QUANTITATIVE APPROXIMATIONS OF EVOLVING PROBABILITY MEASURES AND SEQUENTIAL MARKOV CHAIN MONTE CARLO METHODS

ANDREAS EBERLE AND CARLO MARINELLI

Abstract. We study approximations of evolving probability measures by an interacting particle system. The particle system dynamics is a combination of independent Markov chain moves and importance sampling/resampling steps. Under global regularity conditions, we derive non-asymptotic error bounds for the particle system approximation. In a few simple examples, including high dimensional product measures, bounds with explicit constants of feasible size are obtained. Our main motivation are applications to sequential MCMC methods for Monte Carlo integral estimation.

1. Introduction

1.1. **Evolving probability measures.** Let $(\mu_t)_{t\in[0,\infty)}$ denote a family of mutually absolutely continuous probability measures on a set S. To keep the presentation as simple and non-technical as possible, we assume that S is finite. Motivated by Monte Carlo methods for sequential estimation of expectation values with respect to the probability measures μ_t (see e.g. [5, 9, 10, 19] and references therein), we will recall how to obtain Fokker-Planck type evolution equations on the space of probability measures on S that are satisfied by μ_t, and how to approximate these equations by interacting particle systems. The main purpose of this paper is to bound the error of the particle system approximations by an L^p approach (see Theorems 2.5, 2.6 and 2.10 below).

Sequential Monte Carlo (SMC) methods that combine Markov Chain Monte Carlo (MCMC) and Importance Sampling/Resampling methods to approximate a given sequence (μ_t) of probability measures are used in a variety of applications, see for instance [7, 10, 34] and references therein. There is by now a substantial literature on approximation properties of corresponding particle system discretizations, cf. [5, 9, 14] and the references cited below. Nevertheless, our mathematical understanding of SMC methods is still far more superficial than that of traditional MCMC methods, where, at least for some specific models, sharp bounds for mixing times, approximation errors and dependence on the dimension have been derived. The L^p approach to controlling the approximation error that we propose here is a first step towards more quantitative results that might be useful in particular in studying dimensional dependence. In contrast to most of the literature on SMC methods (see however [14, 35, 36]), we focus on the continuous time case.

Date: July 13, 2011.

2000 *Mathematics Subject Classification.* 65C05, 60J25, 60B10, 47H20, 47D08.

Key words and phrases. Markov Chain Monte Carlo, sequential Monte Carlo, importance sampling, spectral gap, Dirichlet forms, functional inequalities, Feynman-Kac formula.

We would like to thank the anonymous referees for detailed reports and very helpful comments on the first version of this article. This work was partially supported by the Sonderforschungsbereich 611, Bonn. The second-named author also gratefully acknowledges the support of the DAAD.

We assume that the measures are represented in the form

$$\mu_t(x) = \frac{1}{Z_t}\,\exp\left(-\mathcal{U}_t(x)\right)\,\mu_0(x), \qquad t \ge 0, \tag{1.1}$$

where Z_t is a normalization constant, and $(t,x) \mapsto \mathcal{U}_t(x)$ is a given function on $[0,\infty) \times S$ that is continuously differentiable in the first variable. If, for example, $\mathcal{U}_t(x) = t\,\mathcal{U}(x)$ for some function $\mathcal{U} : S \to \mathbb{R}$, then $(\mu_t)_{t\ge 0}$ is the exponential family corresponding to $\mathcal{U}$ and μ_0. Let

$$H_t(x) := -\frac{\partial}{\partial t}\log \mu_t(x) = -\frac{\partial}{\partial t}\log \frac{\mu_t(x)}{\mu_0(x)}$$

denote the negative logarithmic time derivative of the measures μ_t. Note that

$$\mu_t(x) = \exp\left(-\int_0^t H_s(x)\,ds\right)\,\mu_0(x)\,, \tag{1.2}$$

and

$$\langle H_t, \mu_t\rangle = -\frac{d}{dt}\mu_t(S) = 0 \qquad \text{for all } t \ge 0, \tag{1.3}$$

where

$$\langle f, \nu\rangle := \int_S f\,d\nu \;= \sum_{x\in S} f(x)\,\nu(x)$$

denotes the integral of a function $f : S \to \mathbb{R}$ w.r.t. a measure ν on S. In particular,

$$H_t = \frac{\partial}{\partial t}\mathcal{U}_t - \langle \frac{\partial}{\partial t}\mathcal{U}_t, \mu_t\rangle.$$

In the applications we have in mind, the functions $\mathcal{U}_t$ are given explicitly. Hence H_t is known explicitly up to an additive time-dependent constant. The evaluation of this constant, however, would require computing an integral w.r.t. μ_t.

If all the functions H_t, $t \ge 0$, vanish then $\mu_t = \mu_0$ for all $t \ge 0$. In this case the measures are invariant for a Markov transition semigroup $(p_t)_{t\ge 0}$, i.e.,

$$\mu_s p_{t-s} = \mu_t \qquad \text{for any } t \ge s \ge 0,$$

provided the generator $\mathcal{L}$ of $(p_t)_{t\ge 0}$ satisfies $\mu_0\mathcal{L} = 0$, i.e.

$$\sum_{x\in S} \mu_0(x)\mathcal{L}(x,y) = 0 \qquad \text{for any } y \in S.$$

This fact is exploited in Markov Chain Monte Carlo methods for approximating expectation values w.r.t. the measure μ_0. The particle systems studied below can be applied for the same purpose when the measures μ_t are time-dependent.

1.2. **Fokker-Planck equation and particle system approximation.** To obtain approximations of the measures μ_t, we consider generators (Q-matrices) $\mathcal{L}_t$, $t \ge 0$, of a time-inhomogeneous Markov process on S satisfying the detailed balance conditions

$$\mu_t(x)\mathcal{L}_t(x,y) = \mu_t(y)\mathcal{L}_t(y,x) \quad \forall\, t \ge 0,\; x, y \in S. \tag{1.4}$$

For example, $\mathcal{L}_t$ could be the generator of a Metropolis dynamics w.r.t. μ_t, i.e.,

$$\mathcal{L}_t(x,y) \;=\; K_t(x,y)\cdot \min\left(\frac{\mu_t(y)}{\mu_t(x)}, 1\right) \;\text{ for } x \ne y,$$

$\mathcal{L}_t(x,x) = -\sum_{y\neq x} \mathcal{L}_t(x,y)$, where the proposal matrix K_t is a given symmetric transition matrix on S. In the sequel we will use the notation $\mathcal{L}_t^*\mu$ to denote the adjoint action of the generator on a probability measure μ, i.e.,

$$(\mathcal{L}_t^*\mu)(y) := (\mu\mathcal{L}_t)(y) = \sum_{x\in S} \mu(x)\mathcal{L}_t(x,y).$$

By (1.4), $\mathcal{L}_t^*\mu_t = 0$, i.e.,

$$\langle \mathcal{L}_t f, \mu_t\rangle = 0 \qquad \text{for any } f : S \to \mathbb{R} \text{ and } t \geq 0.$$

We fix non-negative constants λ_t, $t \geq 0$, such that $t \mapsto \lambda_t$ is continuous. Since the state space S is finite, the measures μ_t are the *unique* solution of the evolution equation for measures

$$\frac{\partial}{\partial t}\nu_t = \lambda_t\, \mathcal{L}_t^*\nu_t - H_t\nu_t \tag{1.5}$$

with initial condition $\nu_0 = \mu_0$. In general, solutions of (1.5) are not necessarily probability measures, even if ν_0 is a probability measure. Therefore, we consider the equation

$$\frac{\partial}{\partial t}\eta_t \;=\; \lambda_t\, \mathcal{L}_t^*\eta_t \;-\; H_t\eta_t \;+\; \langle H_t, \eta_t\rangle\, \eta_t \tag{1.6}$$

satisfied by the normalized measures $\eta_t = \frac{\nu_t}{\nu_t(S)}$. Note that, by (1.3), μ_t also solves (1.6). Moreover, if η_t is a solution of (1.6), then

$$\nu_t = \exp\left(-\int_0^t \langle H_s, \eta_s\rangle\, ds\right)\, \eta_t$$

is the unique solution of (1.5) with initial condition $\nu_0 = \eta_0$.

The Fokker-Planck equation (1.6) is an evolution equation for probability measures which, in contrast to the unnormalized equation, is not modified by adding constants to the functions H_t. We now introduce interacting particle systems that discretize the evolution equations (1.6) and (1.5). Consider right continuous time-inhomogeneous Markov processes $(X_t^N, \mathbb{P})$, $N \in \mathbb{N}$, with state space S^N and generators at time t given by

$$\begin{aligned}\mathcal{L}_t^N\varphi(x_1,\ldots,x_N) &= \lambda_t \sum_{i=1}^N \mathcal{L}_t^{(i)}\varphi(x_1,\ldots,x_N)\\ &\quad + \frac{1}{N}\sum_{i,j=1}^N (H_t(x_i) - H_t(x_j))^+(\varphi(x^{i\to j}) - \varphi(x)).\end{aligned} \tag{1.7}$$

Here $x = (x_1,\ldots,x_N) \in S^N$ and

$$(x^{i\to j})_k \;=\; \begin{cases} x_k & \text{if } k \neq i,\\ x_j & \text{if } k = i.\end{cases}$$

Moreover, $\mathcal{L}_t^{(i)}$ stands for the operator $\mathcal{L}_t$ applied to the i-th component of x. Thus the components $X_{t,i}^N$, $i = 1,\ldots,N$, of the process X_t^N move like independent Markov processes with generator $\lambda_t\mathcal{L}_t$ and are occasionally replaced by components with a lower value of H_t. Note that to compute the generator (and hence to simulate the Markov process) it is enough to know the functions H_t up to an additive constant.

Discretizations of interacting particle systems of a similar type are widely used in applications, where mostly the time parameter is discrete. Variants appear in the literature under different names, including sequential Monte Carlo methods (e.g. in [10, 18, 19]), population Monte Carlo algorithms [4, 17, 35, 36], Feynman-Kac particle models [6, 9, 14]), particle filters [1, 3, 7]), etc. Theoretical properties of these Monte Carlo methods and, in particular, the asymptotics as $N \to \infty$, have been studied intensively (mostly in discrete time), see e.g. [5, 9] for an overview, and [6, 27] for more recent results. The continuous time case has been investigated in [14, 35, 36].

The Markov processes $(X_t^N, \mathbb{P})$ introduced above are continuous-time analogues of a particular type of sequential Monte Carlo samplers which have been introduced and studied systematically in [10] (cf. also [7, 11, 26, 33]). One major motivation for the use of SMC samplers is the estimation of expectation values with respect to multimodal distributions where traditional MCMC methods fail due to metastability problems. The processes $(X_t^N, \mathbb{P})$ have the additional property that the underlying generator at time t satisfies detailed balance w.r.t μ_t. In this case, the resulting sequential MCMC methods are also related to several multi-level sampling methods, including parallel tempering [22, 25, 31] and the equi-energy sampler [29]. The detailed balance condition is not necessarily required for applications, but it fixes a clear framework that is the foundation for our L^p approach developed below.

It is essentially well-known (see [14]) that if the initial distributions of the Markov processes $(X_t^N, \mathbb{P})$ are the N-fold products π^N of a probability measure π on S, then almost surely, the empirical distributions

$$\eta_t^N = \frac{1}{N} \sum_{i=1}^{N} \delta_{X_{t,i}^N} \tag{1.8}$$

and the reweighted empirical distributions

$$\nu_t^N = \exp\left(- \int_0^t \langle H_s, \eta_s^N \rangle \right) \eta_t^N \tag{1.9}$$

converge to the solutions of the equations (1.6) and (1.5) with initial conditions $\eta_0 = \nu_0 = \pi$, see also Corollary 2.8 below. As a consequence, simulating the Markov process X_t^N with initial distribution μ_0^N yields a Monte Carlo method for approximating sequentially the probability measures μ_t, $t \geq 0$, which can be viewed as a combination of Markov Chain Monte Carlo and Importance Sampling/Resampling.

1.3. **Quantitative convergence bounds.** Our main aim is to quantify more explicitly the approximation properties of the particle systems with initial distribution μ_0^N. There is a substantial literature on asymptotic properties of corresponding particle system approximations, see e.g. [9, 14, 35] and references therein. In particular, a law of large numbers type convergence theorem and a corresponding central limit theorem have been established in [12, 14] for a related particle system approximation, cf. also [36]. A crucial question for algorithmic applications, however, are quantitative bounds on the approximation error

$$\langle f, \eta_t^N \rangle - \langle f, \mu_t \rangle \tag{1.10}$$

for a given function $f : S \to \mathbb{R}$ and fixed N that incorporate some more explicit control of the constants. For example, the dependence of the bounds on the dimension in product models is very relevant.

