 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 $s(d) \to s$. Then:*

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 $n_0 = 0$ gives that:

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

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

$$\mathbb{E}[\,|M_{l_d(s):(l_d(t)-1)}|^{2+\epsilon}\,] \le M\,d^{1+\frac{\epsilon}{2}}\,\mathbb{E}\,[\,V(X_0)\,]\ ,$$

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

$$\mathbb{E}[\,|R_{l_d(s):(l_d(t)-1)}|^{2+\epsilon}\,] \le M\,\mathbb{E}\,[\,V(X_0)\,]\ .$$

One now needs to notice that these 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 $\mathbb{E}\,[\,M_{l_d(s):(l_d(t)-1)}\,] = 0$. It remains to prove (ii). The residual term $R_{l_d(s):(l_d(t)-1)}/\sqrt{d}$ 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 [46]) 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 $t \in [s(d), 1]$:

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

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:

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

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:

$$f_d(s(d), t) \to_t e^{-\sigma^2_{s:t}}\ , \tag{40}$$

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

$$e^{\frac{2}{\sqrt{d}}\overline{S}_{s(d):t}} = 1 + \tfrac{2}{\sqrt{d}}\,\overline{S}_{s(d):t} + \tfrac{2}{d}\,\overline{S}^2_{s(d):t} e^{2\zeta_{d,t}}\ ; \tag{41}$$

$$e^{\frac{1}{\sqrt{d}}\overline{S}_{s(d):t}} = 1 + \tfrac{1}{\sqrt{d}}\,\overline{S}_{s(d):t} + \tfrac{1}{2d}\,\overline{S}^2_{s(d):t}\, e^{\zeta'_{d,t}}\ , \tag{42}$$

where $\zeta_{d,t}, \zeta'_{d,t} \in \left[\,\frac{1}{\sqrt{d}}\,\overline{S}_{s(d):t} \wedge 0\,,\ \frac{1}{\sqrt{d}}\overline{S}_{s(d):t} \vee 0\,\right]$. 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 (41):

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

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 $\mathbb{E}\,[\,|e^{2\zeta_{d,t}} - 1|^q\,] \leq M(q)\,\mathbb{E}\,[\,|\zeta_{d,t}|\,]$ for any $q \geq 1$, 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):

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

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:

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

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 (42) 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 (47) 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 $N \geq 1$, $d \geq 1$, $i \in \{1, \ldots, N\}$, $j \in \{1, \ldots, d\}$:*

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

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 $k \in \{1, \ldots, m^*\}$:

$$\mathbb{E}\,[\,V(X^i_{l_d(t_k(d)),j})\,|\,\mathscr{F}'^{,N}_{t_{k-1}(d)}\,] \leq M\,V(x'^{,i}_{l_d(t_{k-1}(d)),j}) \tag{43}$$

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 resampling) 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):

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

and, via (43), $\mathbb{E}\,[\,V(X'^{,i}_{l_d(t_1(d),j})\,] \leq NM$ 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 ($k \geq 2$), 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):

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

Thus, via (43) and the exchangeability of the particle and dimension index, we obtain that

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

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*

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

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:

$$\mathbb{E}\,\Big|\,\frac{1}{d}\sum_{j=1}^{d} \varphi(X'^{,i}_{l_d(t_1(d)),j}) - \pi_{t_1}(\varphi)\,\Big| \leq N\,\mathbb{E}\,\Big|\,\frac{1}{d}\sum_{j=1}^{d} \varphi(X^i_{l_d(t_1(d)),j}) - \pi_{t_1}(\varphi)\,\Big| \tag{44}$$

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:

$$\mathbb{E}\,\Big|\,\frac{1}{d}\sum_{j=1}^{d} \varphi(X^i_{l_d(t_1(d)),j}) - \mathbb{E}\,[\,\varphi(X^i_{l_d(t_1(d)),j})\,]\,\Big| + |\,\mathbb{E}\,[\,\varphi(X^i_{l_d(t_1(d)),j})] - \pi_{t_1}(\varphi)\,|\ . \tag{45}$$

