

# Sequential Monte Carlo samplers: error bounds and insensitivity to initial conditions

Nick Whiteley\*

## Abstract

This paper addresses finite sample stability properties of sequential Monte Carlo methods for approximating sequences of probability distributions. The results presented herein are applicable in the scenario where the start and end distributions in the sequence are fixed and the number of intermediate steps is a parameter of the algorithm. Under assumptions which hold on non-compact spaces, it is shown that the effect of the initial distribution decays exponentially fast in the number of intermediate steps and the corresponding stochastic error is stable in  $L_p$  norm.

Keywords: non-compact spaces; unbounded functions; sequential Monte Carlo

AMS Subject Classification: 82C80;60J05;

## 1 Introduction

Sequential Monte Carlo (SMC) methods are a class of stochastic algorithms for approximating sequences of probability measures using a population of  $N$  particles. They have been adopted in a variety of application domains, including rare event analysis [8], statistical physics [24], optimal filtering [16, 2] and computational statistics [3, 11]. Various theoretical properties of SMC methods have been studied, and in various contexts, see amongst others [4, 20, 13, 25], and the seminal work of Del Moral [6]. Existing stability results for SMC methods rely on very strong assumptions which typically do not hold on non-compact spaces. This article is concerned with establishing stability properties for a class of SMC algorithms primarily motivated by those of Del Moral et al. [11], under assumptions which do hold on non-compact spaces. The following example indicates a scenario of interest.

---

\*School of Mathematics, University of Bristol, University Walk, Bristol, BS8 1TW, UK

## 1.1 A motivating example

Let  $\pi$  be a distribution on some space  $\mathsf{X}$ , admitting a strictly positive and bounded density with respect to some dominating measure. For ease of presentation, let  $\pi$  also denote this density and let  $\bar{\pi}$  denote the corresponding unnormalised density, i.e. for some  $Z > 0$ ,  $\pi(x) = \bar{\pi}(x)/Z$ . For some  $\underline{\gamma} \in (0, 1)$ , let  $\gamma : [0, 1] \rightarrow [\underline{\gamma}, 1]$  be a non-decreasing Lipschitz function with  $\gamma(0) = \underline{\gamma}$  and  $\gamma(1) = 1$ . For  $n \in \mathbb{N}$  let  $\{G_{n,k}; 0 \leq k < n\}$  be the collection of potential functions on  $\mathsf{X}$  defined by  $G_{n,k}(x) = \bar{\pi}^{\gamma((k+1)/n)}(x)/\bar{\pi}^{\gamma(k/n)}(x)$ . Let  $\{M_{n,k}; 1 \leq k \leq n\}$  be a collection of ergodic Markov kernels also on  $\mathsf{X}$ , where  $M_{n,k}$  admits as its invariant distribution the probability with density proportional to  $\bar{\pi}^{\gamma(k/n)}(x)$  (denoted by  $\pi_{\gamma(k/n)}$ ). For some initial distribution  $\mu$ , consider the sequence of probability distributions  $\{\eta_{n,k}; 1 \leq k \leq n\}$  defined, for a test function  $f$ , by

$$\eta_{n,k}(f) := \frac{\mathbb{E}_{\mu} \left[ \prod_{k=0}^{n-1} G_{n,k}(X_{n,k}) f(X_{n,n}) \right]}{\mathbb{E}_{\mu} \left[ \prod_{k=0}^{n-1} G_{n,k}(X_{n,k}) \right]}, \quad (1)$$

where  $\mathbb{E}_{\mu}$  denotes expectation with respect to the law of the inhomogeneous Markov chain  $\{X_{n,k}; 0 \leq k \leq n\}$ , such that  $X_{0,0} \sim \mu$  and  $X_{n,k+1}|X_{n,k} = x_{n,k} \sim M_{n,k+1}(x_{n,k}, \cdot)$ . The special interest in (1) is that when  $\mu = \pi_{\underline{\gamma}}$ , then  $\eta_{n,n} = \pi$ , for any  $n \geq 1$ . Furthermore, in applications  $Z$  is unknown, the distributions of the form (1) cannot be computed exactly, and one aims to obtain an approximation of  $\pi$  by fixing  $n$  then approximating each of the  $\{\eta_{n,k}; 0 \leq k \leq n\}$  in turn, as follows. Let  $N \in \mathbb{N}$  and let  $\{\zeta_{n,k}; 0 \leq k \leq n\}$  be an inhomogeneous Markov chain, with each  $\zeta_{n,k} = \{\xi_{n,k}^i; 1 \leq i \leq N\}$  an  $N$ -tuple and with each  $\xi_{n,k}^i$  valued in  $\mathsf{X}$ , with  $\{\xi_{n,0}^i; 1 \leq i \leq N\}$  independent and of common distribution  $\mu$ . Given  $\zeta_{n,k}$ , the  $\{\xi_{n,k+1}^i; 1 \leq i \leq N\}$  are independent, with  $\xi_{n,k+1}^i$  drawn from

$$\frac{\sum_{j=1}^N G_{n,k}(\xi_{n,k}^j) M_{n,k+1}(\xi_{n,k}^j, \cdot)}{\sum_{j=1}^N G_{n,k}(\xi_{n,k}^j)}. \quad (2)$$

The particle approximation measure at time step  $k$  is  $\eta_{n,k}^N := \frac{1}{N} \sum_{i=1}^N \delta_{\xi_{n,k}^i}$ , and one takes  $\eta_{n,n}^N(f) := \frac{1}{N} \sum_{i=1}^N f(\xi_{n,n}^i)$  as an approximation of  $\pi(f)$ . The expectation terms in (1) are in the shape of Feynman-Kac formulae, and adopting the terminology of Del Moral [6] throughout the following, a general collection of Markov kernels, potential functions, initial distribution and associated  $\{\eta_{n,k}\}$  are referred to as constituting a Feynman-Kac (FK) model.

The FK model described above has a notable structural characteristic: due to fact that  $\underline{\gamma}$  and  $\pi$  are fixed and  $\gamma(\cdot)$  is continuous, the potential functions each become flat as  $n \rightarrow \infty$  (note that this is not an essential

feature of the general FK models considered by Del Moral [6], nor is it the regime usually considered in the filtering scenario). One might then conjecture that the sequence of measures  $\{\eta_{n,k}; 0 \leq k \leq n\}$  inherit ergodicity properties from the Markov kernels when  $\mu \neq \pi_\gamma$ , and  $\eta_{n,n}(f) - \pi(f)$  goes to zero at some related rate as  $n \rightarrow \infty$ . Perhaps more adventurously, one might further conjecture that this stability property is inherited by the corresponding particle system so that  $\bar{\mathbb{E}}_\mu [|\eta_{n,n}^N(f) - \pi(f)|^p]^{1/p}$  is controlled in  $n$ , and the contribution from the stochastic error diminishes at the usual  $\sqrt{N}$  rate, where  $\bar{\mathbb{E}}_\mu$  denotes expectation w.r.t. the law of the particle process initialised using  $\mu$ .

The results presented in this paper allow it to be shown that this is indeed the case in the context of applications such as [11] on non-compact spaces. The contributions are to establish deterministic stability results for a broad class of FK models, of which the example above is one instance, and to provide  $\mathbb{L}_p$  bounds for the corresponding particle errors.

## 1.2 Summary of results

The present work is built upon generic assumptions about the FK model structure, which can be loosely summarized as follows:

- the Markov kernels  $\{M_{n,k}\}$  are geometrically ergodic, with common Foster-Lyapunov drift function  $V$  and associated constants, and with a common minorization condition
- the potential functions  $\{G_{n,k}\}$  are uniformly bounded above, and are of the form  $G_{n,k}(x) \propto \exp[-\frac{1}{n}U_{n,k}(x)]$ , for  $U_{n,k}$  positive and bounded uniformly over  $n$  and  $k$  in  $V$ -norm .
- $f$  is bounded in  $V$ -norm

Stability properties of general FK semigroups are established in Theorem 1 (section 3). Note here that, in contrast to the above example, no specific form is assumed for  $\{U_{n,k}\}$  or their relations to the invariant distributions of  $\{M_{n,k}\}$ . Theorem 2 (section 4) provides  $\mathbb{L}_p$  error bounds for the corresponding particle systems, under additional assumptions on the drift properties of the particle process.

For the reader's convenience Theorem 3 is now summarized (for the precise statement see section 5), which is an application to the example sketched above. Let  $\mathsf{X} = \mathbb{R}^d$ . Then when  $\pi$  has a sub-exponential density w.r.t. Lebesgue measure with asymptotically regular contours, and when each  $M_{n,k}$  is a random walk Metropolis kernel of invariant distribution  $\pi_{\gamma(k/n)}$ , under a suitable trade-off between  $p \geq 1$ ,  $\alpha \in [0, 1)$  and  $\underline{\gamma}$ , there exists  $\rho < 1$  and constants  $C_1$  and  $C_2$  such that for any  $f$  with  $\|f\|_{V^\alpha} := \sup_x \frac{|f(x)|}{V^\alpha(x)} < +\infty$ ,

any  $N \geq 1$  and  $n \geq 1$ ,

$$\mathbb{E}_\mu \left[ \left| \pi_n^N(f) - \pi(f) \right|^p \right]^{1/p} \leq \|f\|_{V^\alpha} \left( \frac{C_1}{\sqrt{N}} \left( \frac{1 - \rho^n}{1 - \rho} \right) + \rho^n C_2 \mathbb{I}[\mu \neq \pi_\gamma] \right), \quad (3)$$

where  $\pi_n^N := \eta_{n,n}^N$  is the particle approximation as in section 1. The first term on the r.h.s. of (3) corresponds to the stochastic errors from the population of size  $N$  interacting and mutating over  $n$  times steps. The second term on the r.h.s. is a deterministic bias which arises if the initial distribution is mis-specified, and this bias decays exponentially quickly in  $n$ .

### 1.3 Existing work

Despite the fact that the focus here is on models with the “flattening” property, existing SMC stability results cannot be transferred directly to the present scenario under realistic assumptions. They are reviewed below for completeness.

Del Moral [6, Theorem 7.4.4] proved time-uniform  $\mathbb{L}_p$  error bounds, under assumptions which in the present scenario would take the form:

- there exist  $\epsilon_M > 0$ ,  $m \geq 0$  and  $\epsilon_G > 0$  such for all  $n, k$  and  $x, y \in \mathsf{X}$

$$M_{n,k} \dots M_{n,k+m}(x, \cdot) \geq \epsilon_M M_{n,k} \dots M_{n,k+m}(y, \cdot) \quad (4)$$

$$G_{n,k}(x) \geq \epsilon_G G_{n,k}(y), \quad (5)$$

- $f$  is bounded.

Note that these results hold for fully general FK models, not necessarily having the “flattening” property (see also [10, 9, 7, 21, 20, 5, 12] for various results under the same type of assumption as one or both of (4)-(5)). However, these assumptions are very strong. Equation (4) is stronger than uniform ergodicity of the  $m$ -step kernels, and typically is not satisfied for the kernels of interest in [11], such as Metropolis-Hastings kernels on  $\mathbb{R}^d$ . For a toy example which highlights the issue, consider the case that  $\mathsf{X} = \mathbb{R}$  and for some probability measure  $\nu$  on  $\mathsf{X}$  dominated by Lebesgue measure, take  $M_{n,k}(x, \cdot) = a\delta_x(\cdot) + (1 - a)\nu(\cdot)$ . It is easy to check that (4) is violated. Similarly (5) is typically not satisfied in the applications of interest on non-compact spaces and the assumption that  $f$  is bounded is then also rather restrictive. It should be noted however, that with  $a \in (0, 1)$  the Dobrushin co-efficient of this particular  $M_{n,k}$  is in  $(0, 1)$ , and under appropriate

assumptions on  $G_{n,k}$ , stability results such as Theorem 2.6 in [9] then do apply to the underlying model.