The central limit theorem in [14] yields bounds for the approximation error (1.10) asymptotically as $N \to \infty$ (at least for a modified particle system). In [36] corresponding non-asymptotic estimates are given but without quantifying the constants. We also refer to [6] for some more recent non-asymptotic estimates under strong mixing conditions in discrete time. In this respect, several important questions still remain open:

- The expression for the asymptotic variance in the central limit theorem derived in [14] is not very explicit, as it involves L^2 norms of an associated Feynman-Kac semigroup. Methods that allow to bound this expression *efficiently* in a general setup and in concrete models have to be developed.
- For applications it is crucial to derive more explicit *non-asymptotic bounds* (i.e. bounds for fixed N), because the asymptotic estimates could be misleading when only a limited number of particles is available. To the best of our knowledge such bounds have been proven so far only under partially restrictive minorization (see [40]) or strong mixing conditions involving constants that are not very explicit, highly dimension-dependent, and far from optimal. In general, tracking the constants in the proof of the CLT in [14] shows that these could be of order up to $\exp \int_0^t \mathrm{osc}(H_s)\, ds$, where $\mathrm{osc}(H_s) := \sup H_s - \inf H_s$ stands for the oscillation of H_s. In nearly all interesting applications this quantity is extremely large. Hence although the existing results give useful indications on scope and limits of SMC methods, the rigorous verification of a given error bound for a realistic number N of particles/replicas is still an open problem in many simple concrete models.
- Dimensional dependence on product spaces is an important issue, cf. [1, 2, 3]. Rigorous results about the dependence on the dimension of error bounds for SMC methods are still missing, and might be out of reach for the existing techniques.

It is well-known from the theory of reversible Markov processes that a convergence analysis based only on total variation estimates and Dobrushin contraction coefficients is possible but it has several drawbacks. In particular, substantial contractivity w.r.t. the total variation norm often takes place only after a certain number of steps (cutoff phenomena, cf. e.g. [15, 16]). This limits the applicability if one is interested in arguments based on single or even infinitesimal time steps. Moreover, minorization conditions that are often imposed in this context are crude and typically dimension dependent. Therefore, in this article we develop the foundations of an alternative approach to establish non-asymptotic bounds for the particle system approximations, which enables us to prove bounds with a reasonable dependence on the dimension for product models, see Example 2 below. The approach we propose is based on a consequent application of L^p estimates instead of uniform estimates for Feynman-Kac propagators. In [20] (cf. also [39]), an L^2 approach has been considered to quantify asymptotic stability properties of the Fokker-Planck equation. When studying the error of particle system approximations, we are forced to leave the L^2 framework and to work with various L^p norms. A key tool are the L^p estimates for Feynman-Kac propagators that have been derived in [21].

1.4. **Outline.** The main results of our work are stated in Section 2. Here we also consider examples where the approximation errors can be quantified explicitly. Section 3 contains the derivation of an explicit formula for the variances of the estimators $\langle f, \nu_t^N \rangle$, see Proposition 2.1 below. This is based on martingale arguments developed in [14]. In Section 4 we apply the formula to prove Theorem 2.5 below, which is a non-asymptotic bound for the variances. Finally, in Section 5 we combine this bound with the results from [21] to prove the bounds in Theorems 2.6 and 2.10 below.

2. Main results

To state our results in detail let us consider the Markov process $(X_t^N, \mathbb{P})$ with initial distribution μ_0^N. To derive error bounds for the particle system approximation it is convenient to consider at first the error for the Monte Carlo estimates based on the reweighted empirical distributions ν_t^N defined in (1.9). Following closely the reasoning in [14], we first note that, by a martingale argument, it can be shown that $\langle f, \nu_t^N\rangle$ is an unbiased estimator of $\langle f, \mu_t\rangle$ for any function $f : S \to \mathbb{R}$ and $t \geq 0$, and an explicit formula for the variance can be given.

2.1. **An expression for the variance.** To state the formula for the variance, we introduce Feynman-Kac type transition operators $q_{s,t}$ related to the dynamics. For $0 \leq s \leq t < \infty$ and a function $f : S \to \mathbb{R}$, let $q_{s,t}f(x)$ denote the unique solution of the backward equation

$$-\frac{\partial}{\partial s} q_{s,t}f = \lambda_s \mathcal{L}_s q_{s,t}f - H_s q_{s,t}f, \qquad s \in [0,t], \tag{2.1}$$

with terminal condition $q_{t,t}f = f$. It can be shown that $q_{s,t}f$ is also the unique solution of the corresponding forward equation

$$\frac{\partial}{\partial t} q_{s,t}f = q_{s,t}(\lambda_t \mathcal{L}_t f - H_t f), \qquad t \in [s,\infty), \tag{2.2}$$

with initial condition $q_{s,s}f = f$. As a consequence, a probabilistic representation of $q_{s,t}$ is given by the Feynman-Kac formula

$$(q_{s,t}f)(x) = \mathbb{E}_{s,x}\big[e^{-\int_s^t H_r(X_r)\,dr} f(X_t)\big] \qquad \text{for all } x \in S, \tag{2.3}$$

where $(X_t)_{t\geq s}$ is a time-inhomogeneous Markov process w.r.t. $\mathbb{P}_{s,x}$ with generator $\mathcal{L}_t$ and initial condition $X_s = x$ $\mathbb{P}_{s,x}$-a.s., see e.g. [23], [24]. The next proposition is an adaptation of results in [14, §3.3] to our slightly modified setting.

Proposition 2.1. *For any $f : S \to \mathbb{R}$,*

$$\begin{aligned} \mathbb{E}\left[\langle f, \nu_t^N\rangle\right] &= \langle f, \mu_t\rangle, \qquad \textit{and} \\ \mathbb{E}\left[\left|\langle f, \nu_t^N\rangle - \langle f, \mu_t\rangle\right|^2\right] &= \frac{1}{N}\operatorname{Var}_{\mu_t}(f) + \frac{1}{N}\int_0^t \mathbb{E}\left[V_{s,t}^N(f)\right] ds\,, \end{aligned}$$

where

$$\begin{aligned} V_{s,t}^N(f) = &- \langle H_s (q_{s,t}f)^2, \nu_s^N\rangle\langle 1, \nu_s^N\rangle - \langle H_s, \nu_s^N\rangle\langle q_{s,t}f^2 - (q_{s,t}f)^2, \nu_s^N\rangle \\ &+ \frac{1}{2}\iint |H_s(z) - H_s(y)|(q_{s,t}f(z) - q_{s,t}f(y))^2\, \nu_s^N(dy)\, \nu_s^N(dz). \end{aligned} \tag{2.4}$$

Here and in the following $\operatorname{Var}_\mu(f) := \langle f^2, \mu\rangle - \langle f, \mu\rangle^2$ stands for the variance of f with respect to the measure μ. Although the reasoning is very close to [14], a complete proof of Proposition 2.1 is given in Section 3 below for the reader's convenience.

Elementary estimates show that the approximation error (1.10) for estimates based on the empirical distributions η_t^N can be controlled by the variance of estimators based on ν_t^N:

Lemma 2.2. *For all functions $f : S \to \mathbb{R}$ and $t \geq 0$ we have*

$$\mathbb{E}\left[\left|\langle f, \eta_t^N\rangle - \langle f, \mu_t\rangle\right|^2\right] \leq 2\operatorname{Var}\left(\langle f, \nu_t^N\rangle\right) + 2\|f - \langle f, \mu_t\rangle\|_{\sup}^2 \operatorname{Var}\left(\langle 1, \nu_t^N\rangle\right) \tag{2.5}$$

and

$$\mathbb{E}\left[\left|\langle f,\eta_t^N\rangle - \langle f,\mu_t\rangle\right|\right] \le \operatorname{Var}\left(\langle f,\nu_t^N\rangle\right)^{1/2} + \sqrt{2}\|f-\langle f,\mu_t\rangle\|_{\sup}\operatorname{Var}\left(\langle 1,\nu_t^N\rangle\right) + \sqrt{2}\operatorname{Var}\left(\langle f,\nu_t^N\rangle\right)^{1/2}\operatorname{Var}\left(\langle 1,\nu_t^N\rangle\right)^{1/2}, \tag{2.6}$$

where $\|g\|_{\sup} := \sup_{x\in S}|g(x)|$ *for any* $g : S\to\mathbb{R}$.

The proof is given in Section 5 below.

Remark 2.3. A very interesting alternative expression for the variance of normalizing constants similar to $\langle 1,\nu_t^N\rangle$ in discrete time has recently been derived in [6].

2.2. **A quantitative variance bound.** Let $p\in[2,\infty[$. Our goal is to prove quantitative bounds for the approximation errors that hold uniformly for all functions $f : S\to\mathbb{R}$ with L^p norm less than one. Because of Lemma 2.2, the errors can be quantified in terms of the variance bounds

$$\varepsilon_t^{N,p} := \sup\left\{\mathbb{E}\left[\left|\langle f,\nu_s^N\rangle - \langle f,\mu_s\rangle\right|^2\right] \,\middle|\, f : S\to\mathbb{R} \text{ s.t. } \|f\|_{L^p(\mu_s)}\le 1,\ s\in[0,t]\right\} \tag{2.7}$$

with $p\in[2,\infty)$. To efficiently bound the quantities $\varepsilon_t^{N,p}$ we apply estimates of L^p-L^q operator norms for the operators $q_{s,t}$. Corresponding estimates are derived systematically in [21]. We first state a general result that bounds the error in terms of the expression (2.11) and appropriate operator norms, see Theorem 2.5 below.

For p, $q\in[2,\infty]$ with $p\le q$, let us consider the operator norms

$$C_{s,t}(p) := \sup_{f\neq 0}\frac{\|q_{s,t}f\|_{L^p(\mu_s)}}{\|f\|_{L^p(\mu_t)}},$$

$$C_{s,t}(p,q) := \sup_{f\neq 0}\frac{\|q_{s,t}f\|_{L^{2r}(\mu_s)}}{\|f\|_{L^p(\mu_t)}}\vee\sup_{f\neq 0}\frac{\|q_{s,t}f\|_{L^p(\mu_s)}}{\|f\|_{L^{p/2}(\mu_t)}}\vee 1,$$

where $r\in[p,\infty]$ is chosen such that $p^{-1} = q^{-1}+r^{-1}$. Moreover, for $\delta>0$, we set

$$\bar{C}_t(p,q,\delta) := \sup_{\tau\in[0,t]}\int_0^{(\tau-\delta)^+}\|H_s\|_{L^q(\mu_s)}C_{s,\tau}(p,q)^2\,ds.$$

We fix a constant $t_0>0$, and set

$$\omega := \sup_{s\in[0,t_0]}\operatorname{osc}(H_s), \tag{2.8}$$

where $\operatorname{osc}(f) := \sup f - \inf f$. Since $H_s = -\frac{\partial}{\partial s}\log\mu_s$, the constant ω controls the logarithmic time change rate of the measures μ_t. Note that

$$\bar{C}_t(p,q,\delta)\le t\,\omega\sup\left\{C_{s,\tau}(p,q)^2 \,\middle|\, s,\tau\in[0,t] \text{ s.t. } \tau\ge s+\delta\right\}.$$

Remark 2.4. Since we assume that the state space is finite, all the constants are finite, but their numerical values can be very large. It is a straightforward consequence of the forward equation (2.2) that

$$\mu_s q_{s,t} = \mu_t,\qquad 0\le s\le t, \tag{2.9}$$

and hence $C_{s,t}(1)=1$. On the other hand, in contrast to Markov transition operators which are contractions on L^∞, the constants $C_{s,t}(\infty)$ can be extremely large in typical applications. Therefore bounds on $C_{s,t}(p)$ are very sensitive to the choice of p, see [21] for details. The constants $C_{s,t}(p,q)$ and $\bar{C}_t(p,q,\delta)$ are related to hyperbound properties and can only be expected to be bounded in a feasible way if $t-s$ and δ, respectively, are not too small.