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 (45) 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 (44) 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:

$$\mathbb{E}\,\Big|\,\frac{1}{d}\sum_{j=1}^{d}\varphi(X^i_{l_d(t_k(d)),j}) - \frac{1}{d}\sum_{j=1}^{d}\mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}\,[\,\varphi(X^i_{l_d(t_k(d)),j})\,]\,\Big| \quad + \\ \mathbb{E}\,\Big|\,\frac{1}{d}\sum_{j=1}^{d}\mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}\,[\,\varphi(X^i_{l_d(t_k(d)),j})\,] - \pi_{t_k}(\varphi)\,\Big|\;. \tag{46}$$

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 (46) is $\mathcal{O}(d^{-1/2})$. The second term in (46) 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:*

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

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:

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

Proof. We will make use of the Poisson equation and employ the decomposition (33) used in the proof of Theorem A.1. In particular, a straight-forward calculation gives that:

$$R_{i,j} = \sum_{n=n_1+1}^{n_2} \big\{ \,(\,\mathbb{E}_{X_{n_1,j}} - \mathbb{E}_{\pi_{t_{k-1}}}\,)[\,\widehat{g}_n(X^i_{n-1,j}) - \widehat{g}_{n-1}(X^i_{n-1,j})\,]\,\big\}$$
$$+ \,(\,\mathbb{E}_{X_{n_1,j}} - \mathbb{E}_{\pi_{t_{k-1}}}\,)[\,g(X_{n_2,j}) - \widehat{g}_{n_2}(X_{n_2,j})\,] + \widehat{g}_{n_1}(X_{n_1,j}) - \pi_{t_{k-1}}(\widehat{g}_{n_1}) \ , \qquad (47)$$

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 (33) has expectation 0, so is not involved in our manipulations. We will first deal with the sum in the first line of (47), 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 [28]; here, under (A1-2) we have that either:

$$\|\,[\delta_{X_{n_1,j}} - \pi_{t_{k-1}}](k_{n_1+1:n})\,\|_{V^r} \le M\rho^{n-n_1}V(X_{n_1,j})^r + M'\zeta^{n-n_1} \qquad (48)$$

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 [28, pp. 1650]. (Note that this follows from a uniform in time drift condition which follows from Proposition 4 of [28] (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 (48) and Proposition C.3 can give directly that the second term in (47) will vanish in the limit when summing up over j and dividing with d. Finally, for the last term in (47): 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:

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

where we have defined:

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

with:

$$\overline{G}_{i,j} = \sum_n \left\{ g(X^i_{n,j}) - \mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}[\, g(X^i_{n,j})\,] \right\} ;$$

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

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:

$$\begin{aligned} \mathcal{G}_{0,d} &= \sigma(X'^{,l}_{l_d(t_{k-1}(d)),j}, 1 \le j \le d,\, 1 \le l \le N) \ ; \\ \mathcal{G}_{j,d} &= \mathcal{G}_{j-1,d} \bigvee \sigma(X^i_{n,j},\, l_d(t_{k-1}(d)) \le n \le l_d(s_k(d)) - 1) \ , \quad j \ge 1 \ . \end{aligned} \tag{49}$$

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 (49). 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 (49) 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:

$$\begin{aligned} \frac{1}{d^2} \sum_{j=1}^{d} \mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}\,[\,G_{i,j}^2\,] &= \frac{1}{d^2} \sum_{j=1}^{d} \mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}\,[\,M_{i,j}^2 + R_{i,j}^2 + 2\,M_{i,j}R_{i,j}\,] \\ &= \frac{1}{d^2} \sum_{j=1}^{d} \mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}\,[\,M_{i,j}^2\,] + \mathcal{O}(d^{-1/2})\ . \qquad (50) \end{aligned}$$

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):

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

One then only needs to make use of Proposition C.2 to get (50). Now, using the analytical definition of $M_{i,j}$ from Theorem A.1 we have:

$$
\begin{aligned}
\frac{1}{d^2}\sum_{j=1}^{d}\mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}\,[\,M^2_{i,j}\,] &= \frac{1}{d^2}\sum_{j=1}^{d}\sum_{n=n_1+1}^{n_2}\big\{\,\mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}\,[\,\widehat{g}^2_n(X^i_{n,j}) - k^2_n(\widehat{g}_n)(X^i_{n-1,j})\,]\,\big\} \\
&= \frac{1}{d^2}\sum_{j=1}^{d}\sum_{n=n_1}^{n_2-1}\mathbb{E}_{X'^{,i}_{l_d(t_{k-1}(d)),j}}\,[\,\varphi_{n+1}(X^i_{n,j})\,] =: A_d \qquad (51)
\end{aligned}
$$

where:

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

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 (51), and using the above bound and Proposition C.2, we have:

$$
\Big|\,A_d - \frac{1}{d}\sum_{n=n_1}^{n_2-1}\pi_n(\varphi_{n+1})\,\Big| = \mathcal{O}(d^{-1})\ . \qquad (52)
$$

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

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

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

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

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

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

Thus, we also have:

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

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 and we only sketch the proof. The only real difference is the consideration of Proposition A.1. When $t_{m^*} < 1$, the latter result follows by considering d large enough so that $u_d \geq l_d(t_{m^*}(d))$ and by applying Lemma C.2. If $t_{m^*} = 1$, then one can simply integrate out the resampling-time:

$$\mathbb{E}\,[\,\varphi(X_{d,1}) - \pi(\varphi)\,] = \mathbb{E}\,[\,\sum_{i=1}^{N}\overline{w}^i_d\,k_1(\varphi)(X^i_{d,1}) - \pi(\varphi)\,]$$

where $\overline{w}^i_d \equiv \overline{w}_d(x^i_{l_d(t_{m^*-1}(d)):d-1})$ are used as a short-hand for the normalized weights. Then one follows the above argument. □

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:

$$\mathbb{P}\,[\,\Omega\setminus\Omega^N_{m^*(\delta),d}(\upsilon,(a_k)_{1\leq k\leq m^*(\delta)})\,] \leq \sum_{k=0}^{m^*(\delta)-1}\ \sum_{s\in\Delta^{\delta}_{[t^\delta_k(d),t^\delta_{k+1}(d)]}}\mathbb{P}\left[\,\left|\tfrac{1}{N}\,\mathrm{ESS}_{(t^\delta_k(d),s)}(N) - \mathrm{ESS}_{(t^\delta_k(d),s)}\right| \geq \upsilon\left|\mathrm{ESS}_{(t^\delta_k(d),s)} - a_{k+1}\right|\,\right]\ .$$

Define

$$\epsilon(d) := \inf_n\inf_s|\,\mathrm{ESS}_{(t^\delta_k(d),s)} - a_{k+1}\,|\ ;$$

we remark $\lim_{d\to\infty}\epsilon(d) = \epsilon > 0$ (with probability one). Hence we have

$$\mathbb{P}\,[\,\Omega\setminus\Omega^N_{m^*(\delta),d}(\upsilon,(a_k)_{1\leq k\leq m^*(\delta)})\,] \leq \sum_{k=0}^{m^*(\delta)-1}\ \sum_{s\in\Delta^{\delta}_{[t^\delta_k(d),t^\delta_{k+1}(d)]}}\mathbb{P}\left[\,\left|\tfrac{1}{N}\,\mathrm{ESS}_{(t^\delta_k(d),s)}(N) - \mathrm{ESS}_{(t^\delta_k(d),s)}\right| \geq \upsilon\epsilon(d)\right]\ .$$

Application of the Markov inequality, yields

$$\mathbb{P}\,[\,\Omega\setminus\Omega^N_{m^*(\delta),d}(\upsilon,(a_k)_{1\leq k\leq m^*(\delta)})\,] \leq \frac{m^*(\delta)\delta}{\upsilon\epsilon(d)}\max_{k,s}\mathbb{E}\left[\,\left|\tfrac{1}{N}\,\mathrm{ESS}_{(t^\delta_k(d),s)}(N) - \mathrm{ESS}_{(t^\delta_k(d),s)}\right|\,\right]\ .$$

Since k,s lie in a finite set and $\epsilon > 0$, we need only deal with the expectation as d grows. Note, in the expectation, the case $s = t^\delta_k(d)$ is not of interest; the ESS is constant and hence lower-bounds all other cases.