Oudjane and Rubenthaler [22] and Heine and Crisan [17] used truncation approaches to obtain stability results for particle filters in expectation over the observation process, without mixing assumptions, but they respectively introduce a rejection step into the particle algorithm (making its computational cost random) and use restrictive assumptions about the state-spaces involved, the hidden Markov model (HMM) and the particle mutation kernels which are not realistic in the scenario of interest. van Handel [25] proved uniform time-average consistency of particle filters under tightness assumptions for a class of bounded functions. Very recently, Del Moral et al. [12] have studied SMC methods in which the resampling step is applied adaptively over time.

The stability of SMC methods has also been studied in the asymptotic regime  $N \rightarrow \infty$ . Chopin [4] established a CLT for a broad class of SMC algorithms, and showed that under the same type of strong mixing assumptions as in (4)-(5), the asymptotic (in  $N$ ) variance associated with the rescaled stochastic error can be bounded uniformly in  $n$ . Jasra and Doucet [19] focused on the asymptotic variance corresponding to the algorithms proposed by Del Moral et al. [11] for unbounded functions. They obtained a bound on the asymptotic variance under realistic geometric ergodicity assumptions which are the same as those considered in the present work. They do not consider the “flattening” regime in their assumptions, and their bounds on the asymptotic variance are not time uniform. Further details of the relationship between the approach of Jasra and Doucet [19] and some of the ideas in this paper are postponed until section 3.

Finally the particle algorithm considered here is derived from the simplest case of the general transition form studied in Section 2.5.3. of [6]. Other selection mechanisms, notably that discussed in [6, Section 11.8.1], have advantages in the  $n \rightarrow \infty$  regime, and it would be interesting to extend the analysis presented here to that context.

The remainder of the paper is structured in the following manner. Section 2 specifies the general form of the FK models in question, associated semigroups and particle systems. Section 3 deals with the deterministic stability of the sequences of measures arising from the FK models.  $\mathbb{L}_p$  error bounds for the stochastic errors of the particle approximations are derived in Section 4. Section 5 applies the results to the case where  $\mathsf{X} = \mathbb{R}^d$  and when the Markov kernels are of the random walk Metropolis variety.

## 2 Definitions

Consider a *state space*  $X$  and an associated countably generated  $\sigma$ -algebra  $\mathcal{B}(X)$ , Let  $\mathcal{P}(X)$  be the collection of probability measures on  $(X, \mathcal{B}(X))$ . For a measure  $\mu$  on  $(X, \mathcal{B}(X))$ , an integral kernel  $M : X \times \mathcal{B}(X) \rightarrow [0, \infty)$  and a function  $f : X \rightarrow \mathbb{R}$ , define  $\mu f := \int_X f(x)\mu(dx)$ ,  $Mf(x) := \int_X M(x, dy)f(y)$  and  $\mu M(\cdot) = \int_X \mu(dx)P(x, \cdot)$ . The function which assigns 1 to every point in  $X$  is also denoted by 1, and the indicator function on a set  $A$  is denoted by  $\mathbb{I}_A$ .

Let  $W : X \rightarrow [1, \infty)$  and  $f : X \rightarrow \mathbb{R}$  be two measurable functions, then define the norm

$$\|f\|_W = \sup_{x \in X} \frac{|f(x)|}{W(x)}.$$

and let  $\mathcal{L}_W = \{f : \|f\|_W < \infty\}$ . Let  $\mu$  be a signed measure on  $(X, \mathcal{B}(X))$ , then define the norm

$$\|\mu\|_W = \sup_{|\phi| \leq W} |\mu(\phi)|.$$

### 2.1 Feynman-Kac models and associated semigroups

Let  $\mu \in \mathcal{P}(X)$  and for each  $n \in \mathbb{N}$  let  $\{M_{n,k}; 1 \leq k \leq n\}$  be a collection of Markov kernels, each kernel acting  $X \times \mathcal{B}(X) \rightarrow [0, 1]$ . Let  $\{G_{n,k}; 0 \leq k \leq n-1\}$  be a collection of  $\mathcal{B}(X)$ -measurable, real-valued, strictly positive and bounded functions on  $X$ . The notation employed below is directly inspired by that of Del Moral [6], with some important modifications, primarily to the indexing, reflecting the scenario of interest in which there is a different FK model for each  $n$ .

Next, for each  $n \in \mathbb{N}$ , let  $\{Q_{n,k}; 1 \leq k \leq n\}$  be the collection of integral kernels defined by

$$Q_{n,k}(x, dy) = G_{n,k-1}(x)M_{n,k}(x, dy).$$

For each  $n$  and  $0 \leq k \leq \ell \leq n$ , let  $M_{n,k:\ell}$  and  $Q_{n,k:\ell}$  be the semigroups associated with the Markov kernels  $\{M_{n,k}\}$  and the kernels  $\{Q_{n,k}\}$ . These semigroups are defined by

$$\begin{aligned} M_{n,k:\ell} &= M_{n,k+1}M_{n,k+2} \dots M_{n,\ell}, & k < \ell \leq n, \\ Q_{n,k:\ell} &= Q_{n,k+1}Q_{n,k+2} \dots Q_{n,\ell}, & k < \ell \leq n, \end{aligned} \tag{6}$$

and  $M_{n,k:k} = Q_{n,k:k} = Id$ ,  $0 \leq k \leq n$ . Next define the collection of probability measures  $\{\eta_{n,k}; 0 \leq k \leq n\}$

by

$$\eta_{n,k}(A) = \frac{\mu Q_{n,0:k}(A)}{\mu Q_{n,0:k}(1)}, \quad A \in \mathcal{B}(X).$$

Let  $\{\Psi_{n,k}; 0 \leq k < n\}$  and  $\{\Phi_{n,k}; 1 \leq k \leq n\}$  be the collections of mappings, each mapping acting from  $\mathcal{P}(X)$  into  $\mathcal{P}(X)$ , defined for any  $\eta \in \mathcal{P}(X)$  by

$$\Psi_{n,k}(\eta)(dx) = \frac{G_{n,k}(x)}{\eta(G_{n,k})} \eta(dx), \quad \Phi_{n,k}(\eta) = \Psi_{n,k-1}(\eta) M_{n,k},$$

and for  $0 \leq k \leq \ell \leq n$  denote by  $\Phi_{n,k:\ell}$  the semigroup associated with the mappings  $\{\Phi_{n,k}\}$ , defined by

$$\Phi_{n,k:\ell} = \Phi_{n,\ell} \circ \Phi_{n,\ell-1} \circ \dots \circ \Phi_{n,k+1}, \quad k < \ell \leq n,$$

and with the convention  $\Phi_{n,k:k} = Id$ . It is straightforward to check that under these definitions, for any  $0 \leq k \leq \ell \leq n$ ,  $\eta \in \mathcal{P}(X)$  and  $A \in \mathcal{B}(X)$ ,

$$\Phi_{n,k:\ell}(\eta)(A) = \frac{\eta Q_{n,k:\ell}(A)}{\eta Q_{n,k:\ell}(1)} \quad (7)$$

and in particular,

$$\Phi_{n,k:\ell}(\eta_{n,k}) = \eta_{n,\ell}.$$

Lastly, let  $\{S_{n,k}; 1 \leq k \leq n\}$  be the collection of Markov kernels, each kernel acting  $X \times \mathcal{B}(X) \rightarrow [0, 1]$ , defined by

$$S_{n,k}(x, A) = \frac{M_{n,k}(Q_{n,k:n}(1) \mathbb{I}_A)(x)}{M_{n,k}(Q_{n,k:n}(1))(x)}.$$

Under these definitions it is straightforward to check that, in line with (7), we have the alternative description of the mapping  $\Phi_{n,k:n}$  in terms of the Markov kernels  $\{S_{n,k}\}$ : for any  $0 \leq k \leq n$  and any  $\eta \in \mathcal{P}(X)$

$$\Phi_{n,k:n}(\eta)(A) = \frac{\eta(Q_{n,k:n}(1) S_{n,k+1} \dots S_{n,n}(A))}{\eta(Q_{n,k:n}(1))}. \quad (8)$$

Consider the following assumption.

(A1) There exists a finite constant  $\overline{G}_X$  such that for all  $n \in \mathbb{N}$  and  $0 \leq k \leq n-1$ ,

$$G_{n,k}(x) \leq \overline{G}_X, \quad \forall x \in X.$$

When assumption (A1) holds, for  $0 \leq k \leq n-1$  define  $\tilde{G}_{n,k} : X \rightarrow (0, 1]$  by

$$\tilde{G}_{n,k}(x) = \frac{G_{n,k}(x)}{\overline{G}_X},$$

and correspondingly,

$$\tilde{Q}_{n,k}(x, dy) = \tilde{G}_{n,k-1}(x) M_{n,k}(x, dy).$$

Also let  $\tilde{Q}_{n,k:\ell}$  denote the semigroup associated with the kernels  $\tilde{Q}_{n,k}$  in the same manner as (6), with the same convention  $\tilde{Q}_{n,k:k} = Id$ . Furthermore under (A1), let  $U_{n,k} : X \rightarrow [0, \infty)$  be defined by  $U_{n,k}(x) = -n \log \tilde{G}_{n,k}(x)$ .

As stated in section 1, the work presented here is primarily motivated by the models and algorithms considered in [11]. However, the “backward” kernel structure which they consider is not introduced here as it is not essential for our purposes. A specific example is given in section 5 and at that point comment on how this fits with the framework of Del Moral et al. [11] is provided.

## 2.2 Particle systems

An explicit construction of the probability space for the particle systems is not provided here, but this can be carried out by canonical methods, see for example [6, Chapter 3] and should be clear from the following symbolic description. Fix  $N \in \mathbb{N}$ ,  $n \in \mathbb{N}$ , and for  $0 \leq k \leq n$  let  $\zeta_{n,k}^{(N)} := \{\xi_{n,k}^{(N,i)}; 1 \leq i \leq N\}$ , where each  $\xi_{n,k}^{(N,i)}$  is valued in  $X$ . Denote  $\eta_{n,k}^N := \frac{1}{N} \sum_{i=1}^N \delta_{\xi_{n,k}^{(N,i)}}$ . For  $1 \leq i \leq N$  and  $1 \leq k \leq n$  let  $\mathcal{F}_{n,k}^{(N,i)} := \sigma(\zeta_{n,0}^{(N)}, \dots, \zeta_{n,k-1}^{(N)}, \xi_{n,k}^{(N,1)}, \dots, \xi_{n,k}^{(N,i)})$  and  $\mathcal{F}_{n,0}^{(N,i)} := \sigma(\xi_{n,0}^{(N,1)}, \dots, \xi_{n,0}^{(N,i)})$ . The generations of the particle system  $\{\zeta_{n,k}^{(N)}; 0 \leq k \leq n\}$  form a non-homogeneous Markov chain: for  $\mu \in \mathcal{P}(X)$ , the law of this chain is denoted



by  $\mathbb{P}_\mu$  and has transitions given in integral form by:

$$\mathbb{P}_\mu \left( \zeta_{n,0}^{(N)} \in dx \right) = \prod_{i=1}^N \mu(dx^i),$$

$$\mathbb{P}_\mu \left( \zeta_{n,k}^{(N)} \in dx \mid \zeta_{n,k-1}^{(N)} \right) = \prod_{i=1}^N \Phi_{n,k}(\eta_{n,k-1}^N)(dx^i), \quad 1 \leq k \leq n,$$

where  $dx = d(x^1, \dots, x^N)$ . Denote by  $\bar{\mathbb{E}}_\mu$  the expectation corresponding to  $\mathbb{P}_\mu$ . It is easy to check that for any  $0 \leq k \leq n$  and  $1 \leq i \leq N$ ,

$$\bar{\mathbb{E}}_\mu \left[ \mathbb{I}_A \left( \xi_{n,k}^{(N,i)} \right) \right] = \bar{\mathbb{E}}_\mu \left[ \eta_{n,k}^N(A) \right], \quad A \in \mathcal{B}(\mathsf{X}). \quad (9)$$

### 3 Stability of the deterministic measures

This section is concerned with stability properties of the sequences of Markov kernels  $\{S_{n,k}\}$  and operators  $\{\Phi_{n,k}\}$ . The approach is to identify non-homogeneous Foster-Lyapunov drift functions and minorization conditions which arise quite naturally from the structure of the FK model and then to employ the quantitative bounds of Douc et al. [14]. Compared to [19], in the present work the general structure of FK models is exploited more directly and in a way which is fruitful when the models satisfy assumption (A4) below. Douc et al. [15, Section 5] identified drift functions and coupling sets for related operators in some specific HMM's; the present work is concerned with a general FK model structure. The first main idea of this section is illustrated by the following assumption and lemma.