For a function $f : S \to \mathbb{R}$, set

$$V_{s,t}(f) := -\langle H_s(q_{s,t}f)^2, \mu_s\rangle + \iint |H_s(x)|(q_{s,t}f(y) - q_{s,t}f(x))^2 \, \mu_s(dx)\, \mu_s(dy). \quad (2.10)$$

Our first main result shows that for $p > 4$ the asymptotic (as $N \to \infty$) variance of the estimator $\langle f, \nu_t^N\rangle$ is bounded from above by

$$N^{-1}\Big(\mathrm{Var}_{\mu_t}(f) + \int_0^t V_{s,t}(f)\, ds + \|f\|^2_{L^p(\mu_t)}\Big), \quad (2.11)$$

and, more importantly, it gives a non-asymptotic bound for the mean square error $\mathrm{Var}\big(\langle f, \nu_t^N\rangle\big)$ of the same order:

Theorem 2.5. *Fix $q \in]6, \infty]$ and $p \in]\frac{4q}{q-2}, q[$. Let $N \in \mathbb{N}$ be such that*

$$N \geq 25 \max\big(2, \bar{C}_{t_0}(p, q, \delta), \bar{C}_{t_0}(\tilde{p}, q, \delta)\big),$$

where $\tilde{p}$ is defined by $\tilde{p}^{-1} = q^{-1} + (p/2)^{-1}$ and $\delta := (17\omega)^{-1}$. Then, for $t \in [0, t_0]$,

$$\begin{aligned} N\,\mathbb{E}\left[|\langle f, \nu_t^N\rangle - \langle f, \mu_t\rangle|^2\right] \;\leq\; &\mathrm{Var}_{\mu_t}(f) + \int_0^t V_{s,t}(f)\, ds \\ &+ \left[1 + 7\bar{C}_t(p, q, \delta)\varepsilon_t^{N,p}\right] \|f\|^2_{L^p(\mu_t)}. \end{aligned} \quad (2.12)$$

In particular,

$$\varepsilon_t^{N,p} \leq (2 + v_t(p))N^{-1}\big(1 + 10\bar{C}_t(p, q, \delta)N^{-1}\big) \quad (2.13)$$

where

$$v_t(p) := \sup_{\tau\in[0,t]} \sup_{f\neq 0} \frac{\int_0^\tau V_{s,\tau}(f)\, ds}{\|f\|^2_{L^p(\mu_\tau)}}.$$

The proof is given in Section 4 below. To apply Theorem 2.5 we need bounds for the constants $v_t(p)$ and $\bar{C}_t(p, q, \delta)$. We will now discuss how to derive such bounds from Poincaré and logarithmic Sobolev inequalities in the following particular cases:

a) The Markov processes with generators $\mathcal{L}_t$, $t \geq 0$, have "good" global mixing properties (see §2.3).
b) The state space S can be decomposed into disjoint subsets S_i, $i \in I$, such that $\mathcal{L}_t(x, y) = 0$ for all $t \geq 0$, $x \in S_i$ and $y \in S_j$ with $i \neq j$, and "good" mixing properties hold on each of the subsets S_i (see §2.5).

2.3. **Non-asymptotic bounds from global Poincaré and log Sobolev inequalities.** For $t \geq 0$ and $q \in [1, \infty]$ let us define

$$K_t(q) = \int_0^t \|H_s\|_{L^q(\mu_s)}\, ds.$$

The quantities $K_t(q)$ are a way to control how much the measures μ_s change for $s \in [0, t]$. A rough estimate yields

$$v_t(p) \leq 5K_t(2) \sup\left\{C_{s,\tau}(4)^2 \,\middle|\, 0 \leq s \leq \tau \leq t\right\} \qquad \text{for any } p \geq 4, \quad (2.14)$$

$$\bar{C}_t(p, q, \delta) \leq K_t(q) \sup\left\{C_{s,\tau}(p, q)^2 \,\middle|\, 0 \leq s \leq s + \delta \leq \tau \leq t\right\} \quad \text{for any } q \geq p \geq 1. \quad (2.15)$$

Hence estimates for $v_t(p)$ and $\bar{C}_t(p, q, \delta)$ follow from appropriate L^p-L^q bounds for the Feynman-Kac propagators $q_{s,t}$. In [21], we derive such bounds systematically from

Poincaré and logarithmic Sobolev inequalities. To apply these results let us define the weighted Poincaré and log Sobolev constants

$$A_t := \sup_{f\in\mathcal{S}_0} \frac{-\int H_t f^2\, d\mu_t}{\mathcal{E}_t(f)},$$

$$B_t := \sup_{f\in\mathcal{S}_0} \frac{\left|\int H_t f\, d\mu_t\right|^2}{\mathcal{E}_t(f)},$$

$$\gamma_t := \sup_{f\in\mathcal{S}_1} \frac{\int f^2\, \log|f|\, d\mu_t}{\mathcal{E}_t(f)},$$

where $\mathcal{S}_0 = \big\{f : S \to \mathbb{R}|\ \langle f, \mu_t\rangle = 0,\ f \not\equiv 0\big\}$, $\mathcal{S}_1 = \big\{f : S \to \mathbb{R}|\ \langle f^2, \mu_t\rangle = 1,\ f \not\equiv 1\big\}$, and

$$\mathcal{E}_t(f) \ = \ -\int f\mathcal{L}_t f\, d\mu_t \ = \ \frac{1}{2}\sum_{x,y\in S}(f(y) - f(x))^2\mathcal{L}_t(x, y)\mu_t(x)$$

denotes the Dirichlet form of the self-adjoint operator $\mathcal{L}_t$ on $L^2(S, \mu_t)$. We refer to [37] for background on Poincaré and logarithmic Sobolev inequalities and their applications to estimate L^p contractivity properties of transition semigroups and mixing times of reversible time-homogeneous Markov chains. In [21] we apply similar techniques to derive L^p-L^q bounds for Feynman-Kac propagators. We show that $C_{s,t}(p)$ and $C_{s,t}(p, q)$ are small (in particular less than 2) if the intensities λ_s, $0 \le s \le t$, of MCMC moves are sufficiently large in terms of the constants A_s, B_s and γ_s, respectively. By combining these results with Theorem 2.5 we obtain:

Theorem 2.6. *Fix $t_0 \ge 0$, $q \in]6, \infty[$ and $p \in]\frac{4q}{q-2}, q[$. Suppose that*

$$N \ge 40\, \max(K_{t_0}(q), 1), \qquad \textit{and} \tag{2.16}$$

$$\lambda_s \ge \max\left(\frac{pA_s}{4} + \frac{p(p+3)}{4}\, t_0 B_s\ ,\ \frac{17}{4}a(p, q)\, \omega\gamma_s\right) \qquad \textit{for all } s \in [0, t_0], \tag{2.17}$$

where ω is defined by (2.8) and

$$a(p, q) := \log\max\left(\frac{2r-1}{p-1}, \frac{2p-2}{p-2}, \frac{p-1}{\tilde p-1}, \frac{2\tilde p-2}{\tilde p-2}\right),$$

with $\tilde p$ and r determined by $\tilde p^{-1} = q^{-1} + 2p^{-1}$ and $p^{-1} = q^{-1} + r^{-1}$. Then, for $t \in [0, t_0]$,

$$\varepsilon_t^{N,p} \ \le \ (2 + 8\, K_t(2))\, N^{-1}\big(1 + 16K_t(q)N^{-1}\big). \tag{2.18}$$

Note that the assumptions on p and q guarantee that $\tilde p > 2$, so that $a(p, q)$ is finite. The proof of the theorem is given in Section 5 below.

Remark 2.7. (i) The theorem shows that if the intensities λ_s are large enough, then already a limited number of particles/replicas suffices to obtain reasonable error bounds. In particular, if (2.17) holds, then, by (2.18), a number

$$N \ge \frac{3 + 10K_t(q)}{\alpha}$$

of particles guarantees $\varepsilon_t^{N,p} \le \alpha$ for a given $\alpha \in]0, 1/8[$. In particular, as $\alpha \to 0$, a number of particles of order $O(K_t(q)/\alpha)$ is sufficient to bound the error by α.

(ii) Rough bounds for the constants $K_t(q)$, A_t and B_t for $t \in [0, t_0]$ are given by

$$K_t(q) \le t\omega, \qquad A_t \le C_t^{\mathrm{Poi}}\, \max H_t^-, \qquad B_t \le C_t^{\mathrm{Poi}}\, \mathrm{Var}_{\mu_t}(H_t),$$

where ω is defined by (2.8) and

$$C_t^{\mathrm{Poi}} := \sup_{f\in\mathcal{S}_0} \frac{\int f^2\,d\mu_t}{\mathcal{E}_t(f)}$$

denotes the Poincaré constant, i.e., the inverse spectral gap of the generator $\mathcal{L}_t$. Therefore, assumptions (2.16) and (2.17) in Theorem 2.6 are satisfied if

$$N \geq 40\max(t_0\omega, 1) \tag{2.19}$$

and

$$\lambda_s \geq \max\Big(\frac{p}{4}\big(\max H_s^- + t_0(p+3)\operatorname{Var}_{\mu_s}(H_s)\big)C_s^{\mathrm{Poi}},\ \frac{17}{4}a(p,q)\omega\gamma_s\Big). \tag{2.20}$$

Theorems 2.5 and 2.6 provide non-asymptotic bounds on the variances of the Monte Carlo estimators $\langle f, \nu_t^N\rangle$ that hold uniformly over all functions $f \in L^p(\mu_t)$. One can combine these bounds with (2.12) and (2.6) to obtain more precise non-asymptotic error bounds for the Monte Carlo estimators $\langle f, \nu_t^N\rangle$ and $\langle f, \eta_t^N\rangle$ for a fixed function f:

Corollary 2.8. *Suppose that the assumptions of Theorem 2.6 hold, and let $f \in L^p(\mu_t)$. Then*

$$N\,\mathbb{E}\left[\left|\langle f,\nu_t^N\rangle - \langle f,\mu_t\rangle\right|^2\right] \leq \operatorname{Var}_{\mu_t}(f) + \int_0^t V_{s,t}(f)\,ds + \|f\|^2_{L^p(\mu_t)} + R(t,N)\,\|f\|^2_{L^p(\mu_t)},$$

$$N^{1/2}\,\mathbb{E}\left[\left|\langle f,\eta_t^N\rangle - \langle f,\mu_t\rangle\right|\right] \leq \left(\operatorname{Var}_{\mu_t}(f) + \int_0^t V_{s,t}(f)\,ds + \|f - \langle f,\mu_t\rangle\|^2_{L^p(\mu_t)}\right)^{1/2} + \tilde{R}(t,N)\,\|f - \langle f,\mu_t\rangle\|_{\sup}$$

with explicit constants $R(t,N)$ of order $O(N^{-1})$ and $\tilde{R}(t,N)$ of order $O(N^{-1/2})$.

The proof is given in Section 5 below.

2.4. **Scope and Examples.** Summarizing our results, we make the following observations: the derived error bounds of a given size for the particle system approximation rely on the following quantities:

(i) A uniform upper bound on the oscillations of the logarithmic time derivatives $H_t = -\frac{\partial}{\partial t}\log\mu_t$.
(ii) A minimal intensity λ_t of MCMC moves. A lower bound for the required intensity can be given in terms of the constants A_t, B_t and γ_t, or alternatively in terms of ω, C_t^{Poi} and γ_t.
(iii) A minimal number N of particles. On a time interval of length t_0, a number of particles of order $O(\omega t_0\alpha^{-1})$ is sufficient to bound the error $\varepsilon_{t_0}^{N,p}$ by α (provided λ_t is large enough).

We now illustrate range and limits of applicability of the results in two examples. The first is a simple one-dimensional example, while the second discusses the dimensional dependence of the estimates in the case of product measures.

Example 1. Moving Gaussians – one dimensional case. Suppose that $S = \{a, a+1, \ldots, a+\Delta-1\}$ for some $a \in \mathbb{Z}$ and $\Delta \in \mathbb{N}$, and $(\mu_t)_{t\geq 0}$ are probability measures on S such that

$$\mu_t(x) \propto \exp\Big(-\frac{(x-m_t)^2}{2\sigma_t^2}\Big), \qquad x \in S.$$

We assume that $t \mapsto m_t$ and $t \mapsto \sigma_t$ are continuously differentiable functions such that $\sigma_t \in]0,\infty[$ and $m_t \in [a, a+\Delta-1]$ for all $t \geq 0$. Moreover, we assume that the Markov chain moves are given by a Random Walk Metropolis dynamics (in continuous time), that is,

$$\mathcal{L}_t(x,y) = \begin{cases} \frac{1}{2}\min\left(\frac{\mu_t(y)}{\mu_t(x)}, 1\right), & \text{if } |y-x| = 1, \\ 0, & \text{if } |y-x| \geq 1. \end{cases}$$

In this case, the following upper bounds for C_t^{Poi} and γ_t hold (see the Appendix):

$$C_t^{\mathrm{Poi}} \leq 30\big((\sigma_t \wedge \Delta) \vee 2\big)^2 \tag{2.21}$$

$$\gamma_t \leq 300\frac{\Delta^2}{(\sigma_t \wedge 1)^2} + 300\big((\sigma_t \wedge \Delta) \vee 2\big)^2 \log \Delta \tag{2.22}$$

It can be shown that the upper bound for C_t^{Poi} is of the correct order in σ_t and Δ. The upper bound for γ_t could be improved, but γ_t is always bounded from below by a positive multiple of $(\Delta/\sigma_t)^2$. Our results can be applied in the following way. For $t \geq 0$ and x, $y \in S$ we have

$$\begin{aligned} H_t(x) - H_t(y) &= \frac{\partial}{\partial t}\Big(\frac{(x-m_t)^2}{\sigma_t^2} - \frac{(y-m_t)^2}{\sigma_t^2}\Big) \\ &= -\frac{\sigma_t'}{\sigma_t}\frac{(x-y)(x+y-2m_t)}{\sigma_t^2} - m_t'\frac{x-y}{\sigma_t^2} \\ &\leq \Big(2\frac{|\sigma_t'|}{\sigma_t} + \frac{|m_t'|}{\Delta}\Big)\frac{\Delta^2}{\sigma_t^2}. \end{aligned} \tag{2.23}$$

Therefore, if we choose the time scale in such a way that the condition

$$2\frac{|\sigma_t'|}{\sigma_t} + \frac{|m_t'|}{\Delta} \leq \frac{\sigma_t^2}{\Delta^2} \qquad \forall t \in [0, t_0] \tag{2.24}$$

is satisfied, then

$$\omega = \sup_{t \in [0,t_0]} \mathrm{osc}(H_t) \leq 1.$$

Condition (2.24) is an upper bound on the relative change rates of the parameters σ_t and m_t. Note that if Δ is large compared to σ_t, then only small change rates are possible. The reason is that in this case the Gaussian measure μ_t changes too rapidly in the tails, so that our arguments break down.