Application of Theorem 4.1 now yields:

$$\lim_{d\to\infty} \mathbb{E}\left[\,\left|\tfrac{1}{N}\mathrm{ESS}_{(t_k^\delta(d)),s)}(N) - \mathrm{ESS}_{(t_k^\delta(d),s)}\right]\right. = \mathbb{E}\left[\,\left|\tfrac{1}{N}\,\mathrm{ESS}_{(t_k^\delta,s)}(N) - \mathrm{ESS}_{(t_k^\delta,s)}\right|\,\right]$$

where

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

with $X_j^n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0,\sigma^2_{t_k^\delta:s})$. Let

$$\alpha_j^k = \exp\{X_j^k\}\ ; \quad \beta_j^k = \exp\{2X_j^k\}\ ; \quad \alpha^k = \exp\{\tfrac{1}{2}\,\sigma^2_{t_k^\delta:s}\}\ ; \quad \beta^k = \exp\{2\sigma^2_{t_k^\delta:s}\}\ ,$$

then we are to bound

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

Now, we have the decomposition

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

For the first part of the R.H.S. as the ESS divided by N is upper-bounded by 1, we can use Jensen and the Marcinkiewicz-Zygmund inequality. For the second 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 $M(k,\delta,s)$

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

Thus we have proved $\lim_{d\to\infty}\mathbb{P}\left[\,\Omega\setminus\Omega^N_{m^*(\delta),d}(\upsilon,(a_k)_{1\le k\le m^*(\delta)})\,\right] \le \frac{M(m^*(\delta))}{\sqrt{N}}$ as required. □

D Verifying the Assumptions

Proof of Proposition 5.1. We start with (A1)(i)-(ii); to establish uniform (in s) drift and minorization conditions for the kernel k_s. The proof is standard and included for completeness.

It is first noted that, for any $\delta_q > 0$, if $|x-y| < \delta_q$:

$$q_s(x,y) \ge \frac{\phi_0^{1/2}}{\sqrt{2\pi}}\exp\left\{-\frac{s}{2}\delta_q^2\right\} \ge \frac{\phi_0^{1/2}}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\delta_q^2\right\}\ . \tag{53}$$

This property will be used below. To establish the minorization, one can follow the proof of Theorem 2.2 of [44] to show that for any x, with $y\in B(x,\delta_q/2)$ (the open ball, centered x and of radius $\delta_{q/2}$), $A\in\mathscr{B}(\mathbb{R})$, $A\subseteq B(x,\delta_q/2)$

$$k_s(y,A) \ge \eta(x,\delta_q/2)\int_A (q_s(z,y)\wedge q_s(y,z))dz \ge \eta(x,\delta_q/2)\epsilon_q\int_A dz$$

where $\eta(x,\delta_q/2) = \inf_{x\in B(x,\delta_q/2)}\pi_1(x)/\|\pi_{\phi_0}\|_\infty$ and δ_q is as (53), ϵ_q as the RHS of the inequality in (53). Hence, we have the uniform minorization condition.

To prove the drift, we do not require it hold for $s = \phi_0$ as, in the algorithm, we sample exactly from π_{ϕ_0}. None-the-less, by our assumptions there exist a drift condition for k_{ϕ_0} (a symmetric normal random walk Metropolis-kernel of invariant π_{ϕ_0}); write the parameters λ, b. Now, for any $s \in (\phi_0, 1]$, via Lemma 5 of [3] and using that for any x, y, $\frac{q_s(x,y)}{q_{\phi_0}(x,y)} \leq \frac{1}{\sqrt{\phi_0}}$ one has

$$k_s(V)(x) \leq \frac{1}{\sqrt{\phi_0}}(k_{\phi_0}(V)(x) - V(x)) + V(x)$$

where

$$V(x) = \|e^{\phi_0 g}\|_\infty^{1/2} / e^{\frac{\phi_0}{2} g(x)}. \tag{54}$$

Now one can easily find a $\bar{c} \in [(1 - \phi_0^{-1/2}) \wedge (-\lambda/\sqrt{\phi_0}), 1 - \lambda\phi_0^{-1/2}]$ such that $k_s(V)(x) \leq \widetilde{\lambda} V(x) + \widetilde{b}\,\mathbb{I}_C(x)$ with $\widetilde{\lambda} \in (0,1)$, $\widetilde{b} < \infty$. Hence, the uniform drift condition is verified. (A1) (iii) can be verified in a similar manner to e.g. [28] and is omitted.