**(A2)** There exists  $\lambda \in [0, 1)$ , a function  $V : \mathsf{X} \rightarrow [1, \infty)$ ,  $\varepsilon \in (0, 1]$ ,  $b \in (0, \infty)$ ,  $\mathsf{C} \in \mathcal{B}(\mathsf{X})$  and a probability measure  $\nu \in \mathcal{P}(\mathsf{X})$ , such that

$$\inf_{n \geq 1} \inf_{1 \leq k \leq n} M_{n,k}(x, A) \geq \varepsilon \cdot \nu(A), \quad \forall A \in \mathcal{B}(\mathsf{X}), \forall x \in \mathsf{C}, \quad (10)$$

$$\sup_{n \geq 1} \sup_{1 \leq k \leq n} M_{n,k}V(x) \leq \lambda V(x) + b \mathbb{I}_{\mathsf{C}}(x), \quad \forall x \in \mathsf{X}. \quad (11)$$

**Lemma 1.** Assume (A1)-(A2). Then for each  $n \in \mathbb{N}$  and  $1 \leq k \leq n$ ,

$$S_{n,k}(x, \cdot) \geq \epsilon_{n,k} \nu_{n,k}(\cdot), \quad \forall x \in \mathbf{C}, \quad (12)$$

$$S_{n,k} V_{n,k}(x) \leq \lambda V_{n,k-1}(x) + b_{n,k-1} \mathbb{I}_{\mathbf{C}}(x), \quad \forall x \in \mathbf{X}, \quad (13)$$

where

$$\epsilon_{n,k} := \varepsilon \cdot \nu \left( \tilde{Q}_{n,k:n}(1) \right), \quad b_{n,k} := \frac{b}{\varepsilon \cdot \nu \left( \tilde{Q}_{n,k:n}(1) \right)}, \quad (14)$$

and  $\nu_{n,k} \in \mathcal{P}(\mathbf{X})$  and  $V_{n,k} : \mathbf{X} \rightarrow [1, \infty)$  are defined by

$$\nu_{n,k}(A) := \frac{\nu \left( \tilde{Q}_{n,k:n}(1) \mathbb{I}_A \right)}{\nu \left( \tilde{Q}_{n,k:n}(1) \right)}, \quad A \in \mathcal{B}(\mathbf{X}), \quad (15)$$

$$V_{n,k}(x) := \frac{V(x)}{M_{n,k+1} \left( \tilde{Q}_{n,k+1:n}(1) \right) (x)}, \quad x \in \mathbf{X}, \quad 0 \leq k < n, \quad (16)$$

$V_{n,n} := V$ , with  $\lambda, V, b, \nu, \varepsilon$  and  $\mathbf{C}$  as in (A2).

*Proof.* Noting that  $0 < \tilde{Q}_{n,k:n}(1)(x) \leq 1$  for all  $x \in \mathbf{X}$ , we have

$$\begin{aligned} S_{n,k}(x, A) &= \frac{M_{n,k} \left( \tilde{Q}_{n,k:n}(1) \mathbb{I}_A \right) (x)}{M_{n,k} \left( \tilde{Q}_{n,k:n}(1) \right) (x)} \geq M_{n,k} \left( \tilde{Q}_{n,k:n}(1) \mathbb{I}_A \right) (x) \\ &\geq \varepsilon \cdot \nu \left( \tilde{Q}_{n,k:n}(1) \right) \frac{\nu \left( \tilde{Q}_{n,k:n}(1) \mathbb{I}_A \right)}{\nu \left( \tilde{Q}_{n,k:n}(1) \right)}, \quad \forall x \in \mathbf{C}. \end{aligned}$$

and

$$\begin{aligned} S_{n,k} V_{n,k}(x) &= \frac{M_{n,k} \left( \tilde{G}_{n,k} V \right) (x)}{M_{n,k} \left( \tilde{Q}_{n,k:n}(1) \right) (x)} \leq \lambda \frac{V(x)}{M_{n,k} \left( \tilde{Q}_{n,k:n}(1) \right) (x)} + b \frac{\mathbb{I}_{\mathbf{C}}(x)}{M_{n,k} \left( \tilde{Q}_{n,k:n}(1) \right) (x)} \\ &\leq \lambda V_{n,k-1}(x) + b \frac{\mathbb{I}_{\mathbf{C}}(x)}{\varepsilon \cdot \nu \left( \tilde{Q}_{n,k:n}(1) \right)}, \quad \forall x \in \mathbf{X}, \quad (17) \end{aligned}$$

recalling the convention  $\tilde{Q}_{n,n:n}(1) = 1$ . □

The minorization and drift conditions (12)-(13) pave the way to establishing the stability properties of the operators  $\{\Phi_{n,k}\}$ . In order to move further we introduce assumption (A3) below, which is a stricter version of (A2). The extra structure of (A3) allows the construction in Proposition 1 of bi-variate drift and minorization conditions. Consideration of the case in which the small set arises as a sub-level set of the drift function  $V$  is a standard and generic approach to constructing bi-variate drift conditions from their uni-variate counter-parts, see for example [1, 14]. Furthermore, assumption (A3) allows the level in question to be chosen in a very flexible way, and this property is exploited in Proposition 1 when dealing with the specific structure arising from the kernels  $\{S_{n,k}\}$ . Verification of (A3) in a particular application is provided in section 5. Assumption (A4) allows the bounding of non-homogeneous minorization and drift constants. A generic approach to verifying (A4) is presented at the end of this section, where the connection with flattening property mentioned in the introduction is made more explicit.

**(A3)** There exists  $d_0 \geq 1$ ,  $\lambda \in [0, 1)$ , a function  $V : \mathsf{X} \rightarrow [1, \infty)$ , and for all  $d \in [d_0, +\infty)$ , there exists  $\varepsilon_d \in (0, 1]$ ,  $b_d \in (0, \infty)$  and  $\nu_d \in \mathcal{P}(\mathsf{X})$  such that  $\nu_d(\mathsf{C}_d) > 0$ ,  $\nu_d(V) < +\infty$ ,

$$\inf_{n \geq 1} \inf_{1 \leq k \leq n} M_{n,k}(x, A) \geq \varepsilon_d \cdot \nu_d(A), \quad \forall A \in \mathcal{B}(\mathsf{X}), \forall x \in \mathsf{C}_d, \quad (18)$$

$$\sup_{n \geq 1} \sup_{1 \leq k \leq n} M_{n,k}V(x) \leq \lambda V(x) + b_d \mathbb{I}_{\mathsf{C}_d}(x), \quad \forall x \in \mathsf{X}, \quad (19)$$

where  $\mathsf{C}_d := \{x : V(x) \leq d\}$ .

**(A4)** Whenever (A1) and (A3) hold, for any  $\mu \in \mathcal{P}(\mathsf{X})$  with  $\mu(V) < +\infty$  there exists a positive and finite constant  $K(\mu, \lambda, V, b_{d_0})$  such that

$$\inf_{n \geq 1} \inf_{0 \leq k \leq n} \mu \left( \tilde{Q}_{n,k;n}(1) \right) \geq K(\mu, \lambda, V, b_{d_0}),$$

where  $\lambda$ ,  $V$  and  $d_0$  are as in (A3).

The main result of this section is now presented.

**Theorem 1.** *Assume (A1), (A3) and (A4). Then for any  $\alpha \in (0, 1]$ , there exists  $\rho \in (\lambda, 1)$  and a finite constant  $M$  such that for any  $n \in \mathbb{N}$ , any  $1 \leq k \leq n$ , and any  $\mu, \mu' \in \mathcal{P}(\mathsf{X})$  with  $\mu(V_{n,k-1}^\alpha) < +\infty$ ,  $\mu'(V_{n,k-1}^\alpha) < +\infty$ ,*

$$\|\mu S_{n,k} \dots S_{n,n} - \mu' S_{n,k} \dots S_{n,n}\|_{V^\alpha} \leq M \rho^{n-k+1} [\mu(V_{n,k-1}^\alpha) + \mu'(V_{n,k-1}^\alpha)], \quad (20)$$

where  $V_{n,k-1}$  is as given in equation (16). Consequently, for any  $\mu, \mu' \in \mathcal{P}(X)$  with  $\mu(V^\alpha) < +\infty, \mu'(V^\alpha) < +\infty$  and  $0 \leq k \leq n$ ,

$$\|\Phi_{n,k:n}(\mu) - \Phi_{n,k:n}(\mu')\|_{V^\alpha} \leq M\rho^{n-k} \left[ \frac{\mu(\tilde{G}_{n,k}V^\alpha)}{\mu(\tilde{Q}_{n,k:n}(1))} + \frac{\mu'(\tilde{G}_{n,k}V^\alpha)}{\mu'(\tilde{Q}_{n,k:n}(1))} \right]. \quad (21)$$

The proof of Theorem 1 is postponed. It involves the bi-variate drift functions identified in the following proposition.

**Proposition 1.** *Assume (A1) and (A3) and let  $V, d_0$  and  $\lambda$  be as in (A3). Then for all  $d \geq d_0$ , and all  $\bar{\lambda} \in (\lambda, 1)$  there exists  $\bar{d} \geq d$ , and for each  $n \in \mathbb{N}$ , there exist*

- 1) a collection of functions  $\{\bar{V}_{n,k}; 0 \leq k \leq n\}$ , with each  $\bar{V}_{n,k} : X \times X \rightarrow [1, \infty)$  and  $\bar{V}_{n,n}(x, x') = \frac{1}{2}[V(x) + V(x')]$ ;
- 2) collections  $\{\bar{\epsilon}_{n,k}; 1 \leq k \leq n\}$ , and  $\{\bar{b}_{n,k}; 0 \leq k \leq n-1\}$  depending on  $d$  and  $\bar{d}$ , with each  $\bar{\epsilon}_{n,k} \in (0, 1]$  and each  $\bar{b}_{n,k} \in (0, \infty)$ ;
- 3) a collection of probability measures  $\{\bar{\nu}_{n,k}; 1 \leq k \leq n\}$ , depending on  $\bar{d}$ , with each  $\bar{\nu}_{n,k} \in \mathcal{P}(X)$ ;

such that for  $1 \leq k \leq n$ ,

$$S_{n,k}(x, \cdot) \wedge S_{n,k}(x', \cdot) \geq \bar{\epsilon}_{n,k} \bar{\nu}_{n,k}(\cdot), \quad \forall (x, x') \in \bar{C}_{\bar{d}}, \quad (22)$$

$$S_{n,k}^* \bar{V}_{n,k}(x, x') \leq \bar{\lambda} \bar{V}_{n,k-1}(x, x') + \bar{b}_{n,k-1} \mathbb{I}_{\bar{C}_{\bar{d}}}(x, x'), \quad \forall (x, x') \in X \times X, \quad (23)$$

where  $\{\bar{C}_{\bar{d}} := (x, x') : V(x) \leq \bar{d}, V(x') \leq \bar{d}\}$ ,  $S_{n,k}^* : X \times X \times \mathcal{B}(X \times X) \rightarrow [0, 1]$  is defined by

$$S_{n,k}^*((x, x'), d(y, y')) = \begin{cases} S_{n,k}(x, dy) S_{n,k}(x', dy'), & (x, x') \notin \bar{C}_{\bar{d}}, \\ \bar{R}_{n,k}((x, x'), d(y, y')), & (x, x') \in \bar{C}_{\bar{d}} \end{cases}$$

and  $\bar{R}_{n,k} : X \times X \times \mathcal{B}(X \times X) \rightarrow [0, 1]$  is defined by

$$\bar{R}_{n,k}((x, x), d(y, y')) = \frac{1}{(1 - \bar{\epsilon}_{n,k})^2} (S_{n,k}(x, dy) - \bar{\epsilon}_{n,k} \bar{\nu}_{n,k}(dy)) (S_{n,k}(x', dy') - \bar{\epsilon}_{n,k} \bar{\nu}_{n,k}(dy')).$$

Furthermore,

$$\inf_{n \geq 1} \inf_{1 \leq k \leq n} \bar{\epsilon}_{n,k} > 0, \quad \text{and} \quad \sup_{n \geq 1} \sup_{0 \leq k \leq n-1} \bar{b}_{n,k} < +\infty. \quad (24)$$

*Proof.* Throughout the proof, expressions featuring indices  $n$  and  $k$  hold for all  $n \in \mathbb{N}$  and  $1 \leq k \leq n$ , unless stated otherwise.