Assuming (2.24), Theorem 2.6 and Remark 2.7(iii) imply that

$$\varepsilon_t^{N,p} \leq (2+8t)N^{-1}(1+16N^{-1}),$$

provided $N \geq 40(t_0 \vee 1)$, and (2.20) holds with $\omega = 1$, $\max H_s^-$ and $\mathrm{Var}_{\mu_s}(H_s)$ bounded by $\omega = 1$, and C_s^{Poi}, γ_s replaced by the upper bounds in (2.21), (2.22). If $(\sigma_t \wedge 1)/\Delta$ is not too small, this yields reasonably sized (although far from optimal) lower bounds on λ_t and N. On the other hand, if $\sigma_t/\Delta \to 0$, then the upper bounds in both (2.23) and (2.22) degenerate drastically.

Example 2. Product measures – dependence on the dimension. In our second example we study the dependence of (2.19), (2.20) and (2.18) on the dimension in the case when the evolving measures are all product measures. Suppose that

$$S = \prod_{i=1}^{d} S_i, \qquad \mu_t = \bigotimes_{i=1}^{d} \mu_t^{(i)},$$

with probability measures $\mu_t^{(i)}$, $t \geq 0$, $i = 1, \dots, d$, on finite sets S_i such that $t \mapsto \mu_t^{(i)}(x)$ is continuously differentiable and strictly positive for all $1 \leq i \leq d$ and $x \in S_i$. In this case one has

$$H_t(x) = \sum_{i=1}^{d} H_t^{(i)}(x_i),$$

where H_t and $H_t^{(i)}$ denote the negative logarithmic time derivatives of the measures μ_t and $\mu_t^{(i)}$, respectively. If we assume

$$\operatorname{osc}(H_t^{(i)}) \leq 1 \qquad \forall t \in [0, t_0],\ i = 1, \dots, d,$$

then

$$\omega = \sup_{t \in [0,t_0]} \operatorname{osc}(H_t) \leq d, \tag{2.25}$$

and

$$\operatorname{Var}_{\mu_t}(H_t) = \sum_{i=1}^{d} \operatorname{Var}_{\mu_t^{(i)}}(H_t^{(i)}) \leq d.$$

Now suppose that

$$\mathcal{L}_t(x, y) = \sum_{i=1}^{d} \mathcal{L}_t^{(i)}(x_i, y_i)$$

for generators $\mathcal{L}_t^{(i)}$, $t \geq 0$, $i = 1, \dots, d$, of time-inhomogeneous Markov processes on S_i, i.e. $\mathcal{L}_t$ is the generator of the product dynamics on S with component generators $\mathcal{L}_t^{(i)}$. It is well known that $\mathcal{L}_t$ satisfies Poincaré and logarithmic Sobolev inequalities with constants

$$C_t^{\mathrm{Poi}} = \max_{i=1,\dots,d} C_t^{\mathrm{Poi},(i)}, \qquad \gamma_t = \max_{i=1,\dots,d} \gamma_t^{(i)},$$

respectively, where $C_t^{\mathrm{Poi},(i)}$ and $\gamma_t^{(i)}$ are the Poincaré and logarithmic Sobolev constants for the generators $\mathcal{L}_t^{(i)}$. In particular, if the component generators $\mathcal{L}_t^{(i)}$ satisfy Poincaré and logarithmic Sobolev inequalities with constants independent of i, then $\mathcal{L}_t$ satisfies the corresponding inequalities with the same constants – independently of the dimension d. Therefore, in this case, the values of N and λ_s required to satisfy conditions (2.19) and (2.20) are of order $O(d)$. Hence both the number of particles/replicas and the intensity of MCMC moves required are of order $O(d)$. Since simulating from the product dynamics also requires $O(d)$ steps, the total effort to keep track of the evolving product measures up to a given precision is of order $O(d^3)$.

Remark 2.9 (Independent particles)**.** We compare briefly with the particle dynamics without importance sampling/resampling, i.e., when the second summand is omitted in the definition (1.7) of the generator $\mathcal{L}_t^N$. In this case, the particles/replicas move independently according to the time-inhomogeneous Markovian dynamics with generators $\mathcal{L}_t$, $t \geq 0$. Hence the positions of the particles at time t are independent random variables with distribution $\tilde{\mu}_t = \mu_0 p_{0,t}$, where $p_{s,t}$, $0 \leq s \leq t$, is the time-inhomogeneous transition function. A corresponding discrete-time dynamics is used for example in the classical simulated annealing algorithm (see e.g. [13, 30]). Since in general $\tilde{\mu}_t \neq \mu_t$, the empirical distribution of the independent particle system is an asymptotically biased estimator for μ_t. However, under strong mixing conditions as imposed above, the difference between $\tilde{\mu}_t$ and μ_t, and hence the asymptotic bias, will be small. Therefore it is possible that, for fixed N, the empirical distribution of the independent particles process is a better estimate for μ_t than η_t^N. On the other hand, if the mixing properties break down, the bias of the independent particles estimator will not be small, whereas the empirical measures ν_t^N and η_t^N may still be suitable estimators. This will be demonstrated now in a particular case.

2.5. **Non-asymptotic bounds from local estimates.** With suitable modifications the above analysis can also be applied to derive bounds when good mixing properties hold only locally. As an illustration, we consider another extreme case in which the state space is decomposed into several components that are not connected by the underlying Markovian dynamics. Suppose that

$$S = \bigcup_{i \in I} S_i,$$

is a decomposition of S into disjoint non-empty subsets S_i, $i \in I$, such that

$$\mathcal{L}_t(x, y) \;=\; 0 \qquad \text{for any } t \geq 0,\; x \in S_i \text{ and } y \in S_j \text{ with } i \neq j.$$

Let $\mu_t^i := \mu_t(\cdot|S_i)$ denote the measure μ_t conditioned by S_i. Then we can apply the arguments above with the L^p norm replaced by the stronger norm

$$\|f\|_{\widetilde{L^p(\mu_t)}} \;:=\; \max_{i \in I} \|f\|_{L^p(S_i, \mu_t^i)}.$$

Since Hölder's inequality and related estimates hold for these modified L^p norms as well, the assertion of Theorem 2.5 still remains true if $\varepsilon_t^{N,p}$ is replaced by

$$\tilde{\varepsilon}_t^{N,p} := \sup \left\{ \mathbb{E}\left[\left| \langle f, \nu_s^N \rangle - \langle f, \mu_s \rangle \right|^2 \right] \;\middle|\; f : S \to \mathbb{R} \text{ s.t. } \|f\|_{\widetilde{L^p(\mu_s)}} \leq 1,\; s \in [0, t] \right\},$$

and the constants $C_{s,t}(p, q)$ and $\bar{C}_t(p, q, \delta)$ are defined w.r.t. the modified L^p and L^q norms as well. Moreover, the representations (1.2) and (1.3) hold for μ_t^i in place of μ_t if H_t is replaced by

$$H_t^i \;:=\; H_t - \langle H_t, \mu_t^i \rangle .$$

Let A_t^i, B_t^i and γ_t^i denote the Poincaré and logarithmic Sobolev constants defined as above but with S, μ_t and H_t replaced by S_i, μ_t^i and H_t^i, respectively. Let us also set

$$\tilde{A}_t := \max_{i \in I} A_t^i, \qquad \tilde{B}_t := \max_{i \in I} B_t^i, \qquad \tilde{\gamma}_t := \max_{i \in I} \gamma_t^i,$$

$$\tilde{K}_t(q) \;:=\; \int_0^t \|H_s\|_{\widetilde{L^q(\mu_s)}}\, ds, \qquad \text{and}$$

$$\tilde{M}_t \;:=\; \max_{i \in I} \sup_{0 \leq r \leq s \leq t} \frac{\mu_s(S_i)}{\mu_r(S_i)} .$$

Then, by estimating L^p norms separately on each component, we can prove the following extension of Theorem 2.6:

Theorem 2.10. *Fix $t_0 \geq 0$, $q \in]6, \infty[$ and $p \in]\frac{4q}{q-2}, q[$. Suppose that*

$$N \geq 40 \max(\tilde{K}_{t_0}(q), 1), \qquad \textit{and}$$

$$\lambda_s \geq \max\left(\frac{p\tilde{A}_s}{4} + \frac{p(p+3)}{4} t_0 \tilde{B}_s, \frac{17}{4} a(p, q) \omega \tilde{\gamma}_s \right) \qquad \textit{for all } s \in [0, t_0]. \tag{2.26}$$

Then, for $t \in [0, t_0]$, one has

$$\tilde{\varepsilon}_t^{N,p} \;\leq\; \left(2 + 8K_t(2)\tilde{M}_t^2\right) N^{-1} \left(1 + 16\tilde{K}_t(q)\tilde{M}_t^2\, N^{-1}\right).$$

Remark 2.11. (i) If there is only one component, the assertion of Theorem 2.10 reduces to that of Theorem 2.6.
(ii) Error bounds for the estimators $\langle f, \nu_t^N \rangle$ and $\langle f, \eta_t^N \rangle$ for a fixed function f hold analogously to Corollary 2.8.

2.6. **Open problems.** 1) The cases discussed in Sections 2.3 and 2.5 are extreme cases. In many typical applications, one would expect the state space to split up as time evolves into more and more components that get almost disconnected by the dynamics (local modes, metastable states). The study of such more complicated situations is an important topic for future research.

2) We have discussed here a setup with discrete state space and continuous time. In continuous time, particle systems on more general state spaces can in principle be treated by similar techniques, although of course additional technical considerations are required (cf. for instance [36]). For algorithmic applications, the case of discrete time and a continuous state space is probably the most interesting one. For an overview of the substantial literature and some more recent results in this case we refer to [2, 5, 8, 9, 14, 10, 18, 27] and references therein. An L^p approach similar to the one presented here is developed for the discrete time case in the PhD thesis of N. Schweizer [38].

3. Variances of weighted empirical averages

In this section we will prove Proposition 2.1, which shows that $\langle f, \nu_t^N \rangle$ is an unbiased estimator for $\langle f, \mu_t \rangle$ and gives an explicit formula for the variance. The proof follows the arguments developed in [14] relying on the identification of appropriate martingales.

Recall that the carré du champ (square field) operator Γ_t^N associated to $\mathcal{L}_t^N$ is defined for functions $\varphi : S^N \to \mathbb{R}$ by

$$\Gamma_t^N(\varphi) \;=\; \mathcal{L}_t^N \varphi^2 - 2\varphi \mathcal{L}_t^N \varphi,$$

i.e.,

$$\Gamma_t^N(\varphi)(x) \;=\; \sum_{y \in S} \mathcal{L}_t^N(x,y)\,(\varphi(y) - \varphi(x))^2 \qquad \forall x \in S^N. \tag{3.1}$$

It is well-known that the processes

$$M_t^\varphi \;=\; \varphi(t, X_t^N) - \varphi(0, X_0^N) - \int_0^t \left(\frac{\partial}{\partial s} + \mathcal{L}_s^N\right)\varphi(s, X_s^N)\,ds, \qquad \text{and} \tag{3.2}$$

$$N_t^\varphi \;=\; (M_t^\varphi)^2 - \int_0^t \Gamma_s^N(\varphi(s,\,\cdot\,))(X_s^N)\,ds \tag{3.3}$$

are martingales w.r.t. the filtration induced by the process X_t^N for any function $\varphi : \mathbb{R}^+ \times S^N \to \mathbb{R}$ that is twice continuously differentiable in the first variable, cf. e.g. [28, Appendix 1, Lemma 5.1]. For $x \in S^N$ let

$$\eta(x) = \frac{1}{N} \sum_{i=1}^N \delta_{X_i}$$

denote the corresponding empirical average. In the next lemma we derive expressions for $\mathcal{L}_t^N$ and Γ_t^N acting on linear functions on S^N of the form

$$\varphi_f(x) \;=\; \langle f, \eta(x) \rangle \;=\; N^{-1} \sum_{i=1}^N f(x_i).$$

Lemma 3.1. *For any function $f : S \to \mathbb{R}$ and $t \geq 0$, one has*

$$\mathcal{L}_t^N \langle f, \eta\rangle \;=\; \lambda_t \langle \mathcal{L}_t f, \eta\rangle + \langle H_t, \eta\rangle\langle f, \eta\rangle - \langle H_t f, \eta\rangle$$

and

$$\Gamma_t^N(\langle f, \eta\rangle) \;=\; \frac{\lambda_t}{N}\langle \Gamma_t(f), \eta\rangle + \frac{1}{N}\iint \big(H_t(y) - H_t(z)\big)^+ \big(f(z) - f(y)\big)^2\, \eta(dy)\, \eta(dz),$$

where Γ_t denotes the carré du champ operator w.r.t. $\mathcal{L}_t$.