Now to (A2), which is a little more complex. Recall, we want to establish that there exist an $M < \infty$ such that for any $s, t \in (\phi_0, 1]$, $|||k_s - k_t|||_V \leq M|s - t|$. For simplicity, we will consider only the increment of proposal (via change of variables), so q_s is a zero mean normal density, with variance $1/s$. For any fixed $x \in \mathbb{R}$ q_s is a bounded-continuous function of $s \in [\phi_0, 1]$ and further, the first derivative w.r.t. s is upper-bounded by $\frac{1}{2\sqrt{2\pi\phi_0}} e^{-\phi_0 x^2/2}$ hence it follows that for any $x \in \mathbb{R}$, $s, t \in [\phi_0, 1]$:

$$|q_s(x) - q_t(x)| \leq \left(\frac{1}{2\sqrt{2\pi\phi_0}} e^{-\phi_0 x^2/2}\right)|s - t| \ . \tag{55}$$

Now central to our proof is the consideration of the acceptance probability, which is $\alpha_s(x, z) = 1 \wedge \exp\{s(g(x + z) - g(x))\}$. Let

$$A(x) = \{z : g(x + z) - g(x) > 0\} \tag{56}$$

then if $z \in A(x)$, $\alpha_s(x, z) = 1$. We begin by considering the acceptance part of the kernel. The difficult issue is when $z \in A(x)^c$ which is dealt with now:

$$\int_{A(x)^c} \varphi(x + z) \exp\{-sG(x, z)\} q_s(z) dz - \int_{A(x)^c} \varphi(x + z) \exp\{-tG(x, z)\} q_t(z) dz \tag{57}$$

where $\varphi \in \mathscr{L}_V$. Now for any fixed x, $z \in A(x)^c$ one has that

$$|\exp\{-sG(x, z)\} - \exp\{-tG(x, z)\}| \leq \left(G(x, z) e^{-\frac{\phi_0}{2} G(x,z)}\right)|s - t| \tag{58}$$

for every $s, t \in [\phi_0, 1]$. Then, returning to (57), it can be decomposed into the sum of

$$\int_{A(x)^c} \varphi(x + z)[\exp\{-sG(x, z)\} - \exp\{-tG(x, z)\}] q_s(z) dz \tag{59}$$

and

$$\int_{A(x)^c} \varphi(x+z)\exp\{-tG(x,z)\}[q_s(z)-q_t(z)]dz\ . \tag{60}$$

First consider (59). Applying (58), it follows that (59) is upper-bounded by $C_{\phi_0}|\varphi|_V|s-t|\int_{A(x)^c} e^{-\frac{\phi_0}{2}g(x+z)}G(x,z)e^{-\frac{\phi_0}{2}G(x,z)}q_s(z)dz$ where C_{ϕ_0} is associated to the Lyapunov function (54). Now as $e^{-\frac{\phi_0}{2}g(x+z)}e^{-\frac{\phi_0}{2}G(x,z)} = e^{-\frac{\phi_0}{2}g(x)}$ which is controlled by $V(x)$ and by assumption (31), $\int_{A(x)^c} G(x,z)q_s(z)dz$ is dealt with; hence (59) divided by $V(x)$ is upper-bounded by $C_{\phi_0}|\varphi|_V|s-t|C^*$. Our next task is (60). Applying (55), it is upper-bounded by

$$C_{\phi_0}|\varphi|_V\int_{A(x)^c} e^{-\frac{\phi_0}{2}g(x+z)}e^{-tG(x,z)}|s-t|\frac{1}{2\sqrt{2\pi\phi_0}}e^{-\phi_0 z^2/2}dz =$$
$$C_{\phi_0}|\varphi|_V e^{-tg(x)}\int_{A(x)^c} e^{(t-\frac{\phi_0}{2})g(x+z)}|s-t|\frac{1}{2\sqrt{2\pi\phi_0}}e^{-\phi_0 z^2/2}dz\ ;$$

on dividing by the Lyapunov function, we are to deal with the expression $\exp\{-(t-\frac{\phi_0}{2})G(x,z)\}$. Now, $t>\phi_0/2$ and for any $x,\ z\in A(x)^c$, one has that $G(x,z)>0$ hence this latter expression is upper-bounded by 1. This leaves the term $\int_{A(x)^c}\frac{1}{2\sqrt{2\pi\phi_0}}e^{-\phi_0 z^2/2}dz$ which is finite. Hence, putting together the above arguments, we have shown that there exists an $M<\infty$ such that for any $s,t\in[\phi_0,1]$, $x\in\mathbb{R}$ one has

$$\Big|\int_{A(x)^c}\varphi(x+z)e^{-sG(x,z)}q_s(z)dz - \int_{A(x)^c}\varphi(x+z)e^{-tG(x,z)}q_t(z)dz\Big|\ /\ V(x) \leq M|s-t|\ ,$$

where we have applied (55).