Fix  $d \geq d_0$  and  $\bar{\lambda} \in (\lambda, 1)$ . Then let  $\varepsilon_d$ ,  $b_d$  and  $\nu_d$  be the corresponding constants and minorizing measure from (A3). Let  $K(\nu_d, \lambda, V, b_{d_0})$  be the constant of assumption (A4) corresponding to  $\nu_d$  and  $b_d$ , and set

$$\bar{d} := \left[ \frac{b_d}{\varepsilon_d (\bar{\lambda} - \lambda) K(\nu_d, \lambda, V, b_{d_0})} - 1 \right] \vee d.$$

Then for equation (22), under assumption (A3), there exists  $\varepsilon_{\bar{d}}$  and  $\nu_{\bar{d}}$  such that

$$S_{n,k}(x, A) \geq \varepsilon_{\bar{d}} \cdot \nu_{\bar{d}}(\tilde{Q}_{n,k:n}(1) \mathbb{I}_A) = \varepsilon_{\bar{d}} \cdot \nu_{\bar{d}}(\tilde{Q}_{n,k:n}(1)) \frac{\nu_{\bar{d}}(\tilde{Q}_{n,k:n}(1) \mathbb{I}_A)}{\nu_{\bar{d}}(\tilde{Q}_{n,k:n}(1))},$$

for all  $x \in \mathbb{C}_{\bar{d}}$ . Then setting

$$\bar{\epsilon}_{n,k} := \varepsilon_{\bar{d}} \cdot \nu_{\bar{d}}(\tilde{Q}_{n,k:n}(1)), \quad \bar{\nu}_{n,k}(A) := \frac{\nu_{\bar{d}}(\tilde{Q}_{n,k:n}(1) \mathbb{I}_A)}{\nu_{\bar{d}}(\tilde{Q}_{n,k:n}(1))}, \quad A \in \mathcal{B}(\mathbb{X}), \quad (25)$$

establishes (22). The first part of (24) is an immediate consequence of (A4) and (25):

$$\varepsilon_{\bar{d}} \cdot \nu_{\bar{d}}(\tilde{Q}_{n,k:n}(1)) \geq \varepsilon_{\bar{d}} K(\nu_{\bar{d}}, \lambda, V, b_{d_0}). \quad (26)$$

Let  $\{V_{n,k}\}$  be as defined in (16). Consider the collection of bi-variate drift functions  $\{\bar{V}_{n,k}; 0 \leq k \leq n\}$ , with each  $\bar{V}_{n,k} : \mathbb{X} \times \mathbb{X} \rightarrow [1, \infty)$  defined by

$$\bar{V}_{n,k}(x, x') := \frac{1}{2} [V_{n,k}(x) + V_{n,k}(x')]. \quad (27)$$

We now proceed to establish the bi-variate drift condition of equation (23). First, following the same arguments as in the proof of Lemma 1, under (A3),

$$S_{n,k}V_{n,k}(x) \leq \lambda V_{n,k-1}(x) + b_d \frac{\mathbb{I}_{C_d}(x)}{\varepsilon_d \cdot \nu_d \left( \tilde{Q}_{n,k:n}(1) \right)}, \quad \forall x \in \mathcal{X}. \quad (28)$$

From (28), for  $(x, x') \notin \bar{C}_{\bar{d}}$  we have

$$\begin{aligned} & S_{n,k}^* \bar{V}_{n,k}(x, x') \\ & \leq \frac{\lambda}{2} [V_{n,k-1}(x) + V_{n,k-1}(x')] + \frac{b_d}{\varepsilon_d K(\nu_d, \lambda, V, b_{d_0})} \frac{1}{2} [\mathbb{I}_{C_d}(x) + \mathbb{I}_{C_d}(x')] \\ & \leq \frac{\lambda}{2} [V_{n,k-1}(x) + V_{n,k-1}(x')] + (\bar{\lambda} - \lambda) (\bar{d} + 1) \frac{1}{2} [\mathbb{I}_{C_{\bar{d}}}(x) + \mathbb{I}_{C_{\bar{d}}}(x')] \\ & \leq \frac{\lambda}{2} [V_{n,k-1}(x) + V_{n,k-1}(x')] + (\bar{\lambda} - \lambda) \frac{1}{2} [V(x) + V(x')] [\mathbb{I}_{C_{\bar{d}}}(x) + \mathbb{I}_{C_{\bar{d}}}(x')] \\ & \leq \frac{\lambda}{2} [V_{n,k-1}(x) + V_{n,k-1}(x')] + (\bar{\lambda} - \lambda) \frac{1}{2} [V_{n,k-1}(x) + V_{n,k-1}(x')] [\mathbb{I}_{C_{\bar{d}}}(x) + \mathbb{I}_{C_{\bar{d}}}(x')] \\ & \leq \bar{\lambda} \bar{V}_{n,k-1}(x, x'), \end{aligned}$$

where for the first inequality (A4) has been applied, the second inequality is due to the definition of  $\bar{d}$  and the penultimate inequality is due to the definition of  $\bar{V}_{n,k-1}$ . For all  $(x, x') \in \bar{C}_{\bar{d}}$ ,

$$\begin{aligned} & S_{n,k}^* \bar{V}_{n,k}(x, x') \\ & = \bar{R}_{n,k} \bar{V}_{n,k}(x, x') \\ & = \frac{1}{2(1 - \bar{\varepsilon}_{n,k})} [S_{n,k}V_{n,k}(x) + S_{n,k}V_{n,k}(x') - 2\varepsilon_{n,k} \bar{\nu}_{n,k}(V_{n,k})] \\ & \leq \frac{\lambda}{2(1 - \bar{\varepsilon}_{n,k})} [V_{n,k-1}(x) + V_{n,k-1}(x')] + \frac{b_d}{2(1 - \bar{\varepsilon}_{n,k})\varepsilon_d \nu_d \left( \tilde{Q}_{n,k:n}(1) \right)} [\mathbb{I}_{C_d}(x) + \mathbb{I}_{C_d}(x')] \\ & \leq \frac{\lambda \bar{d}}{(1 - \bar{\varepsilon}_{n,k})} \frac{1}{\inf_{x:V(x) \leq \bar{d}} M_{n,k} \left( \tilde{Q}_{n,k:n}(1) \right) (x)} + \frac{b_d}{(1 - \bar{\varepsilon}_{n,k})\varepsilon_d \nu_d \left( \tilde{Q}_{n,k:n}(1) \right)} =: \bar{b}_{n,k-1}, \quad (29) \end{aligned}$$

where equation (28) has been used. This concludes the proof of equation (23). Applying (A4) to the denominator terms in (29) and using (25)-(26) establishes the remaining part of equation (24).  $\square$

*Proof. (Theorem 1).* Let  $d_0$  and  $\lambda$  be as in (A3). Set  $d \geq d_0$ ,  $\bar{\lambda} \in (\lambda, 1)$  and let  $\bar{d}$ ,  $\{\bar{\varepsilon}_{n,k}\}$ ,  $\{\bar{b}_{n,k}\}$ ,  $\{\bar{R}_{n,k}\}$ , and  $\{\bar{V}_{n,k}\}$  be as in Proposition 1. The latter verifies conditions (NS1) and (NS2) of Douc et al. [14]. Consequently [14, Theorem 8] may be applied. For  $\alpha = 1$ , the uniform bounds in equation (24) of

Proposition 1, combined with standard manipulations of the bounds of Douc et al. [14, Theorem 8] (details omitted for brevity) show that there exists a finite constant  $M$  and  $\rho < 1$  such that

$$\|\mu S_{n,k} \dots S_{n,n} - \mu' S_{n,k} \dots S_{n,n}\|_V \leq M \rho^{n-k+1} [\mu(V_{n,k-1}) + \mu'(V_{n,k-1})], \quad (30)$$

Noting that from equation (8),

$$\Phi_{n,k:n}(\mu)(A) = \frac{\mu \left( \tilde{G}_{n,k} M_{n,k+1} \left( \tilde{Q}_{n,k+1:n}(1) \right) S_{n,k+1} \dots S_{n,n}(A) \right)}{\mu \left( \tilde{Q}_{n,k:n}(1) \right)},$$

equation (21) holds due to (30) and the definition of  $V_{n,k-1}$  given in equation (16). For the case  $\alpha \in (0, 1)$ , due to Jensen's inequality and the fact that for any two non-negative reals  $a, b$  and  $\alpha \in [0, 1]$ ,  $(a+b)^\alpha \leq a^\alpha + b^\alpha$ , we have that whenever equation (23) of Proposition 1 holds,

$$\begin{aligned} S_{n,k}^* (\bar{V}_{n,k}^\alpha)(x, x') &\leq (\bar{\lambda} \bar{V}_{n,k-1}(x, x') + \bar{b}_{n,k-1} \mathbb{I}_{\bar{C}}(x, x'))^\alpha \\ &\leq \bar{\lambda}^\alpha \bar{V}_{n,k-1}^\alpha(x, x') + \bar{b}_{n,k-1}^\alpha \mathbb{I}_{\bar{C}}(x, x'), \quad \forall (x, x') \in \mathsf{X} \times \mathsf{X}. \end{aligned}$$

and for  $\bar{V}_{n,k-1}$  given in equation (27),

$$\bar{V}_{n,k-1}^\alpha(x, x') \leq \frac{1}{2^\alpha} [V_{n,k}^\alpha(x) + V_{n,k}^\alpha(x')].$$

The arguments as for the case  $\alpha = 1$  are then repeated essentially replacing  $\bar{V}_{n,k}$ ,  $\bar{\lambda}$ ,  $\bar{b}_{n,k-1}$  by  $\bar{V}_{n,k}^\alpha$ ,  $\bar{\lambda}^\alpha$  and  $\bar{b}_{n,k-1}^\alpha$  respectively, in order to establish equation (20) and thus (21). The details are omitted for brevity.  $\square$

### 3.1 Verifying assumption (A4)

The following lemma illustrates that (A4) can be verified under a generic condition on the decay in  $x$  of the potential functions  $\{G_{n,k}\}$  specified via  $\{U_{n,k}\}$ , relative to the drift function  $V$  of assumption (A3).