Proof. The definition of $\mathcal{L}_t^N$ immediately yields

$$\mathcal{L}_t^N \langle f, \eta\rangle(x) \;=\; \frac{\lambda_t}{N}\sum_{i=1}^N \mathcal{L}_t f(x_i) + \frac{1}{N^2}\sum_{i,j=1}^N (H_t(x_i) - H_t(x_j))^+ (f(x_j) - f(x_i)). \tag{3.4}$$

Moreover,

$$\begin{aligned}
&\sum_{i,j=1}^N (H_t(x_i) - H_t(x_j))^+ (f(x_j) - f(x_i)) \\
&\qquad = \sum_{i,j: H_t(x_i) > H_t(x_j)} (H_t(x_i) - H_t(x_j))(f(x_j) - f(x_i)) \\
&\qquad = \sum_{i,j: H_t(x_j) > H_t(x_i)} (H_t(x_j) - H_t(x_i))(f(x_i) - f(x_j)) \\
&\qquad = \sum_{i,j: H_t(x_j) > H_t(x_i)} (H_t(x_i) - H_t(x_j))(f(x_j) - f(x_i)) \\
&\qquad = -\sum_{i,j=1}^N (H_t(x_i) - H_t(x_j))^- (f(x_j) - f(x_i)),
\end{aligned}$$

and hence

$$\sum_{i,j=1}^N (H_t(x_i) - H_t(x_j))(f(x_j) - f(x_i)) = 2\sum_{i,j=1}^N (H_t(x_i) - H_t(x_j))^+ (f(x_j) - f(x_i)).$$

Therefore the second term on the right hand side of (3.4) is equal to

$$\begin{aligned}
&\frac{1}{2N^2}\sum_{i,j=1}^N (H_t(x_i) - H_t(x_j))(f(x_j) - f(x_i)) \\
&\qquad = \Big(\frac{1}{N}\sum_{i=1}^N H_t(x_i)\Big)\Big(\frac{1}{N}\sum_{j=1}^N f(x_j)\Big) - \frac{1}{N}\sum_{i=1}^N H_t(x_i) f(x_i) \\
&\qquad = \langle H_t, \eta(x)\rangle\langle f, \eta(x)\rangle - \langle H_t f, \eta(x)\rangle,
\end{aligned}$$

from which the first claim follows.

Furthermore, since

$$\langle f, \eta(x^{i\to j})\rangle - \langle f, \eta(x)\rangle \;=\; N^{-1}\,(f(x_j) - f(x_i))\,,$$

(3.1) and (1.7) imply

$$\begin{aligned}\Gamma_t^N\langle f,\eta\rangle(x) &= \frac{\lambda_t}{N^2}\sum_{i=1}^N\sum_{y\in S}\mathcal{L}_t(x_i,y)(f(y)-f(x_i))^2\\ &\quad+\frac{1}{N^3}\sum_{i,j=1}^N\big(H_t(x_i)-H_t(x_j)\big)^+\big(f(x_j)-f(x_i)\big)^2,\end{aligned}$$

from which the second claim follows noting that the first term on the right hand side of the previous expression is equal to

$$\frac{\lambda_t}{N^2}\sum_{i=1}^N\Gamma_t(f)(x_i) \;=\; \frac{\lambda_t}{N}\langle\Gamma_t(f),\eta(x)\rangle. \qquad \square$$

Now let us define

$$\bar{A}^f_{s,t} \;=\; \langle q_{s,t}f,\eta^N_s\rangle \;=\; \frac{1}{N}\sum_{i=1}^N(q_{st}f)(X^N_{s,i})\,.$$

As a consequence of Lemma 3.1 we obtain:

Proposition 3.2. *The processes $\bar{M}^f_u$ and $\bar{N}^f_u$, $u\in[0,t]$, defined by*

$$\begin{aligned}\bar{M}^f_u &= \bar{A}^f_{u,t}-\bar{A}^f_{0,t}-\int_0^u\langle H_s,\eta^N_s\rangle\langle q_{s,t}f,\eta^N_s\rangle\,ds,\\ \bar{N}^f_u &= (\bar{M}^f_u)^2-\frac{1}{N}\int_0^u\lambda_s\,\langle\Gamma_s(q_{s,t}f),\eta^N_s\rangle\,ds\\ &\quad-\frac{1}{N}\int_0^u\iint\big(H_s(y)-H_s(z)\big)^+\big(q_{s,t}f(z)-q_{s,t}f(y)\big)^2\,\eta^N_s(dy)\,\eta^N_s(dz)\,ds\end{aligned}$$

are martingales w.r.t. the filtration $\mathcal{F}_t=\sigma(X^N_s\mid s\in[0,t])$.

Proof. Note that $\bar{A}^f_s=\varphi(s,X^N_s)$, where

$$\varphi(s,x) \;=\; N^{-1}\sum_{i=1}^N q_{st}f(x_i).$$

By the backward equation (2.1),

$$\begin{aligned}\frac{\partial}{\partial s}\varphi(s,x) &= -\frac{\lambda_s}{N}\sum_{i=1}^N\mathcal{L}_sq_{st}f(x_i)+\frac{1}{N}\sum_{i=1}^N H_sq_{st}f(x_i)\\ &= -\lambda_s\langle\mathcal{L}_sq_{st}f,\eta(x)\rangle+\langle H_sq_{st}f,\eta(x)\rangle,\end{aligned}$$

and by lemma 3.1,

$$\left(\mathcal{L}^N_s\varphi\right)(s,x) \;=\; \lambda_s\langle\mathcal{L}_sq_{s,t}f,\eta(x)\rangle+\langle H_s,\eta(x)\rangle\langle q_{s,t}f,\eta(x)\rangle-\langle H_sq_{s,t}f,\eta(x)\rangle$$

Hence

$$\left(\frac{\partial}{\partial s}+\mathcal{L}^N_s\right)\varphi(s,x) \;=\; \langle H_s,\eta(x)\rangle\langle q_{s,t}f,\eta(x)\rangle,$$

which proves that $\bar{M}^f = M^\varphi$ is a martingale, cf. (3.2). Similarly, by Lemma 3.1,

$$\begin{aligned}\Gamma_s^N(\varphi)(s,x) &= \frac{\lambda_s}{N}\langle \Gamma_s(q_{s,t}f), \eta(x)\rangle \\ &\quad + \frac{1}{N}\iint \big(H_s(y)-H_s(z)\big)^+ \big(q_{s,t}f(z)-q_{s,t}f(y)\big)^2\, \eta_s^N(dy)\, \eta_s^N(dz),\end{aligned}$$

which proves that $\bar{N}^f = N^\varphi$ is a martingale, cf. (3.3). □

Since in general, $\bar{A}^f_{s,t}$ is not a martingale, $\langle f, \eta_t^N\rangle$ is not an unbiased estimator for $\langle f, \mu_t\rangle$. This motivates considering $\langle f, \nu_t^N\rangle$ instead. Let

$$A^f_{s,t} \;=\; \langle q_{s,t}f, \nu_s^N\rangle \;=\; e^{-\int_0^s \langle H_r, \eta_r^N\rangle\, dr}\, \bar{A}^f_{s,t}. \tag{3.5}$$

Proposition 3.3. *The process $A^f_{u,t}$, $u \in [0,t]$, is a martingale with increasing process given by*

$$\begin{aligned}\langle A^f_{\bullet,t}\rangle_u &= \frac{1}{N}\int_0^u \lambda_s\, \langle 1, \nu_s^N\rangle \langle \Gamma_s(q_{s,t}f), \nu_s^N\rangle\, ds \\ &\quad + \frac{1}{N}\int_0^u \iint \big(H_s(x)-H_s(y)\big)^+ \big(q_{s,t}f(y)-q_{s,t}f(x)\big)^2\, \nu_s^N(dx)\, \nu_s^N(dy)\, ds.\end{aligned}$$

Proof. By the integration by parts formula for Stieltjes integrals and Proposition 3.2, we get

$$\begin{aligned}A^f_{u,t} - A^f_{0,t} &= \int_0^u e^{-\int_0^s \langle H_r, \eta_r^N\rangle\, dr} d\bar{A}^f_{s,t} - \int_0^u \langle H_s, \eta_s^N\rangle e^{-\int_0^s \langle H_r, \eta_r^N\rangle\, dr} \bar{A}^f_{s,t}\, ds \\ &= \int_0^u e^{-\int_0^s \langle H_r, \eta_r^N\rangle\, dr} d\bar{M}^f_s + \langle H_s, \eta_s^N\rangle A^f_s\, ds - \langle H_s, \eta_s^N\rangle A^f_s\, ds.\end{aligned}$$

Hence $[0,t] \ni s \mapsto A^f_{s,t}$ is a martingale whose increasing process can be written as

$$\langle A^f_{\bullet,t}\rangle_u \;=\; \int_0^u e^{-2\int_0^s \langle H_r, \eta_r^N\rangle\, dr}\, d\langle \bar{M}^f\rangle_s.$$

The result now follows by Proposition 3.2 and Equation (1.9). □

The purpose of the next lemma is to obtain an alternative representation (modulo martingale terms) of the term involving the carré du champ operator in the expression for $\langle A^f_{\bullet,t}\rangle$.

Lemma 3.4. *The following decomposition holds:*

$$\begin{aligned}&\int_0^u \lambda_s \langle 1, \nu_s^N\rangle \langle \Gamma_s(q_{st}f), \nu_s^N\rangle\, ds \\ &= \tilde{M}_u + \langle 1, \nu_u^N\rangle \left\langle (q_{ut}f)^2, \nu_u^N\right\rangle + \int_0^u \langle H_s, \nu_s^N\rangle \langle (q_{st}f)^2, \nu_s^N\rangle\, ds \\ &\quad - \int_0^u \langle 1, \nu_s^N\rangle \langle H_s(q_{st}f)^2, \nu_s^N\rangle\, ds,\end{aligned}$$

where $\tilde{M}$ is a martingale.

Proof. Let

$$Y_u := \langle 1, \nu_u^N\rangle \langle (q_{ut}f)^2, \nu_u^N\rangle = e^{-2\int_0^u \langle H_r, \eta_r^N\rangle\, dr} \left\langle (q_{ut}f)^2, \eta_u^N\right\rangle.$$

By applying the martingale problem to the functions $\varphi(s,x) = \big\langle (q_{st}f)^2, \eta(x)\big\rangle$, we obtain

$$Y_u = e^{-2\int_0^u \langle H_r,\eta_r^N\rangle\,dr}\,\big\langle (q_{ut}f)^2, \eta_u^N\big\rangle \;\sim\; -2\int_0^u e^{-2\int_0^s \langle H_r,\eta_r^N\rangle\,dr}\langle H_s,\eta_s^N\rangle\,\big\langle (q_{st}f)^2,\eta_s^N\big\rangle\,ds + \int_0^u e^{-2\int_0^s \langle H_r,\eta_r^N\rangle\,dr}\left(\frac{\partial}{\partial s}+\mathcal{L}_s^N\right)\varphi(s,X_s^N)\,ds.$$

Here and in the following we write $Y_u \sim Z_u$ if the processes Y_u and Z_u differ only by a martingale term. Proceeding as in the proof of proposition 3.2, we get that

$$\frac{\partial}{\partial s}\varphi(s,X_s^N) \;=\; 2\langle q_{st}f\frac{\partial}{\partial s}q_{st}f,\eta_s^N\rangle \;=\; -2\lambda_s\langle q_{st}f\mathcal{L}_s q_{st}f,\eta_s^N\rangle + 2\langle H_s(q_{st}f)^2,\eta_s^N\rangle,$$

and

$$\mathcal{L}_s^N\varphi(s,X_s^N) \;=\; \lambda_s\langle \mathcal{L}_s(q_{st}f)^2,\eta_s^N\rangle + \langle H_s,\eta_s^N\rangle\langle (q_{st}f)^2,\eta_s^N\rangle - \langle H_s(q_{st}f)^2,\eta_s^N\rangle.$$

Recalling that $\mathcal{L}_s(q_{st}f)^2 - 2q_{st}f\mathcal{L}_s q_{st}f = \Gamma_s(q_{s,t}f)$ and $\nu_s^N = \exp(-\int_0^s\langle H_r,\nu_r^N\rangle dr)\,\eta_s^N$, we conclude

$$\langle 1,\nu_u^N\rangle\langle (q_{ut}f)^2,\nu_u^N\rangle \;\sim\; -\int_0^u \langle H_s,\nu_s^N\rangle\langle (q_{st}f)^2,\nu_s^N\rangle\,ds + \int_0^u \langle 1,\nu_s^N\rangle\langle H_s(q_{st}f)^2,\nu_s^N\rangle\,ds + \int_0^u \lambda_s\,\langle 1,\nu_s^N\rangle\langle \Gamma_s(q_{st}f),\nu_s^N\rangle\,ds,$$

which proves the assertion. □

Lemma 3.5. *For all $t \geq 0$,*

$$\mathbb{E}\Big[\langle 1,\nu_t^N\rangle\langle f^2,\nu_t^N\rangle\Big] \;=\; \langle f^2,\mu_t\rangle - \mathbb{E}\left[\int_0^t \langle H_s,\nu_s^N\rangle\langle q_{st}f^2,\nu_s^N\rangle\,ds\right].$$