Turning to the acceptance part of the kernel on $A(x)$, we have

$$\int_{A(x)}\varphi(x+z)[q_s(z)-q_t(z)]dz \leq C_{\phi_0}|\varphi|_V\int_{A(x)} V(x+z)|s-t|\frac{1}{2\sqrt{2\pi\phi_0}}e^{-\phi_0 z^2/2}dz\ .$$

As $V(x+z)\leq V(x)$ on $A(x)$, it follows that the term of interest is upper-bounded by $M|s-t|V(x)$ for some $M<\infty$. Hence the acceptance part of the kernel, divided by V, is upper bounded by $M|s-t|$. In the rejection part of the kernel, we have to control:

$$\varphi(x)\Big[\int_{A(x)^c}[\alpha_t(x,z)-\alpha_t(x,z)]q_t(z)dz+\int_{A(x)^c}[q_t(z)-q_s(z)]\alpha_s(x,z)dz+\int_{A(x)}[q_t(z)-q_s(z)]dz\Big]\ .$$

Now, as φ is controlled by V, we need to consider the continuity of the terms in the bracket. The latter two terms, via (55), are upper-bounded by $M|s-t|$. The first term is upper-bounded by $|s-t|\int_{A(x)^c}G(x,z)e^{-\phi_0/2G(x,z)}q_t(z)dz$ using (58). As $e^{-\phi_0/2G(x,z)}\leq 1$, we can use (31) to complete the argument. □

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. & Thoms J. (2008). A tutorial on adaptive MCMC. *Statist. Comp.*, **18**, 343–373.

[3] Andrieu, C., Breyer, L. & Doucet, A. (2001). Convergence of simulated annealing using Foster-Lyapunov criteria. *J. Appl. Prob.*, **4**, 975–994.

[4] Andrieu, C., Jasra, A., Doucet, A. & Del Moral P. (2011). On non-linear Markov chain Monte Carlo. *Bernoulli*, (to appear).

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

[6] 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.

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

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

[9] 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.

[10] Beskos, A., Pillai, N., Roberts, G. O., Sanz-Serna, J. M. & Stuart A. (2010). Optimal tuning of the hybrid Monte Carlo algorithm. Technical Report, University College London.

[11] 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.

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

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

[14] Cérou, F., Del Moral, P. & Guyader, A. (2011). A non-asymptotic variance theorem for unnormalized Feynman-Kac particle models. *Ann. Inst. Henri Poincare*, (to appear).

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

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

[17] Chopin, N. (2007). Inference and model choice for sequentially ordered hidden Markov models. *J. R. Statist. Soc. B*, **69**, 269–284.

[18] Chopin, N. & Jacob, P. (2011). Free energy Sequential Monte Carlo, application to mixture modelling. *Bayesian Statistics 9*, (to appear).

[19] Chopin, N., Jacob, P. & Papaspiliopoulos, O. (2011). SMC^2: A sequential Monte Carlo algorithm with particle Markov chain Monte Carlo updates. Technical Report, ENSAE.

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

[21] Dean, T. & Dupuis, P. (2009). Splitting for rare event simulation: A large deviations approach to design and analysis. *Stoch. Proc. Appl.*, **119**, 562–587.

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

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

[24] Del Moral, P., Doucet, A. & Jasra, A. (2009). An adaptive sequential Monte Carlo method for approximate Bayesian computation. Technical Report, Imperial College London.

[25] Del Moral, P., Doucet, A. & Jasra, A. (2011). On adaptive resampling procedures for sequential Monte Carlo methods. *Bernoulli*, (to appear).

[26] 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.

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

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

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

[30] Gelman, A. & Meng, X. L.. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statist. Sci.*, **13**, 163–185.

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

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

[33] Jarner, S. F. & Hansen E. (2000). Geometric ergodicity of Metropolis algorithms. *Stoch. Proc. Appl.*, **85**, 341–361.

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

[35] 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.

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

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

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

[39] 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.

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

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

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

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

[44] Roberts, G. O. & Tweedie R. L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, **83**, 95–110.

[45] 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.

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

[47] Schäfer, C. & Chopin, N. (2010). Adaptive Monte Carlo on binary sampling spaces. Technical Report, ENSAE.

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

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

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