**Lemma 2.** *Assume (A1) and (A3). Let  $V$ ,  $d_0$  and  $\lambda$  be as in (A3) and assume*

$$\sup_{n \geq 1} \sup_{0 \leq k \leq n-1} \sup_{x \in \mathsf{X}} \frac{U_{n,k}(x)}{V(x)} < +\infty.$$

*Then there exists a positive, finite constant  $C$  depending only on  $\lambda$  and  $b_{d_0}$  such that for any  $\mu \in \mathcal{P}(\mathsf{X})$  with*

$\mu(V) < +\infty$ ,

$$\inf_{n \geq 1} \inf_{0 \leq k \leq n} \mu \left( \tilde{Q}_{n,k:n}(1) \right) \geq \exp[-C\mu(V)]. \quad (31)$$

*Proof.* Firstly,  $\mu \left( \tilde{Q}_{n,n:n}(1) \right) = \mu(1)$  and by Jensen's inequality,

$$\mu \left( \tilde{Q}_{n,n-1:n}(1) \right) = \mu \left( \tilde{G}_{n,n-1} \right) \geq \exp \left[ -\frac{1}{n} \mu(U_{n,n-1}) \right] \geq \exp \left[ -\mu(V) \|U_{n,n-1}\|_V \right].$$

For  $1 \leq k < n-1$ , by Jensen's inequality,

$$\begin{aligned} \mu \left( \tilde{Q}_{n,k:n}(1) \right) &= \int_{\mathcal{X}^{n-k+1}} \exp \left( -\frac{1}{n} \sum_{\ell=k}^{n-1} U_{n,\ell}(x_\ell) \right) \mu(dx_k) \prod_{\ell=k+1}^n M_{n,\ell}(x_{\ell-1}, dx_\ell) \\ &\geq \exp \left[ -\frac{1}{n} \int_{\mathcal{X}^{n-k+1}} \left( \sum_{\ell=k}^{n-1} U_{n,\ell}(x_\ell) \right) \mu(dx_k) \prod_{\ell=k+1}^n M_{n,\ell}(x_{\ell-1}, dx_\ell) \right] \\ &= \exp \left[ -\frac{1}{n} \mu(U_{n,k}) - \frac{1}{n} \sum_{\ell=k+1}^{n-1} \int_{\mathcal{X}} U_{n,\ell}(x_\ell) \mu M_{n,k:\ell}(dx_\ell) \right]. \end{aligned} \quad (32)$$

Iteration of the drift inequality in (A3) shows that for any  $1 \leq k < \ell < n$ ,

$$\int_{\mathcal{X}} V(x_\ell) M_{n,k:\ell}(x_k, dx_\ell) \leq \lambda^{\ell-k} V(x) + b_{d_0} \sum_{j=0}^{\ell-k-1} \lambda^j. \quad (33)$$

It follows from (33) that

$$\int_{\mathcal{X}} U_{n,\ell}(x_\ell) \mu M_{n,k-1:\ell}(dx_\ell) \leq \|U_{n,\ell}\|_V \int_{\mathcal{X}} V(x_\ell) \mu M_{n,k-1:\ell}(dx_\ell) \leq \mu(V) + b_{d_0} \frac{1}{1-\lambda}$$

which combined with (32) implies the desired result.  $\square$

## 4 $\mathbb{L}_p$ error bounds for the particle measures

Making use of the results of section 3, the following theorem presents an  $\mathbb{L}_p$  bound on the error  $\eta_{n,n}^N(f) - \eta_{n,n}(f)$ , for some possibly unbounded  $f$ . This theorem rests on assumptions about the moments of the mean particle drift,  $\eta_{n,k}^N(V)$ , and a related normalization quantity, which are used in the proof to bound the moments of Martingale increments associated with the particle approximation. Verification of (35) is given in the application of section 5.



**Theorem 2.** Assume (A1), (A3) and (A4). Let  $V$  be as in (A3) and for  $s > 0$  an independent parameter let  $t := \frac{1+s}{s}$ . Let  $p \geq 1$  and  $\alpha \in [0, 1]$  be such that  $\alpha tp \leq 1$  and  $(1+s)p \leq 1$ , and for  $\mu \in \mathcal{P}(X)$  assume

$$\sup_{N \geq 1} \sup_{n \geq 1} \sup_{1 \leq k \leq n} \bar{\mathbb{E}}_{\mu} \left[ \eta_{n,k}^N \left( \tilde{Q}_{n,k:n}(1) \right)^{-(1+s)p} \right] < +\infty, \quad (34)$$

$$\sup_{N \geq 1} \sup_{n \geq 1} \sup_{1 \leq k \leq n} \bar{\mathbb{E}}_{\mu} \left[ \eta_{n,k}^N (V^{\alpha tp}) \right] < +\infty. \quad (35)$$

Then there exists  $\rho < 1$  and a finite constant  $C$  depending on  $\alpha, \mu, V$ , and the constants in (A1), (A3), and (A4) such that for any  $f \in \mathcal{L}_{V^{\alpha}}$ ,  $n \in \mathbb{N}$  and  $N \in \mathbb{N}$ ,

$$\bar{\mathbb{E}}_{\mu} \left[ \left| (\eta_{n,n}^N - \eta_{n,n}) (f) \right|^p \right]^{1/p} \leq C \|f\|_{V^{\alpha}} \left( \frac{1 - \rho^n}{1 - \rho} \right) \frac{1}{\sqrt{N}}. \quad (36)$$

*Proof.* Throughout the proof  $C$  denotes a constant whose value may change on each appearance. Consider the telescoping decomposition

$$(\eta_{n,n}^N - \eta_{n,n}) (f) = \sum_{k=0}^n [\Phi_{n,k:n}(\eta_{n,k}^N) - \Phi_{n,k:n}(\Phi_{n,k}(\eta_{n,k-1}^N))] (f), \quad (37)$$

with the convention  $\Phi_{n,0}(\eta_{n,-1}^N) := \mu$ . For any of the terms in the summation of equation (37), following the approach of Del Moral [6, page 245], we have

$$\begin{aligned} & [\Phi_{n,k:n}(\eta_{n,k}^N) - \Phi_{n,k:n}(\Phi_{n,k}(\eta_{n,k-1}^N))] (f) \\ &= \left[ \frac{\eta_{n,k}^N \tilde{Q}_{n,k:n}}{\eta_{n,k}^N \tilde{Q}_{n,k:n}(1)} - \frac{\Phi_{n,k}(\eta_{n,k-1}^N) \tilde{Q}_{n,k:n}}{\Phi_{n,k}(\eta_{n,k-1}^N) \tilde{Q}_{n,k:n}(1)} \right] (f) \\ &= \frac{1}{\eta_{n,k}^N \tilde{Q}_{n,k:n}(1)} \left[ \eta_{n,k}^N \tilde{Q}_{n,k:n} - \eta_{n,k}^N \tilde{Q}_{n,k:n}(1) \frac{\Phi_{n,k}(\eta_{n,k-1}^N) \tilde{Q}_{n,k:n}}{\Phi_{n,k}(\eta_{n,k-1}^N) \tilde{Q}_{n,k:n}(1)} \right] (f) \\ &= \frac{1}{\eta_{n,k}^N \tilde{Q}_{n,k:n}(1)} [\eta_{n,k}^N - \Phi_{n,k}(\eta_{n,k-1}^N)] \tilde{Q}_{n,k:n}^N (f), \end{aligned} \quad (38)$$

where

$$\tilde{Q}_{n,k:n}^N (f)(x) := \tilde{Q}_{n,k:n}(f)(x) - \tilde{Q}_{n,k:n}(1)(x) \frac{\Phi_{n,k}(\eta_{n,k-1}^N) \tilde{Q}_{n,k:n}(f)}{\Phi_{n,k}(\eta_{n,k-1}^N) \tilde{Q}_{n,k:n}(1)}.$$

From equations (37) and (38), for  $p \geq 1$ ,

$$\begin{aligned}
& \bar{\mathbb{E}}_\mu \left[ \left| (\eta_{n,n}^N - \eta_{m,n}) (f) \right|^p \right]^{1/p} \\
& \leq \sum_{k=0}^n \bar{\mathbb{E}}_\mu \left[ \left| \frac{1}{\eta_{n,k}^N \tilde{Q}_{n,k:n}(1)} [\eta_{n,k}^N - \Phi_{n,k}(\eta_{n,k-1}^N)] \tilde{Q}_{n,k:n}^N(f) \right|^p \right]^{1/p} \\
& \leq \sum_{k=0}^n \bar{\mathbb{E}}_\mu \left[ \left| \frac{1}{\eta_{n,k}^N \tilde{Q}_{n,k:n}(1)} \right|^{(1+s)p} \right]^{1/[(1+s)p]} \bar{\mathbb{E}}_\mu \left[ \left| [\eta_{n,k}^N - \Phi_{n,k}(\eta_{n,k-1}^N)] \tilde{Q}_{n,k:n}^N(f) \right|^{pt} \right]^{1/(tp)}, \quad (39)
\end{aligned}$$

where Minkowski's and Hölder's inequalities have been applied. We next proceed to bound each of the factors in the summands of equation (39).

Denoting

$$[\eta_{n,k}^N - \Phi_{n,k}(\eta_{n,k-1}^N)] \tilde{Q}_{n,k:n}^N(f) = \frac{1}{N} \sum_{i=1}^N T_{n,k}^{(i)},$$

where

$$T_{n,k}^{(i)} := \tilde{Q}_{n,k:n}^N(f)(\xi_{n,k}^{(N,i)}) - \Phi_{n,k}(\eta_{n,k-1}^N) \tilde{Q}_{n,k:n}^N(f),$$

(with the dependence of  $T_{n,k}^{(i)}$  on  $N$  suppressed) we have that for any  $n \in \mathbb{N}$ ,  $1 \leq k \leq n$  and  $1 \leq i \leq N$ ,

$$\bar{\mathbb{E}}_\mu \left[ T_{n,k}^{(i)} \mid \mathcal{F}_{n,k}^{(N,i-1)} \right] = 0,$$

with the convention that  $\mathcal{F}_{n,k}^{(N,0)} = \mathcal{F}_{n,k-1}^{(N,N)}$ . Next, there exists a constant  $C$  such that for any  $p \geq 1$ ,

$$\begin{aligned}
& \bar{\mathbb{E}}_\mu \left[ \left| T_{n,k}^{(i)} \right|^p \right]^{1/p} \\
&= \bar{\mathbb{E}}_\mu \left[ \left| \tilde{Q}_{n,k:n}(1) \left( \xi_{n,k}^{(N,i)} \right) \left[ S_{n,k+1} \cdots S_{n,n}(f) \left( \xi_{n,k}^{(N,i)} \right) \right. \right. \right. \\
&\quad \left. \left. \left. - \frac{\Phi_{n,k} \left( \eta_{n,k-1}^N \right) \left( \tilde{Q}_{n,k:n}(1) S_{n,k+1} \cdots S_{n,n}(f) \right)}{\Phi_{n,k} \left( \eta_{n,k-1}^N \right) \left( \tilde{Q}_{n,k:n}(1) \right)} \right] \right|^p \right]^{1/p} \\
&\leq \rho^{n-k} M \|f\|_{V^\alpha} \bar{\mathbb{E}}_\mu \left[ \left| \tilde{Q}_{n,k:n}(1) \left( \xi_{n,k}^{(N,i)} \right) \left[ V_{n,k}^\alpha \left( \xi_{n,k}^{(N,i)} \right) + \frac{\Phi_{n,k} \left( \eta_{n,k-1}^N \right) \left( \tilde{Q}_{n,k:n}(1) V_{n,k}^\alpha \right)}{\Phi_{n,k} \left( \eta_{n,k-1}^N \right) \left( \tilde{Q}_{n,k:n}(1) \right)} \right] \right|^p \right]^{1/p} \\
&\leq \rho^{n-k} M \|f\|_{V^\alpha} \bar{\mathbb{E}}_\mu \left[ V^{\alpha p} \left( \xi_{n,k}^{(N,i)} \right) \right]^{1/p} \\
&\quad + \rho^{n-k} M \|f\|_{V^\alpha} \bar{\mathbb{E}}_\mu \left[ \left| \tilde{Q}_{n,k:n}(1) \left( \xi_{n,k}^{(N,i)} \right) \frac{\Phi_{n,k} \left( \eta_{n,k-1}^N \right) \left( \tilde{Q}_{n,k:n}(1) V_{n,k}^\alpha \right)}{\Phi_{n,k} \left( \eta_{n,k-1}^N \right) \left( \tilde{Q}_{n,k:n}(1) \right)} \right|^p \right]^{1/p} \\
&\leq \rho^{n-k} M \|f\|_{V^\alpha} \bar{\mathbb{E}}_\mu \left[ V^{\alpha p} \left( \xi_{n,k}^{(N,i)} \right) \right]^{1/p} \\
&\quad + \rho^{n-k} M \|f\|_{V^\alpha} \bar{\mathbb{E}}_\mu \left[ \frac{\Phi_{n,k} \left( \eta_{n,k-1}^N \right) \left( \tilde{Q}_{n,k:n}(1) V_{n,k}^{\alpha p} \right)}{\Phi_{n,k} \left( \eta_{n,k-1}^N \right) \left( \tilde{Q}_{n,k:n}(1) \right)} \bar{\mathbb{E}}_\mu \left[ \tilde{Q}_{n,k:n}(1) \left( \xi_{n,k}^{(N,i)} \right) \middle| \mathcal{F}_{n,k}^{(N,i-1)} \right] \right]^{1/p} \\
&\leq \rho^{n-k} M \|f\|_{V^\alpha} \left( \bar{\mathbb{E}}_\mu \left[ \eta_{n,k}^N (V^{\alpha p}) \right]^{1/p} + \bar{\mathbb{E}}_\mu \left[ \Phi_{n,k} \left( \eta_{n,k-1}^N \right) (V^{\alpha p}) \right]^{1/p} \right) \\
&\leq \rho^{n-k} C \|f\|_{V^\alpha}, \tag{40}
\end{aligned}$$