Proof. By the product rule for Stieltjes integrals,

$$\begin{aligned}\langle 1,\nu_s^N\rangle\langle q_{st}f^2,\nu_s^N\rangle &= e^{-\int_0^s\langle H_r,\eta_r^N\rangle\,dr}A_{s,t}^{f^2}\\ &= \int_0^s e^{-\int_0^u\langle H_r,\eta_r^N\rangle\,dr}dA_{u,t}^{f^2} - \int_0^s \langle H_u,\nu_u^N\rangle A_{u,t}^{f^2}\,du.\end{aligned}$$

Since $s\mapsto A_{s,t}^{f^2}$ is a martingale,

$$\mathbb{E}\left[\langle 1,\nu_t^N\rangle\langle f^2,\nu_t^N\rangle\right] \;=\; \langle q_{0,t}f^2,\mu_0\rangle - \mathbb{E}\left[\int_0^t \langle H_u,\nu_u^N\rangle A_{u,t}^{f^2}\,du\right].$$

The proof is completed by noting that $\langle q_{0,t}f^2,\mu_0\rangle = \langle f^2,\mu_t\rangle$. □

Proof of Proposition 2.1. Fix a function $f: S\to\mathbb{R}$ and $t\geq 0$. Recalling that, by (2.9), $\langle f,\mu_t\rangle = \langle q_{0,t}f,\mu_0\rangle$, we have

$$\begin{aligned}\langle f,\nu_t^N\rangle - \langle f,\mu_t\rangle &= \langle q_{t,t}f,\nu_t^N\rangle - \langle q_{0,t}f,\nu_0^N\rangle + \langle q_{0,t}f,\nu_0^N\rangle - \langle q_{0,t}f,\mu_0\rangle\\ &= A_{t,t}^f - A_{0,t}^f + \langle q_{0,t}f,\nu_0^N\rangle - \langle q_{0,t}f,\mu_0\rangle.\end{aligned}$$

Taking expectations on both sides, we immediately obtain

$$\mathbb{E}\left[\langle f,\nu_t^N\rangle\right] \;=\; \langle f,\mu_t\rangle,$$

because $s\mapsto A_{s,t}f$ is a martingale by Proposition 3.3, and ν_0^N is the empirical distribution of N i.i.d. random variables with distribution μ_0. Moreover, by Proposition 3.3 and

Lemma 3.4,

$$\begin{aligned}
N\,\mathbb{E}\Big[\big|\langle f,\nu_t^N\rangle-\langle f,\mu_t\rangle\big|^2\Big] \;&=\; N\,\mathbb{E}\Big[(A^f_{t,t}-A^f_{0,t})^2\Big] \;+\; N\,\mathbb{E}\Big[\big(\langle q_{0,t}f,\nu_0^N\rangle-\langle q_{0,t}f,\mu_0\rangle\big)^2\Big]\\
&=\; N\,\mathbb{E}\Big[\langle A^f_{\bullet,t}\rangle_t\Big]+\mathrm{Var}_{\mu_0}(q_{0,t}f)\\
&=\; \mathbb{E}\Big[\langle 1,\nu_t^N\rangle\langle f^2,\nu_t^N\rangle-\langle (q_{0,t}f)^2,\nu_0^N\rangle\Big] \;+\; \mathrm{Var}_{\mu_0}(q_{0,t}f)\\
&\quad+\mathbb{E}\int_0^t\langle H_s,\nu_s^N\rangle\langle (q_{st}f)^2,\nu_s^N\rangle\,ds \;-\; \mathbb{E}\int_0^t\langle 1,\nu_s^N\rangle\langle H_s(q_{st}f)^2,\nu_s^N\rangle\,ds\\
&\quad+\mathbb{E}\int_0^t\iint\big(H(x)-H(y)\big)^+\big(q_{s,t}f(y)-q_{s,t}f(x)\big)^2\,\nu_s^N(dx)\,\nu_s^N(dy)\,ds.
\end{aligned}$$

The assertion now follows from Lemma 3.5 observing that

$$\begin{aligned}
-\mathbb{E}\Big[\langle (q_{0,t}f)^2,\nu_0^N\rangle\Big] \;+\; \mathrm{Var}_{\mu_0}(q_{0,t}f) \;&=\; -\langle (q_{0,t}f)^2,\mu_0\rangle \;+\; \mathrm{Var}_{\mu_0}(q_{0,t}f)\\
&=\; -\langle q_{0,t}f,\mu_0\rangle^2=-\langle f,\mu_t\rangle^2.
\end{aligned}$$

□

4. Proof of Theorem 2.5

Proposition 4.1. *Let p, q, $r\in[1,\infty]$ be such that $p^{-1}=q^{-1}+r^{-1}$. Then, for $0\le s\le t$,*

$$\begin{aligned}
\mathbb{E}\left[V^N_{s,t}(f)\right] \le\, &V_{s,t}(f)\\
&+\big(6\|H_s\|_{L^q(\mu_s)}\|q_{s,t}f\|^2_{L^{2r}(\mu_s)}+\|H_s\|_{L^p(\mu_s)}\|q_{s,t}f^2\|_{L^p(\mu_s)}\big)\varepsilon_s^{N,p}.
\end{aligned}$$

Proof. Since $\langle f,\nu_s^N\rangle$ and $\langle g,\nu_s^N\rangle$ are unbiased estimators of $\langle f,\mu_s\rangle$ and $\langle g,\mu_s\rangle$, respectively, we have, by the Cauchy-Schwarz inequality,

$$\begin{aligned}
&\big|\mathbb{E}[\langle f,\nu_s^N\rangle\langle g,\nu_s^N\rangle]-\langle f,\mu_s\rangle\langle g,\mu_s\rangle\big|\\
&\quad=\big|\mathbb{E}\big[(\langle f,\nu_s^N\rangle-\langle f,\mu_s\rangle)(\langle g,\nu_s^N\rangle-\langle g,\mu_s\rangle)\big]\big|\\
&\quad\le\big(\mathbb{E}\big|\langle f,\nu_s^N\rangle-\langle f,\mu_s\rangle\big|^2\big)^{1/2}\big(\mathbb{E}\big|\langle g,\nu_s^N\rangle-\langle g,\mu_s\rangle\big|^2\big)^{1/2}\\
&\quad\le\varepsilon_s^{N,p}\,\|f\|_{L^p(\mu_s)}\,\|g\|_{L^p(\mu_s)}
\end{aligned}\tag{4.1}$$

for all $0\le s\le t$ and all functions f, $g:S\to\mathbb{R}$. Since the last term on the right-hand side of (2.4) can be bounded by

$$\iint|H_s(y)|(q_{s,t}f(z)-q_{s,t}f(y))^2\,\nu_s^N(dz)\,\nu_s^N(dy),$$

an application of (4.1) yields, by (1.3) and (2.10),

$$\begin{aligned}
\mathbb{E}\big[V^N_{s,t}(f)\big]&\le-\langle H_s(q_{s,t}f)^2,\mu_s\rangle\langle 1,\mu_s\rangle-\langle H_s,\mu_s\rangle\langle q_{s,t}f^2-(q_{s,t}f)^2,\mu_s\rangle\\
&\quad+\iint|H_s(y)|(q_{s,t}f(z)-q_{s,t}f(y))^2\,\mu_s(dz)\,\mu_s(dy)\,+\varepsilon_s^{N,p}R_{s,t}(f)\\
&=V_{s,t}(f)+\varepsilon_s^{N,p}R_{s,t}(f),
\end{aligned}$$

where

$$\begin{aligned}R_{s,t}(f) &= \|H_s(q_{s,t}f)^2\|_{L^p(\mu_s)} + \|H_s\|_{L^p(\mu_s)}\big\|q_{s,t}f^2 - (q_{s,t}f)^2\big\|_{L^p(\mu_s)}\\ &\quad + \|H_s\|_{L^p(\mu_s)}\big\|(q_{s,t}f)^2\big\|_{L^p(\mu_s)} + 2\|H_s q_{s,t}f\|_{L^p(\mu_s)}\|q_{s,t}f\|_{L^p(\mu_s)}\\ &\quad + \big\|H_s(q_{s,t}f)^2\big\|_{L^p(\mu_s)}\\ &\le \|H_s\|_{L^p(\mu_s)}\|q_{s,t}f^2\|_{L^p(\mu_s)} + 6\|H_s\|_{L^q(\mu_s)}\|q_{s,t}f\|^2_{L^{2r}(\mu_s)}.\end{aligned}$$

□

In order to bound $V^N_{s,t}(f)$ uniformly over $f \in L^p(\mu_t)$ with $\|f\|_{L^p(\mu_t)} \le 1$, one needs to be able to control $\|q_{s,t}f\|_{L^{2r}(\mu_t)}$ in terms of $\|f\|_{L^p(\mu_t)}$. This is possible if hypercontractivity holds and $t-s$ is sufficiently large. Over short time intervals $[s,t]$ we apply in a first step another rough estimate instead:

Lemma 4.2. *Let $p \ge 2$ and $N \in \mathbb{N}$. Then for $0 \le s \le t$,*

$$\frac{1}{N}\,\mathbb{E}\big[V^N_{s,t}(f)\big] \le 4\,\mathrm{osc}(H_s)\Big(1 + \varepsilon^{N,p}_s \exp\Big(2\int_s^t \mathrm{osc}(H_r)\,dr\Big)\Big)\|f\|^2_{L^p(\mu_t)}.$$

Proof. Setting

$$A^f_t \;:=\; \langle f, \nu^N_t\rangle \;=\; \langle f, \eta^N_t\rangle \exp\Big(-\int_0^t \langle H_s, \eta^N_s\rangle\,ds\Big),$$

we have $A^f_t = \langle f, \eta^N_t\rangle\, A^1_t$ for all $f : S \to \mathbb{R}$. Since

$$\langle f^2, \eta^N_t\rangle = \frac{1}{N}\sum_{i=1}^N f(X_{t,i})^2 \le \frac{1}{N}\Big(\sum_{i=1}^N |f(X_{t,i})|\Big)^2 = N\langle |f|, \eta^N_t\rangle^2,$$

we obtain, recalling that η^N_t is a probability measure,

$$\begin{aligned}V^N_{s,t}(f) \;&\le\; N\,(A^1_s)^2\Big((\max H^-_s + \max H^+_s)\langle |q_{s,t}f|, \eta^N_s\rangle^2 + \max H^-_s\,\langle (q_{s,t}f^2)^{1/2}, \eta^N_s\rangle^2\\ &\qquad + 2\,\mathrm{osc}(H_s)\langle |q_{s,t}f|, \eta^N_s\rangle^2\Big)\\ &\le\; N\,\mathrm{osc}(H_s)\Big(3\langle q_{s,t}|f|, \nu^N_s\rangle^2 + \langle (q_{s,t}f^2)^{1/2}, \nu^N_s\rangle^2\Big).\end{aligned}\tag{4.2}$$

Moreover, by inequality (4.1),

$$\mathbb{E}\left[\langle f, \nu^N_t\rangle^2\right] \le \langle f, \mu_t\rangle^2 + \varepsilon^{N,p}_t\|f\|^2_{L^p(\mu_t)},$$

hence, taking expectations on both sides of (4.2), we obtain

$$\begin{aligned}\frac{1}{N}\,\mathbb{E}\big[V^N_{s,t}(f)\big] &\le 3\,\mathrm{osc}(H_s)\big[\langle q_{s,t}|f|, \mu_s\rangle^2 + \varepsilon^{N,p}_s\big\|q_{s,t}|f|\big\|^2_{L^p(\mu_s)}\big]\\ &\quad + \mathrm{osc}(H_s)\big[\langle q_{s,t}f^2, \mu_s\rangle + \varepsilon^{N,p}_s\big\|q_{s,t}f^2\big\|^2_{L^{p/2}(\mu_t)}\big]\\ &\le 4\,\mathrm{osc}(H_s)\,\left[\langle f^2, \mu_t\rangle + \varepsilon^{N,p}_s \exp\Big(2\int_s^t \mathrm{osc}(H_r)\,dr\Big)\,\|f\|^2_{L^p(\mu_t)}\right],\end{aligned}$$

where we have used the fact that $\langle q_{s,t}f, \mu_s\rangle = \langle f, \mu_t\rangle$, and the estimate

$$\big\|q_{s,t}f\big\|_{L^p(\mu_t)} \;\le\; \exp\Big(\int_s^t \mathrm{osc}(H_r)\,dr\Big)\,\|f\|_{L^p(\mu_s)}.\tag{4.3}$$

The proof of (4.3) is elementary and can be found in [21]. □

Combining Proposition 4.1 and Lemma 4.2 we obtain the following (rough) a priori estimate:

Lemma 4.3. *Let p, q, $r \in [2,\infty]$ be such that $p^{-1} = q^{-1} + r^{-1}$, and choose δ as in Theorem 2.5. If*

$$N \geq 25 \max\left(1, \bar{C}_t(p,q,\delta)\right)$$

then

$$\varepsilon_t^{N,p} < 1.$$

Proof. Note that, by (2.10),

$$V_{s,t}(f) \leq 5\, \|H_s\|_{L^q(\mu_s)}\, \|q_{s,t}f\|_{L^{2r}(\mu_s)}$$

for any $f : S \to \mathbb{R}$ and $0 \leq s \leq t$. Hence Proposition 4.1 implies

$$\mathbb{E}\big[V_{s,t}^N(f)\big] \leq \|H_s\|_{L^q(\mu_s)} C_{s,t}(p,q)^2 \|f\|^2_{L^p(\mu_t)} \big(5 + 7\varepsilon_s^{N,p}\big).$$

Choosing N as stated we get

$$\frac{1}{N}\int_0^{(t-\delta)^+} \mathbb{E}\big[V_{s,t}^N(f)\big]\, ds \leq \frac{12}{25}\|f\|^2_{L^p(\mu_t)} \max\left(\varepsilon_t^{N,p}, 1\right).$$

On the other hand, by Lemma 4.2 and since $17\,\delta\, \mathrm{osc}(H_s) \leq 1$ for any $s \leq t$, we obtain

$$\begin{aligned}\frac{1}{N}\int_{(t-\delta)^+}^{t} \mathbb{E}\big[V_{s,t}^N(f)\big]\, ds &\leq \frac{4}{17}\big(1 + \varepsilon_t^{N,p} e^{2/17}\big)\|f\|^2_{L^p(\mu_t)} \\ &< \frac{1}{2}\|f\|^2_{L^p(\mu_t)} \max\left(\varepsilon_t^{N,p}, 1\right).\end{aligned}$$

Hence by Proposition 2.1, since $N \geq 50$, we get

$$\begin{aligned}\varepsilon_t^{N,p} &= \sup\left\{\frac{1}{N}\mathrm{Var}_{\mu_t}(f) + \frac{1}{N}\int_0^t \mathbb{E}\big[V_{s,t}^N(f)\big]\, ds \,\middle|\, f : S \to \mathbb{R} \text{ with } \|f\|_{L^p(\mu_r)} \leq 1,\ r \in [0,t]\right\} \\ &< \left(\frac{1}{50} + \frac{12}{25} + \frac{1}{2}\right) \max\left(\varepsilon_t^{N,p}, 1\right).\end{aligned}$$

□

The a priori estimate just obtained can be used instead of Lemma 4.2 to estimate $\mathbb{E}\left[V_{s,t}^N(f)\right]$ when $t - s$ is small:

Lemma 4.4. *Let $q \in]6,\infty]$ and $p \in]4q/(q-2), \infty[$. Suppose that*

$$N \geq 25 \max\left(1, \bar{C}_t(\tilde{p},q,\delta)\right),$$

where $\tilde{p}$ is defined by $\tilde{p}^{-1} = q^{-1} + (p/2)^{-1}$. Then for $0 \leq s \leq t \leq t_0$,

$$\mathbb{E}\left[V_{s,t}^N(f)\right] \;\leq\; V_{s,t}(f) + 7\, \exp\left(2\int_s^t \mathrm{osc}(H_r)\, dr\right) \|H_s\|_{L^q(\mu_s)} \|f\|^2_{L^p(\mu_t)}.$$

Proof. Note that $\tilde{p}^{-1} = q^{-1} + (p/2)^{-1} < 1/2$ by the assumptions on p and q. Applying Proposition 4.1 with p, q, r replaced by $\tilde{p}$, $\tilde{q} := q$, and $\tilde{r} := p/2$, respectively, yields

$$\mathbb{E}\big[V_{s,t}^N(f)\big] \leq V_{s,t}(f) + \left(\|H_s\|_{L^{\tilde{p}}(\mu_s)}\|q_{s,t}f^2\|_{L^{\tilde{p}}(\mu_s)} + 6\|H_s\|_{L^q(\mu_t)}\|q_{s,t}f\|^2_{L^p(\mu_s)}\right) \varepsilon_s^{N,\tilde{p}}$$

Since $\tilde{p} < \min(q, p/2)$, the claim follows by Lemma 4.3 and the estimate (4.3). □

We are now ready to prove the theorem:

Proof of Theorem 2.5. By Proposition 4.1 we have
$$\mathbb{E}\big[V_{s,t}^N(f)\big] \le V_{s,t}(f) + 7\|H_s\|_{L^q(\mu_s)} C_{s,t}(p,q)^2 \|f\|^2_{L^p(\mu_t)} \varepsilon_t^{N,p}$$
for any $f : S \to \mathbb{R}$ and $0 \le s \le t$. Therefore by Proposition 2.1, Lemma 4.4, and the choice of δ,
$$\begin{aligned} N\,\mathbb{E}\big|\langle f, \nu_t^N\rangle - \langle f, \mu_t\rangle\big|^2 &= \mathrm{Var}_{\mu_t}(f) + \int_0^{(t-\delta)^+} \mathbb{E}\big[V_{s,t}^N(f)\big]\,ds + \int_{(t-\delta)^+}^t \mathbb{E}\big[V_{s,t}^N(f)\big]\,ds \\ &\le \mathrm{Var}_{\mu_t}(f) + \int_0^t V_{s,t}(f)\,ds \\ &\quad + \Big[7\bar{C}_t(p,q,\delta)\varepsilon_t^{N,p} + 7e^{2/17}\int_{(t-\delta)^+}^t \|H_s\|_{L^q(\mu_s)}\,ds\Big]\,\|f\|^2_{L^p(\mu_t)}. \end{aligned}$$
Observing that $\|H_s\|_{L^q(\mu_s)} \le \mathrm{osc}(H_s)$ and that $7\,e^{2/17}/17 < 1$, we obtain (2.12).

Furthermore, by maximizing (2.12) over all $f : S \to \mathbb{R}$ such that $\|f\|_{L^p(\mu_t)} \le 1$ and over t, we get
$$N\varepsilon_t^{N,p} \le 2 + v_t^p + 7\bar{C}_t(p,q,\delta)\varepsilon_t^{N,p}$$
for all $t \in [0, t_0]$. Recalling that $N > 25\bar{C}_t(p,q,\delta)$ by assumption, we obtain
$$\begin{aligned} \varepsilon_t^{N,p} &\le \frac{2+v_t^p}{N - 7\bar{C}_t(p,q,\delta)} = (2+v_t^p)\left(\frac{1}{N} + \frac{7\bar{C}_t(p,q,\delta)}{N(N-7\bar{C}_t(p,q,\delta))}\right) \\ &\le (2+v_t^p)\,N^{-1}\Big(1 + \frac{7\cdot 25}{18}\bar{C}_t(p,q,\delta)N^{-1}\Big), \end{aligned}$$
which implies (2.13). □

5. Proofs of Theorems 2.6 and 2.10

Proof of Theorem 2.6. By the estimates in [21] we have, for $0 \le s \le t \le t_0$,
$$\|q_{s,t} f\|_{L^p(\mu_s)} \le 2^{1/4}\|f\|_{L^p(\mu_s)}$$
for all $f : S \to \mathbb{R}$, provided
$$\lambda_s \;\ge\; \frac{p}{4}\,A_s \;+\; \frac{p(p+3)}{4}\,t_0\,B_s \qquad \text{for all } s \in [0, t_0]. \tag{5.1}$$
Hence, under this condition, we get $C_{s,t}(p) \le 2^{1/4}$. Moreover, by [21],
$$\|q_{t-\delta,t} f\|_{L^q(\mu_{t-\delta})} \;\le\; \exp\Big(\int_{t-\delta}^t \max H_r^-\,dr\Big)\,\|f\|_{L^p(\mu_t)}$$
for all $f : S \to \mathbb{R}$ and $0 \le \delta \le t \le t_0$, provided
$$\lambda_s \ge \frac{\gamma_s}{4\delta}\log\frac{q-1}{p-1} \qquad \text{for all } s \in [0, t_0]. \tag{5.2}$$
Choosing $\delta = (17\omega)^{-1}$, we obtain that, for $s \le t - \delta$,
$$\|q_{s,t} f\|_{L^p(\mu_s)} \;=\; \|q_{s,t-\delta} q_{t-\delta,t} f\|_{L^p(\mu_s)} \;\le\; 2^{1/4}e^{1/17}\|f\|_{L^q(\mu_t)},$$
if both (5.1) and (5.2) hold. Hence
$$C_{s,t}(p,q) \le 2^{1/4}e^{1/17}$$
provided (5.1) holds and
$$\lambda_s \ge \frac{\gamma_s}{4\delta}\log\max\Big(\frac{2r-1}{p-1}, \frac{2p-2}{p-2}\Big) \qquad \text{for all } s \in [0, t_0].$$

Since $2 < \tilde{p} < p$ and $\tilde{p}^{-1} = q^{-1} + (p/2)^{-1}$, we obtain similarly that $C_{s,t}(\tilde{p}, q) \leq 2^{1/4} e^{1/17}$ provided (5.1) holds and

$$\lambda_s \geq \frac{\gamma_s}{4\delta} \log \max\big(\frac{p-1}{\tilde{p}-1}, \frac{2\tilde{p}-2}{\tilde{p}-2}\big) \qquad \text{for all } s \in [0, t_0].$$

Hence by (2.14) and (2.15) we obtain

$$v_t(p) \leq 5 \cdot 2^{1/2}\, K_t(2), \qquad \bar{C}_t(p,q,\delta) \leq 2^{1/2}\, e^{2/17}\, K_t(q), \qquad \bar{C}_t(\tilde{p},q,\delta) \leq 2^{1/2}\, e^{2/17}\, K_t(q)$$

for any $t \leq t_0$. The assertion now follows from Theorem 2.5. □

Proof of Lemma 2.2. For a function $f : S \to \mathbb{R}$ and $t \geq 0$ let $f_t := f - \langle f, \mu_t \rangle$. Then

$$\langle f_t, \eta_t^N \rangle = \langle f, \eta_t^N \rangle - \langle f, \mu_t \rangle$$

and, by (1.9),

$$\langle f_t, \nu_t^N \rangle = \langle 1, \nu_t^N \rangle \langle f_t, \eta_t^N \rangle. \tag{5.3}$$

Hence

$$\begin{aligned}
\mathbb{E}\big[\langle f_t, \eta_t^N \rangle^2\big] &\leq 2\,\mathbb{E}\Big[\big(\langle f_t, \eta_t^N \rangle - \langle f_t, \nu_t^N \rangle\big)^2\Big] + 2\,\mathbb{E}\big[\langle f_t, \nu_t^N \rangle^2\big] \\
&= 2\,\mathbb{E}\Big[\big(\langle 1, \nu_t^N \rangle - 1\big)^2 \langle f_t, \eta_t^N \rangle^2\Big] + 2\,\mathbb{E}\big[\langle f_t, \nu_t^N \rangle^2\big] \\
&\leq 2\,\|f_t\|_{\sup}^2\, \mathbb{E}\left[\big(\langle 1, \nu_t^N \rangle - 1\big)^2\right] + 2\,\mathbb{E}\left[\langle f_t, \nu_t^N \rangle^2\right].
\end{aligned}$$

Applying this bound and (5.3), we obtain the L^1 estimate:

$$\begin{aligned}
\mathbb{E}\big[|\langle f_t, \eta_t^N \rangle|\big] &= \mathbb{E}\big[|\langle f_t, \eta_t^N \rangle \big(1 - \langle 1, \nu_t^N \rangle\big)|\big] + \mathbb{E}\big[|\langle f_t, \nu_t^N \rangle|\big] \\
&\leq \mathbb{E}\left[\langle f_t, \eta_t^N \rangle^2\right]^{1/2} \mathbb{E}\big[\big(\langle 1, \nu_t^N \rangle - 1\big)^2\big]^{1/2} + \mathbb{E}\left[\langle f_t, \nu_t^N \rangle^2\right]^{1/2} \\
&\leq \mathbb{E}\left[\langle f_t, \nu_t^N \rangle^2\right]^{1/2} + \sqrt{2}\,\|f_t\|_{\sup} \mathbb{E}\left[\big(\langle 1, \nu_t^N \rangle - 1\big)^2\right] \\
&\quad + \sqrt{2}\,\mathbb{E}\left[\langle f_t, \nu_t^N \rangle^2\right]^{1/2} \mathbb{E}\left[\big(\langle 1, \nu_t^N \rangle - 1\big)^2\right]^{1/2}.
\end{aligned}$$

This proves Lemma 2.2. □

Proof of Corollary 2.8. The first assertion is an immediate consequence of (2.12) and (2.18). The second assertion follows by the first one and (2.6). □