where Theorem 1, followed by Minkowski's inequality, Jensen's inequality, the exchangeability property of equation (9), Jensen's inequality again and the assumption of equation (35) have been applied. Thus for fixed  $N$ ,  $\left\{ \left( \sum_{j=1}^i T_{n,k}^{(j)}, \mathcal{F}_{n,k}^{(N,i)} \right); 1 \leq i \leq N \right\}$  is a Martingale sequence with increments bounded in  $\mathbb{L}_p$ . It follows that when  $tp \geq 2$ , by the Burkholder-Davis inequality and Minkowski's inequality, there exists a constant  $C$  such that

$$\begin{aligned}
\bar{\mathbb{E}}_\mu \left[ \left| \left[ \eta_{n,k}^N - \Phi_{n,k} \left( \eta_{n,k-1}^N \right) \right] \tilde{Q}_{n,k:n}^N(f) \right|^{tp} \right]^{1/(tp)} &\leq CN^{-1} \bar{\mathbb{E}}_\mu \left[ \left| \sum_{i=1}^N \left( T_{n,k}^{(i)} \right)^2 \right|^{tp/2} \right]^{1/(tp)} \\
&\leq CN^{-1} \left( \sum_{i=1}^N \bar{\mathbb{E}}_\mu \left[ \left| T_{n,k}^{(i)} \right|^{tp} \right]^{2/(tp)} \right)^{1/2},
\end{aligned}$$

and when  $1 < tp < 2$ , using the fact that for any  $a, b \geq 0$  and  $0 \leq r \leq 1$ ,  $(a + b)^r \leq a^r + b^r$ ,

$$\begin{aligned} \bar{\mathbb{E}}_\mu \left[ \left| [\eta_{n,k}^N - \Phi_{n,k}(\eta_{n,k-1}^N)] \tilde{Q}_{n,k:n}^N(f) \right|^{tp} \right]^{1/(tp)} &\leq CN^{-1} \bar{\mathbb{E}}_\mu \left[ \left| \sum_{i=1}^N (T_{n,k}^{(i)})^2 \right|^{tp/2} \right]^{1/(tp)} \\ &\leq CN^{-1} \left( \sum_{i=1}^N \bar{\mathbb{E}}_\mu \left[ (T_{n,k}^{(i)})^{tp} \right] \right)^{1/(tp)}. \end{aligned}$$

Combining with (40), we conclude that there exists a constant  $C$  such that

$$\bar{\mathbb{E}}_\mu \left[ \left| [\eta_{n,k}^N - \Phi_{n,k}(\eta_{n,k-1}^N)] \tilde{Q}_{n,k:n}^N(f) \right|^{tp} \right]^{1/(tp)} \leq \rho^{n-k} C \|f\|_{V^\alpha} N^{-\{tp/2 \wedge (tp-1)\}/(tp)},$$

for all  $n \geq 1$  and  $1 \leq k \leq n$ .

The remaining terms in (39) are treated directly by the assumption of equation (34), and therefore upon returning to (39) we conclude that there exists a constant  $C$  such that

$$\bar{\mathbb{E}}_\mu \left[ |(\eta_{m,n}^N - \eta_{m,n})|^p \right]^{1/p} \leq C \|f\|_{V^\alpha} \frac{1}{\sqrt{N}} \sum_{k=0}^n \rho^{n-k} < C \|f\|_{V^\alpha} \left( \frac{1 - \rho^n}{1 - \rho} \right) \frac{1}{\sqrt{N}},$$

and the result holds. □

## 5 Application

This section is concerned with the case in which  $\mathbf{X} = \mathbb{R}^d$ ,  $\mathcal{B}(\mathbb{R}^d)$  is the corresponding Borel  $\sigma$ -algebra and throughout we consider the following structural definitions and assumptions.

- Let  $\pi \in \mathcal{P}(\mathbf{X})$  be a target distribution admitting a density with respect to Lebesgue measure. Also denote by  $\pi$  its density. In applications of interest, this density will only be known up to a multiplicative constant,  $Z$ , and denote by  $\bar{\pi}$  the unnormalised density, i.e.  $\pi(x) = \bar{\pi}(x)/Z$ ,  $x \in \mathbb{R}^d$ .
- For  $\underline{\gamma} \in (0, 1]$  a constant, let  $\gamma : [0, 1] \rightarrow [\underline{\gamma}, 1]$  be a non-decreasing, Lipschitz function.
- Let  $\{\pi_\gamma; \gamma \in [\underline{\gamma}, 1]\}$  be the family of probability measures defined by

$$\pi_\gamma(A) := \frac{\int_A \bar{\pi}^\gamma(x) dx}{\int_{\mathbf{X}} \bar{\pi}^\gamma(x) dx}, \quad A \in \mathcal{B}(\mathbf{X}).$$

- Let  $q \in \mathcal{P}(X)$  be an increment distribution admitting a density with respect to Lebesgue measure, also denoted by  $q$ . For each  $n \geq 1$  and  $1 \leq k \leq n$ , let  $M_{n,k}$  be a random walk Metropolis (RWM) kernel of invariant distribution  $\pi_{\gamma(k/n)}$  and proposal kernel  $q$ , i.e.

$$M_{n,k}(x, A) = \int_{A-x} \left[ 1 \wedge \frac{\bar{\pi}^{\gamma(k/n)}(x+y)}{\bar{\pi}^{\gamma(k/n)}(x)} \right] q(y) \lambda^{\text{Leb}}(dy) \\ + \delta_x(A) \int_{X-x} \left[ 1 - \left[ 1 \wedge \frac{\bar{\pi}^{\gamma(k/n)}(x+y)}{\bar{\pi}^{\gamma(k/n)}(x)} \right] \right] q(y) \lambda^{\text{Leb}}(dy), \quad A \in \mathcal{B}(X).$$

where for any set  $C$ ,  $C - x := \{z \in X; z + x \in C\}$  (note that in applications it may be of interest to allow  $q$  to depend on  $n$  and  $k$ , for example via  $\gamma(k/n)$ , but for simplicity this issue is not pursued further here).

- Let  $\{G_{n,k}; n \geq 1, 0 \leq k \leq n-1\}$  be a collection of potential functions defined by

$$G_{n,k}(x) = \exp \left[ \frac{1}{n} \log \bar{\pi}(x) \left( \frac{\gamma((k+1)/n) - \gamma(k/n)}{1/n} \right) \right].$$

Consider the following assumptions on the target density  $\pi$  and increment density  $q$ .

- The density  $\pi$  is strictly positive, bounded and has continuous first derivatives such that

$$\lim_{r \rightarrow \infty} \sup_{|x| \geq r} n(x) \cdot \nabla \log \pi(x) = -\infty, \quad \lim_{r \rightarrow \infty} \sup_{|x| \geq r} n(x) \cdot \frac{\nabla \pi(x)}{|\nabla \pi(x)|} < 0, \quad (41)$$

where  $n(x) := x/|x|$ .

- For all  $r > 0$  there exists  $\epsilon_r > 0$  such that

$$|x| \leq r \quad \Rightarrow \quad q(x) \geq \epsilon_r. \quad (42)$$

The assumptions of equations (41)-(42) are standard types of assumptions ensuring geometric ergodicity of RWM kernels [23, 18]. The assumption of equation (42) is stronger than the standard one in [18], but is flexible enough to verify (A3) which involves a family of minorization measures/constants, indexed over a range of levels of  $V$ .

The interest in the specific FK models of this section arises from the choice of the initial distribution  $\mu$  addressed in the following lemma. This FK model corresponds to a particular choice of the “backwards”

kernels in [11], and in the corresponding SMC algorithm the order of the weighting and resampling steps is reversed.

**Lemma 3.** *Consider the operators  $\{\Phi_{n,k}\}$  associated with  $\{G_{n,k}\}$  and  $\{M_{n,k}\}$  of section 5. Then for all  $n \geq 1$  and  $0 \leq k \leq n$ ,  $\Phi_{n,0:k}(\pi_\gamma) = \pi_{\gamma(k/n)}$ .*

*Proof.* Fix  $n$  arbitrarily and suppose the result holds at rank  $0 < k < n$ . Then

$$\begin{aligned} \Phi_{n,0:k+1}(\pi_\gamma)(A) &= \Phi_{n,k+1} \circ \Phi_{n,0:k}(\pi_\gamma)(A) \\ &= \frac{\int \frac{\bar{\pi}^{\gamma((k+1)/n)}(x)}{\bar{\pi}^{\gamma(k/n)}(x)} M_{n,k+1}(A)(x) \pi_{\gamma(k/n)}(dx)}{\int \frac{\bar{\pi}^{\gamma((k+1)/n)}(x)}{\bar{\pi}^{\gamma(k/n)}(x)} \pi_{\gamma(k/n)}(dx)} \\ &= \pi_{\gamma((k+1)/n)}(A), \quad A \in \mathcal{B}(\mathbf{X}), \end{aligned}$$

due to the property that  $M_{n,k}$  is invariant for  $\pi_{\gamma(k/n)}(dx)$ . The proof is complete upon noting that for all  $n$ ,  $\Phi_{n,0:0} = Id$  by convention.  $\square$

We have the following result.

**Theorem 3.** *Consider the collection of FK models specified in section 5. Let  $s > 0$  be an independent parameter and set  $t = \frac{1+s}{s}$ . Let  $\alpha \in [0, 1]$ , and  $p \geq 1$  be such that  $\alpha p t \leq 1$  and  $(1+s)p(1-\gamma)/\gamma < 1$ . Then there exist finite constants  $C_1(p, \mu)$ , and  $C_2(\mu, \pi_\gamma)$  (depending implicitly on  $\pi$  and  $\gamma(\cdot)$ ), and constants  $\beta \in (0, 1)$  and  $\rho \in [0, 1)$ , such that for any  $f \in \mathcal{L}_{V^\alpha}$ ,  $n \geq 1$  and  $N \geq 1$ ,*

$$\mathbb{E}_\mu \left[ |(\pi_n^N - \pi)(f)|^p \right]^{1/p} \leq \|f\|_{V^\alpha} \left( \frac{C_1(p, \mu)}{\sqrt{N}} + \rho^n C_2(\mu, \pi_\gamma) \mathbb{I}[\mu \neq \pi_\gamma] \right),$$

where  $V(x) \propto \pi^{-\beta\gamma}(x)$  and for each  $n$ ,  $\pi_n^N := \eta_{n,n}^N$ .

The proof of Theorem 3 is postponed until after the following proposition regarding the verification of assumptions.

**Proposition 2.** *Consider the setting of section 5. Then (A1), (A3) and (A4) hold.*

*Proof.* As the density  $\pi$  is bounded and  $\gamma(\cdot)$  is Lipschitz, (A1) holds by the mean value theorem. We now turn to the verification of (A3). Various arguments are adopted from Andrieu et al. [1] and the manipulations are fairly standard, but are included here for completeness. The main difference is that we need to explicitly

verify the drift and minorization conditions of (A3) which hold over a range of sub-levels for  $V$  and the proof below involves verification of some assumptions taken as given in [1, Lemma 4].

Firstly, due to the definition of  $\pi_{\underline{\gamma}}$  we have that  $\nabla \log \pi_{\underline{\gamma}}(x) = \underline{\gamma} \nabla \log \bar{\pi}(x) = \underline{\gamma} \nabla \log \pi(x)$  and  $\frac{\nabla \pi(x)}{|\nabla \pi(x)|} = \frac{\bar{\pi}(x) \nabla \log \bar{\pi}(x)}{|\bar{\pi}(x) \nabla \log \bar{\pi}(x)|} \frac{\underline{\gamma} \bar{\pi}^{\underline{\gamma}-1}(x)}{\underline{\gamma} \bar{\pi}^{\underline{\gamma}-1}(x)} = \frac{\nabla \pi_{\underline{\gamma}}(x)}{|\nabla \pi_{\underline{\gamma}}(x)|}$  and so

$$\lim_{r \rightarrow \infty} \sup_{|x| \geq r} n(x) \cdot \nabla \log \pi_{\underline{\gamma}}(x) = -\infty, \quad \lim_{r \rightarrow \infty} \sup_{|x| \geq r} n(x) \cdot \frac{\nabla \pi_{\underline{\gamma}}(x)}{|\nabla \pi_{\underline{\gamma}}(x)|} < 0. \quad (43)$$

In order to verify (A3) we first verify a drift condition for  $M_0(x, dy)$ , defined to be the RWM kernel reversible w.r.t.  $\pi_{\underline{\gamma}}(x)$  with increment density  $q$  as above, i.e.

$$\begin{aligned} M_0(x, A) &= \int_{A-x} \left[ 1 \wedge \frac{\bar{\pi}^{\underline{\gamma}}(x+y)}{\bar{\pi}^{\underline{\gamma}}(x)} \right] q(y) \lambda^{\text{Leb}}(dy) \\ &\quad + \delta_x(A) \int_{\mathbb{X}-x} \left[ 1 - \left[ 1 \wedge \frac{\bar{\pi}^{\underline{\gamma}}(x+y)}{\bar{\pi}^{\underline{\gamma}}(x)} \right] \right] q(y) \lambda^{\text{Leb}}(dy), \quad A \in \mathcal{B}(\mathbb{X}). \end{aligned}$$

Let  $\beta \in (0, 1)$  and define  $V : \mathbb{X} \rightarrow [1, +\infty)$  by

$$V(x) := \frac{\pi^{-\underline{\gamma}\beta}(x)}{\inf_x \pi^{-\underline{\gamma}\beta}(x)}. \quad (44)$$

The results of Jarner and Hansen [18] show that when (43) holds, then for  $M_0$  with increment density  $q$  satisfying equation (42), it holds that  $\lim_{r \rightarrow \infty} \sup_{|x| \geq r} \frac{M_0 V(x)}{V(x)} < 1$ . Thus there exist  $\lambda < 1$  and  $\rho_\lambda < +\infty$  such that

$$|x| \geq \rho_\lambda \Rightarrow \frac{M_0 V(x)}{V(x)} \leq \lambda. \quad (45)$$

Due to (41), there exists  $\epsilon > 0$  and  $\rho_\epsilon > 0$  such that

$$|x| \geq \rho_\epsilon \Rightarrow n(x) \cdot \nabla \log \pi_{\underline{\gamma}}(x) \leq -\epsilon.$$

Now set  $r_0 = \rho_\lambda \vee \rho_\epsilon$  and  $d_0 := \sup_{|x| \leq r_0} V(x)$ . Note that  $d_0 < +\infty$  due to the definition of  $V$  and as the density  $\pi$  is continuous and positive. We now proceed to verify the drift part of (A3).

For any  $d \geq d_0$  let  $C_d := \{x : V(x) \leq d\}$ . We then have

$$\begin{aligned} \sup_{x \in C_d} M_0 V(x) &\leq d \sup_{x \in C_d} \frac{M_0 V(x)}{V(x)} = d \sup_{x \in C_d} \left\{ \int_{A(x)} \frac{\bar{\pi}^{-\gamma\beta}(x+y)}{\bar{\pi}^{-\gamma\beta}(x)} q(y) \lambda^{\text{Leb}}(dy) \right. \\ &\quad \left. + \int_{R(x)} \left[ 1 - \frac{\bar{\pi}^{\underline{\gamma}}(x+y)}{\bar{\pi}^{\underline{\gamma}}(x)} + \frac{\bar{\pi}^{\underline{\gamma}(1-\beta)}(x+y)}{\bar{\pi}^{\underline{\gamma}(1-\beta)}(x)} \right] q(y) \lambda^{\text{Leb}}(dy) \right\} \end{aligned} \quad (46)$$

where  $A(x) := \{y \in X : \pi(x+y) \geq \pi(x)\}$ ,  $R(x) := \{y \in X : \pi(x+y) < \pi(x)\}$ . As each of the ratios in (46) is less than or equal to 1, then we conclude that there exists a constant  $C_b < +\infty$  such that

$$\sup_{x \in C_d} M_0 V(x) \leq d C_b =: b_d. \quad (47)$$

Noting the definition of  $d_0$ , and combining (45) and (47) we obtain

$$\begin{aligned} M_0 V(x) &= M_0 V(x) \mathbb{I}[|x| > r_0] + M_0 V(x) \mathbb{I}[|x| \leq r_0] \\ &\leq \lambda V(x) + M_0 V(x) \mathbb{I}[V(x) \leq d] \\ &\leq \lambda V(x) + b_d \mathbb{I}_{C_d}(x), \end{aligned} \quad (48)$$

for any  $d \geq d_0$  and  $x \in X$ . The arguments of Andrieu et al. [1, Lemma 5] then give  $M_{n,k} V(x) \leq M_0 V(x)$  and from this, (48) and (47), we obtain for any  $d \geq d_0$ ,

$$\sup_{n \geq 1} \sup_{1 \leq k \leq n} M_{n,k} V(x) \leq \lambda V(x) + b_d \mathbb{I}_{C_d}(x),$$

which establishes the drift part of (A3). It remains to show the minorization part. To this end we first show that for any  $d \geq d_0$ ,  $C_d$  is bounded. Recalling the definition of  $r_0$ , we have that for any  $x$  such that  $|x| - r_0 \geq 0$ ,

$$\begin{aligned} \frac{V(x)}{V(n(x)r_0)} &= \left( \frac{\pi(x)}{\pi(n(x)r_0)} \right)^{-\gamma\beta} \\ &= \exp \left[ -\underline{\gamma}\beta (|x| - r_0) \int_0^1 n(x) \cdot \nabla \log \pi(tx + (1-t)n(x)r_0) dt \right] \\ &\geq \exp [\underline{\gamma}\beta (|x| - r_0) \epsilon] \end{aligned}$$

from which we see that  $\lim_{r \rightarrow \infty} \inf_{|x| \geq r} V(x) = +\infty$ , which in turn implies that for all  $d \geq d_0$  there exists  $r_d \geq 0$  such that  $V(x) \leq d \Rightarrow |x| \leq r_d$ . Then for any  $n \geq 1$ ,  $0 \leq k \leq n$  and  $r_d \geq 0$ , whenever  $x \in C_d$



$$\begin{aligned}
M_{n,k}(x, A) &\geq \int_{A-x} \left[ 1 \wedge \frac{\pi(x+y)}{\pi(x)} \right] q(y) \lambda^{\text{Leb}}(dy) \\
&\geq \int_{(A \cap B(0, r_d) - x)} \left[ 1 \wedge \frac{\pi(x+y)}{\pi(x)} \right] q(y) \lambda^{\text{Leb}}(dy) \\
&\geq \epsilon_{2r_d} \frac{\inf_{y \in B(0, r_d)} \pi(y)}{\sup_{y \in B(0, r_d)} \pi(y)} \int_{(A \cap B(0, r_d) - x)} \lambda^{\text{Leb}}(dy) \\
&= \epsilon_{2r_d} \frac{\inf_{y \in B(0, r_d)} \pi(y)}{\sup_{y \in B(0, r_d)} \pi(y)} \left( \int_{B(0, r_d)} \lambda^{\text{Leb}}(dy) \right) \frac{\int_{A \cap B(0, r_d)} \lambda^{\text{Leb}}(dy)}{\int_{B(0, r_d)} \lambda^{\text{Leb}}(dy)} \\
&=: \epsilon_d \cdot \nu_d(A),
\end{aligned}$$

where the third inequality holds due to the properties of  $q$  in (42). As  $\pi$  is strictly positive and continuous,  $V$  is bounded on compact sets and therefore  $\nu_d(V) < +\infty$ . Also,  $\mathbf{C}_d \supseteq \mathbf{C}_{d_0}$  and then due to the definition of  $d_0$ ,  $\nu_d(\mathbf{C}_d) \geq \nu_d(\mathbf{C}_{d_0}) \geq \nu_d(B(0, r_0)) > 0$ . This concludes the verification of (A3).

For (A4), from the definition of  $G_{n,k}$  in this situation, we observe that  $U_{n,k}$  may be taken as

$$U_{n,k}(x) = -\log \bar{\pi}(x) \left( \frac{\gamma((k+1)/n) - \gamma(k/n)}{1/n} \right) + C_\gamma \sup_y \log \bar{\pi}(y)$$

where  $C_\gamma$  is the Lipschitz constant for  $\gamma(\cdot)$  and we observe that  $\sup_{n \geq 1} \sup_{0 \leq k \leq n-1} \sup_{x \in \mathbf{X}} \frac{U_{n,k}(x)}{V(x)} < +\infty$ . Assumption (A4) is then satisfied upon application of Lemma 2.  $\square$

*Proof. (Theorem 3)*

Throughout the proof, we denote by  $C$  a constant whose value may change upon each appearance. Consider the error decomposition

$$\bar{\mathbb{E}}_\mu \left[ |(\pi_n^N - \pi)(f)|^p \right]^{1/p} = \bar{\mathbb{E}}_\mu \left[ |(\eta_{n,n}^N - \eta_{n,n})(f)|^p \right]^{1/p} + |(\eta_{n,n} - \pi)(f)|. \quad (49)$$

Choose  $\beta$  such that  $(1+s)p(1-\underline{\gamma})/(\underline{\gamma}\beta) \leq 1$  and take  $V$  to be defined as in equation (44). By proposition 2, the FK model of section 5 satisfies assumptions (A1), (A3) and (A4). The second term on the r.h.s. of (49) is treated by application of Theorem 1. Noting that by definition,  $\eta_{n,n} = \Phi_{n,0:n}(\mu)$  and by Lemma 3,

$\pi = \Phi_{n,0:n}(\pi_\gamma)$  we obtain from Theorem 1 that there exist constants  $\rho$ ,  $M$  and  $C_2$  such that

$$\begin{aligned} \|\eta_{n,n} - \pi\|_{V^\alpha} &\leq M\rho^n \left[ \frac{\mu(\tilde{G}_{n,k}V^\alpha)}{\mu(\tilde{Q}_{n,k:n}(1))} + \frac{\pi_\gamma(\tilde{G}_{n,k}V^\alpha)}{\pi_\gamma(\tilde{Q}_{n,k:n}(1))} \right] \mathbb{I}[\mu \neq \pi_\gamma] \\ &\leq \rho^n C_2 (\mu, \pi_\gamma) \mathbb{I}[\mu \neq \pi_\gamma]. \end{aligned}$$

where the constant  $C_2$  arises from assumption (A4) applied to the denominator terms and implicitly depends on  $\pi_\gamma$ ,  $V$ , and the constants in (A3).

In order to apply Theorem 2 to the first term on the r.h.s. of (49) it remains to verify the assumptions of equations (34)-(35). We start by addressing the latter. From the definition of  $V$  and due to the assumption that  $\gamma(\cdot)$  is non-decreasing we observe that for all  $x, x' \in \mathsf{X}$ ,

$$[G_{n,k}(x) - G_{n,k}(x')][V(x) - V(x')] \leq 0$$

and therefore as  $G_{n,k}$  and  $V$  are positive, Lemma 4 (in the appendix) shows that for all (possibly random)  $\eta \in \mathcal{P}(\mathsf{X})$ , we have  $\eta(G_{n,k}V) \leq \eta(G_{n,k})\eta(V)$ . Then for any  $n \geq 1$  and  $1 \leq k \leq n$ .

$$\bar{\mathbb{E}}_\mu \left[ \eta_{n,k}^N(V) \middle| \mathcal{F}_{n,k-1}^{(N,N)} \right] \leq \lambda \frac{\eta_{n,k-1}^N(G_{n,k-1}V)}{\eta_{n,k-1}^N(G_{n,k-1})} + b_{d_0} \leq \lambda \eta_{n,k-1}^N(V) + b_{d_0}, \quad (50)$$

where (A3) has been applied with  $d_0$  is defined below equation (45) in the proof of proposition 2. Standard iteration of the particle drift inequality (50) (details omitted for brevity) combined with the fact that  $\{\xi_{n,0}^{(N,i)}; i = 1, \dots, N\}$  are independent and each distributed according to  $\mu$  shows that

$$\sup_{N \geq 1} \sup_{n \geq 1} \sup_{1 \leq k \leq n} \bar{\mathbb{E}}_\mu [\eta_{n,k}^N(V)] < +\infty, \quad (51)$$

and noting that  $\alpha pt \leq 1$ , equation (35) then holds by two applications of Jensen's inequality.

We now turn to the verification of equation (34). From previous considerations we notice that for some finite constant  $C$ ,

$$U_{n,k}(x) = \frac{1}{\beta \underline{\gamma}} \left( \frac{\gamma((k+1)/n) - \gamma(k/n)}{1/n} \right) \log V(x) + C$$

and therefore

$$\begin{aligned}
& \eta_{n,k}^N \left( \tilde{Q}_{n,k:n}(1) \right)^{-1} \\
&= \left[ \int \exp \left( -\frac{1}{n} \sum_{j=k}^{n-1} U_{n,j}(x_j) \right) \eta_{n,k}^N(dx_k) \prod_{j=k+1}^n M_{n,j}(x_{j-1}, dx_j) \right]^{-1} \\
&\leq \exp \left( \frac{1}{n} \sum_{j=k}^{n-1} \int U_{n,j}(x_j) \eta_{n,k}^N M_{n,k:j}(dx_j) \right) \\
&= \exp \left( C + \frac{1}{n} \frac{1}{\beta \underline{\gamma}} \sum_{j=k}^{n-1} \left( \frac{\gamma((j+1)/n) - \gamma(j/n)}{1/n} \right) \int \log V(x_j) \eta_{n,k}^N M_{n,k:j}(dx_j) \right) \\
&\leq \exp \left( C + \frac{1}{n} \frac{1}{\beta \underline{\gamma}} \sum_{j=k}^{n-1} \left( \frac{\gamma((j+1)/n) - \gamma(j/n)}{1/n} \right) \log \left[ \int V(x_j) \eta_{n,k}^N M_{n,k:j}(dx_j) \right] \right) \\
&\leq \exp \left( C + \frac{1 - \gamma(k/n)}{\beta \underline{\gamma}} \log [\eta_{n,k}^N(V)] \right) \\
&\leq \exp(C) [\eta_{n,k}^N(V)]^{\frac{1-\gamma}{\beta \underline{\gamma}}}
\end{aligned}$$

where the penultimate inequality holds due to standard iteration of the drift inequality in (A3). Therefore

$$\bar{\mathbb{E}}_{\mu} \left[ \eta_{n,k}^N \left( \tilde{Q}_{n,k:n}(1) \right)^{-(1+s)p} \right] \leq C \bar{\mathbb{E}}_{\mu} [\eta_{n,k}^N(V)]^q$$

due to Jensen's inequality and where  $q := \frac{(1-\gamma)}{\beta \underline{\gamma}}(1+s)p \leq 1$  by assumption of the theorem. Equation (34) then follows upon combining this with equation (51). This completes the proof.  $\square$

## Acknowledgments

The author thanks Christophe Andrieu for discussions which lead to the consideration of this work and an anonymous referee for helpful comments.

## 6 Appendix

**Lemma 4.** *Let  $f : X \rightarrow (0, \infty)$ ,  $g : X \rightarrow (0, \infty)$  be two measurable functions such that  $|\eta(fg)| < +\infty$ ,  $|\eta(f)| < +\infty$ ,  $|\eta(g)| < +\infty$ , for any  $\eta \in \mathcal{P}(X)$ . Then for any  $\delta \in [0, \infty)$ ,*

$$\eta(fg) \leq (1 + \delta)\eta(f)\eta(g)$$

if

$$\frac{[f(x) - f(x')][g(x) - g(x')]}{[f(x) + f(x')][g(x) + g(x')]} \leq \frac{\delta}{2 + \delta}, \quad \forall (x, x') \in \mathbf{X}^2, \quad (52)$$

and only if

$$\frac{[f(x) - f(x')][g(x) - g(x')]}{[f(x) + f(x')][g(x) + g(x')]} \leq \frac{3\delta}{2 + \delta}, \quad \forall (x, x') \in \mathbf{X}^2.$$

*Proof.* For any  $\eta, f, g$  as specified in the statement and  $\epsilon \in (0, 1)$ , consider the identity:

$$\begin{aligned} & (1 - \epsilon)\eta(fg) - (1 + \epsilon)\eta(f)\eta(g) \\ &= \frac{1}{2} \int_{\mathbf{X}} \int_{\mathbf{X}} ([f(x) - f(x')][g(x) - g(x')] - \epsilon[f(x) + f(x')][g(x) + g(x')]) \eta(dx)\eta(dx') \\ &= \frac{1}{2} \int_{\mathbf{X}^2} ([f(x) - f(x')][g(x) - g(x')] - \epsilon[f(x) + f(x')][g(x) + g(x')]) \eta(dx) \otimes \eta(dx'), \end{aligned}$$

where the final inequality is due to Fubini's theorem, which is applicable under the hypotheses of the lemma.

The sufficiency part then follows directly upon setting

$$\epsilon = \frac{\delta}{2 + \delta} \Leftrightarrow (1 + \delta) = \frac{1 + \epsilon}{1 - \epsilon}.$$

For the necessity part, suppose on the contrary that there exists  $(y, y') \in \mathbf{X}^2$  such that

$$[f(y) - f(y')][g(y) - g(y')] > \frac{3\delta}{2 + \delta} [f(y) + f(y')][g(y) + g(y')],$$

then setting  $\eta = \frac{1}{2} [\delta_y + \delta_{y'}]$  and  $\epsilon = \frac{\delta}{2 + \delta} \geq 0$ , we obtain

$$\begin{aligned}
& \int_{\mathbf{X}} \int_{\mathbf{X}} ([f(x) - f(x')] [g(x) - g(x')] - \epsilon [f(x) + f(x')] [g(x) + g(x')]) \eta(dx) \eta(dx') \\
&= \int_{\mathbf{X}^2} ([f(x) - f(x')] [g(x) - g(x')] - \epsilon [f(x) + f(x')] [g(x) + g(x')]) \frac{1}{4} [\delta_y(dx) \otimes \delta_{y'}(dx') + \delta_{y'}(dx) \otimes \delta_y(dx')] \\
&+ \int_{\mathbf{X}^2} ([f(x) - f(x')] [g(x) - g(x')] - \epsilon [f(x) + f(x')] [g(x) + g(x')]) \frac{1}{4} [\delta_y(dx) \otimes \delta_{y'}(dx') + \delta_{y'}(dx) \otimes \delta_y(dx')] \\
&= -\epsilon [f(y)g(y) + f(y')g(y')] \\
&+ \frac{1}{2} ([f(y) - f(y')] [g(y) - g(y')] - \epsilon [f(y) + f(y')] [g(y) + g(y')]) \\
&\geq \frac{1}{2} ([f(y) - f(y')] [g(y) - g(y')] - 3\epsilon [f(y) + f(y')] [g(y) + g(y')]) > 0,
\end{aligned}$$

which completes the proof. □

*Remark 1.* Note that the sufficient condition is always met, for example, when  $g = \varphi \circ f$  for some positive, strictly decreasing and invertible function  $\varphi$ .

## References

- [1] C. Andrieu, L. A. Breyer, and A. Doucet. Convergence of simulated annealing using Foster-Lyapunov criteria. *Journal of Applied Probability*, 38(4), 2001.
- [2] O. Cappé, E. Moulines, and T. Ryden. *Inference in hidden Markov models*. Springer Series in Statistics. Springer, New York, 2005.
- [3] N. Chopin. A sequential particle filter method for static models. *Biometrika*, 89(3), 2002.
- [4] N. Chopin. Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Annals of Statistics*, 32(6), 2004.
- [5] N. Chopin, P. Del Moral, and S. Rubenthaler. Stability of Feynman Kac formulae with path-dependent potentials. *Stochastic Processes and their Applications*, 121(1):38–60, 2011.
- [6] P. Del Moral. *Feynman-Kac Formulae. Genealogical and interacting particle systems with applications*. Probability and its Applications. Springer Verlag, New York, 2004.
- [7] P. Del Moral and A. Doucet. Particle motions in absorbing medium with hard and soft obstacles. *Stochastic Analysis and Applications*, 22:1175–1207, 2004.

- [8] P. Del Moral and J. Garnier. Genealogical particle analysis of rare events. *The Annals of Applied Probability*, 15(4), 2005.
- [9] P. Del Moral and A. Guionnet. On the stability of interacting processes with applications to filtering and genetic algorithms. *Annales de l'Institut Henri Poincaré (B) Probability and Statistics*, 37(2), 2001.
- [10] P. Del Moral and L. Miclo. Branching and interacting particle systems approximation of Feynman-Kac formulae with applications to non-linear filtering. *Séminaire de Probabilités XXXIV. Lecture Notes in Mathematics*, 1729:1–145, 2000.
- [11] P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society, Series B*, 68(3):411–436, 2006.
- [12] P. Del Moral, A. Doucet, and A. Jasra. On adaptive resampling strategies for sequential Monte Carlo methods. *Bernoulli*, 2011. To appear.
- [13] R. Douc and E. Moulines. Limit theorems for weighted samples with applications to sequential Monte Carlo methods. *Annals of Statistics*, 36(5), 2008.
- [14] R. Douc, E. Moulines, and J.S. Rosenthal. Quantitative bounds on convergence of time-inhomogeneous Markov chains. *The Annals of Applied Probability*, 14(4):1643–1665, 2004.
- [15] R. Douc, E. Moulines, and Y. Ritov. Forgetting of the initial condition in general state-space hidden Markov chain: a coupling approach. *Electronic Journal of Probability*, 14:27–49, 2009.
- [16] A. Doucet, N. De Freitas, and N. Gordon, editors. *Sequential Monte Carlo methods in practice*. Springer, New York, 2001.
- [17] K. Heine and D. Crisan. Uniform approximations of discrete-time filters. *Advances in Applied Probability*, 40(4):979–1001, 2008.
- [18] S.F. Jarner and E. Hansen. Geometric ergodicity of Metropolis algorithms. *Stochastic Processes and their Applications*, 85(2):341–361, February 2000.
- [19] A. Jasra and A. Doucet. Stability of sequential Monte Carlo samplers via the Foster Lyapunov condition. *Statistics and Probability Letters*, 78(17), 2008.
- [20] H.R. Künsch. Recursive Monte Carlo filters: algorithms and theoretical analysis. *Annals of Statistics*, 33(5), 2005.

- [21] F. Le Gland and N. Oudjane. Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filter. *The Annals of Applied Probability*, 14(1):144–187, 2004.
- [22] N. Oudjane and S. Rubenthaler. Stability and uniform particle approximation of nonlinear filters in case of non ergodic signals. *Stochastic analysis and applications*, 23(3):421–448, 2005.
- [23] G.O. Roberts and R. L. Tweedie. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, 83(1):95–120, March 1996.
- [24] M. Rousset and G. Stoltz. Nonequilibrium sampling from equilibrium dynamics. *Journal of Statistical Physics*, 123(6):1251–1272, 2006.
- [25] R. van Handel. Uniform time average consistency of Monte Carlo particle filters. *Stochastic Processes and their Applications*, 119(11), 2009.