Proof of Theorem 2.10. Fix $i \in I$ and define

$$h_t(i) := \langle H_t, \mu_t^i \rangle = \int_{S_i} H_t \, d\mu_t / \mu_t(S_i).$$

Note that

$$h_t(i) = -\frac{d}{dt} \log \mu_t(S_i).$$

Since (1.2) and (1.3) hold, $H_t^i = H_t - h_t(i)$ is the negative logarithmic time derivative of μ_t^i. If we define $q_{s,t}^i f$ for functions $f : S_i \to \mathbb{R}$ in the same way as $q_{s,t} f$ with H_t replaced by H_t^i, then

$$q_{s,t} f(x) \;=\; \exp\big(-\int_s^t h_r(i)\, dr\big)\, q_{s,t}^i f(x) \;=\; \frac{\mu_t(S_i)}{\mu_s(S_i)}\, q_{s,t}^i f(x).$$

In particular, for $p \in [1, \infty]$, we have

$$\|q_{s,t} f\|_{\widetilde{L^p(\mu_s)}} = \max_{i \in I} \|q_{s,t} f\|_{L^p(\mu_s^i)} \le \max_{i \in I} \frac{\mu_t(S_i)}{\mu_s(S_i)} \|q_{s,t}^i f\|_{L^p(\mu_s^i)}. \tag{5.4}$$

Assuming Poincaré and log Sobolev inequalities with respect to the measures μ_t^i and the functions H_t^i, we obtain the same type of L^p-L^q bounds for the operators $q_{s,t}^i$ as we did for the operators $q_{s,t}$ in the proof of Theorem 2.6. Because of (5.4) the assertion then follows similarly as above. □

Appendix A. Spectral gap and LSI for 1D Metropolis

In this appendix we prove upper bounds for the Poincaré and logarithmic Sobolev constants for Random Walk Metropolis algorithms on a finite subset S of $\mathbb{Z}$. Let $S := \{a, a+1, \dots, -1, 0, 1, \dots, a+\Delta-1\}$ with $a \in \mathbb{Z}$ and $\Delta \in \mathbb{N}$ such that $0 \in S$. We assume that μ is a probability measure on S satisfying

(i) $\mu(x) \le \rho\mu(y)$ for any x, $y \in [-s, s]$;
(ii) $\mu(x+1) \le \alpha\mu(x)$ for any $x \ge s$, and $\mu(x-1) \le \alpha\mu(x)$ for any $x \le -s$,

for appropriate constants $s \in \mathbb{Z}_+$, $\rho \in [1, +\infty[$, and $\alpha \in]0, 1[$. For notational convenience, we set

$$b := a + \Delta - 1, \qquad r := \frac{1}{1-\alpha} \wedge \Delta, \qquad u := s \wedge \Delta.$$

The Random Walk Metropolis chain for sampling from μ is the Markov chain on S with generator $\mathcal{L}$ satisfying

$$\mathcal{L}(x, y) = \begin{cases} \frac{1}{2} \min\left(\frac{\mu(y)}{\mu(x)}, 1\right), & \text{if } |y-x| = 1, \\ 0, & \text{if } |y-x| > 1. \end{cases}$$

To estimate the Poincaré constant for this dynamics, we can apply a general upper bound for one-dimensional Markov chains due to Miclo [32], which implies in our case

$$C^{\mathrm{Poi}} \le 4 \max(B^+, B^-), \tag{A.1}$$

where

$$B^+ := \max_{1 \le k \le b} B_k^+, \qquad B_k^+ := \sum_{x=1}^{k} \frac{1}{\mu(x-1) \wedge \mu(x)} \sum_{x=k}^{b} \mu(x),$$

$$B^- := \max_{a \le k \le -1} B_k^-, \qquad B_k^- := \sum_{x=k}^{-1} \frac{1}{\mu(x+1) \wedge \mu(x)} \sum_{x=a}^{k} \mu(x).$$

The bound is sharp up to a factor 4, see [32]. We are going to estimate B_k^+ in the cases $k > s$ and $k \le s$ separately. Corresponding bounds hold for B_k^-. Let us assume first that $k > s$. Then we have, by (ii),

$$\sum_{x=s+1}^{k} \frac{1}{\mu(x-1) \wedge \mu(x)} = \sum_{x=s+1}^{k} \frac{1}{\mu(x)} \le \frac{1}{\mu(k)} \sum_{i=0}^{k-s-1} \alpha^i \le \frac{r}{\mu(k)}.$$

and, by (i) and (ii),

$$\sum_{x=1}^{s} \frac{1}{\mu(x-1) \wedge \mu(x)} \le \frac{\rho u}{\mu(s)} \le \frac{\alpha^{k-s} \rho u}{\mu(k)}.$$

Hence

$$\sum_{x=1}^{k} \frac{1}{\mu(x-1)\wedge\mu(x)} \le (r+\alpha^{k-s}\rho u)\frac{1}{\mu(k)}. \tag{A.2}$$

Similarly, by (ii),

$$\sum_{x=k}^{b} \mu(x) \le \mu(k)\sum_{i=0}^{b-k}\alpha^{i} \le r\mu(k). \tag{A.3}$$

Therefore (A.2) and (A.3) yield

$$B_k^+ \le r\big(r+\alpha^{k-s}\rho u\big) \le r^2+\rho\, u\, r \qquad \text{for any } k>s. \tag{A.4}$$

Let us now consider the case $k \le s$: by (i) and since $s\wedge b \le u$, we have

$$\sum_{x=1}^{k} \frac{1}{\mu(x-1)\wedge\mu(x)} \sum_{x=k}^{s\wedge b-1}\mu(x) = \sum_{x=1}^{k}\sum_{y=k}^{s\wedge b-1} \frac{\mu(y)}{\mu(x-1)\wedge\mu(x)} \le \rho k(u-k) \le \rho u^2/4.$$

Moreover, similarly to (A.3), we have

$$\sum_{x=s\wedge b}^{b} \mu(x) \le r\mu(s\wedge b),$$

hence, by (i) and since $k \le s$ and $k \le \Delta$,

$$\sum_{x=1}^{k} \frac{1}{\mu(x-1)\wedge\mu(x)} \sum_{x=s\wedge b}^{b}\mu(x) \le r\sum_{x=1}^{k} \frac{\mu(s\wedge b)}{\mu(x-1)\wedge\mu(x)} \le \rho\, k\, r \le \rho\, u\, r.$$

Combining these estimates, we obtain

$$B_k^+ \le \frac{1}{4}\rho u^2 + \rho u r, \qquad \text{for any } k \le s. \tag{A.5}$$

By (A.4) and (A.5), we finally obtain

$$B^+ := \max_{k=1,\dots,b} B_k^+ \le \rho u r + \max(r^2, \rho u^2/4).$$

Observing that the same estimate holds for B^-, we have shown:

Theorem A.1. *The Poincaré constant C^{Poi} for the Random Walk Metropolis chain with stationary distribution μ satisfies*

$$C^{\mathrm{Poi}} \le 4\rho u r + \max(4r^2, \rho u^2)$$

Proof. The result holds by the upper bound (A.1). □

For the corresponding logarithmic Sobolev constant the following upper bound follows from the results in [32]:

$$\gamma \le 20\max(\beta^+, \beta^-),$$

where

$$\beta^+ := \max_{1\le k\le b}\beta_k^+, \qquad \beta_k^+ := \sum_{x=1}^{k} \frac{2}{\mu(x-1)\wedge\mu(x)} \sum_{x=k}^{b}\mu(x)\Big|\log\sum_{x=k}^{b}\mu(x)\Big|,$$

$$\beta^- := \max_{a\le k\le -1}\beta_k^-, \qquad \beta_k^- := \sum_{x=k}^{-1} \frac{2}{\mu(x+1)\wedge\mu(x)} \sum_{x=a}^{k}\mu(x)\Big|\log\sum_{x=a}^{k}\mu(x)\Big|.$$

Again, the bound is sharp up to an explicit numerical constant. A rough estimate for β_k^+ can easily be obtained observing that

$$\Big|\log \sum_{x=k}^{b} \mu(x)\Big| = \log\Big(\sum_{x=k}^{b}\mu(x)\Big)^{-1} \leq \log\frac{1}{\mu(k)} \leq \log\frac{1}{\mu_*},$$

where $\mu_* = \min_x \mu(x)$. In fact, this implies

$$\beta_k^+ \leq 2B_k^+ \log\frac{1}{\mu_*},$$

hence upper bounds for β^+ and β^- can be obtained from the corresponding bounds for B^+ and B^- simply by multiplying by a factor $2\log\mu_*^{-1}$. In particular, the upper bound for C^{Poi} derived above yields an upper bound for γ:

Theorem A.2. *One has*

$$\gamma \leq 10\big(4\rho ur + \max(\rho u^2, 4r^2)\big)\log\frac{1}{\mu_*}.$$

Example: A discrete Gauss model. Assume that

$$\mu(x) \propto \exp\Big(-\frac{x^2}{2\sigma^2}\Big)$$

for some finist constant $\sigma > 0$. Then one can check that (i) and (ii) above are satisfied with

$$s = \lfloor\sigma\rfloor, \qquad \rho = e^{1/2}, \qquad \alpha = \frac{\mu(s+1)}{\mu(s)} = \exp\Big(-\frac{\lfloor\sigma\rfloor + 1/2}{\sigma^2}\Big).$$

Note that $\alpha \leq e^{-1/2}$ for $\sigma < 1$ and $\alpha \leq e^{-3/4\sigma}$ for $\sigma \geq 1$. Applying the elementary inequality $1 - e^{-x} \geq \min(2x/3, 1/2)$, we obtain $1-\alpha \geq 1/(2\sigma)$ if $\sigma > 1$ and $1 - \alpha \geq 1/3$ if $\sigma \leq 1$. Hence

$$r = \frac{1}{1-\alpha} \wedge \Delta \leq (2\sigma \vee 3) \wedge \Delta \leq 2\big((\sigma\wedge\Delta)\vee 2\big).$$

By Theorem A.1, we then obtain

$$C^{\text{Poi}} \leq 30\,\big((\sigma\wedge\Delta)\vee 2\big)^2.$$

Moreover, since $-\Delta \leq a \leq b \leq \Delta$, one has

$$\frac{\mu(k)}{\mu(0)} = \exp\Big(-\frac{k^2}{2\sigma^2}\Big) \geq \exp\Big(-\frac{1}{2}\frac{\Delta^2}{\sigma^2}\Big) \qquad \text{for any } k \in S,$$

and thus

$$\log\frac{1}{\mu_*} \leq \frac{1}{2}(\Delta/\sigma)^2 + \log\frac{1}{\mu(0)} \leq \frac{1}{2}(\Delta/\sigma)^2 + \log\Delta.$$

Therefore we obtain, by Theorem A.2,

$$\begin{aligned}\gamma &\leq 150\big((\sigma\wedge\Delta)\vee 2\big)^2(\Delta/\sigma)^2 + 300\big((\sigma\wedge\Delta)\vee 2\big)^2\log\Delta\\ &\leq 300\Big(\frac{\Delta}{\sigma\wedge 1}\Big)^2 + 300\big((\sigma\wedge\Delta)\vee 2\big)^2\log\Delta.\end{aligned}$$

References

1. Th. Bengtsson, P. Bickel, and Bo Li, *Curse-of-dimensionality revisited: collapse of the particle filter in very large scale systems*, Probability and statistics: essays in honor of David A. Freedman, Inst. Math. Stat. Collect., vol. 2, Inst. Math. Statist., Beachwood, OH, 2008, pp. 316–334. MR 2459957 (2009k:93144)
2. A. Beskos, D. Crisan, and A. Jasra, *On the stability of a class of sequential Monte Carlo methods in high dimensions*, Tech. report, Imperial College, London, 2011.
3. P. Bickel, Bo Li, and Th. Bengtsson, *Sharp failure rates for the bootstrap particle filter in high dimensions*, Pushing the limits of contemporary statistics: contributions in honor of Jayanta K. Ghosh, Inst. Math. Stat. Collect., vol. 3, Inst. Math. Statist., Beachwood, OH, 2008, pp. 318–329. MR 2459233 (2010c:93107)
4. O. Cappé, A. Guillin, J. M. Marin, and C. P. Robert, *Population Monte Carlo*, J. Comput. Graph. Statist. **13** (2004), no. 4, 907–929. MR 2109057
5. O. Cappé, E. Moulines, and T. Rydén, *Inference in hidden Markov models*, Springer Series in Statistics, Springer, New York, 2005. MR MR2159833 (2006e:60002)
6. F. Cérou, P. Del Moral, and A. Guyader, *A nonasymptotic theorem for unnormalized Feynman-Kac particle models*, Ann. Inst. H. Poincar Probab. Statist. **47** (2011), no. 3, 629–649.
7. N. Chopin, *A sequential particle filter method for static models*, Biometrika **89** (2002), no. 3, 539–551. MR 1929161
8. ———, *Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference*, Ann. Statist. **32** (2004), no. 6, 2385–2411. MR 2153989 (2006b:60033)
9. P. Del Moral, *Feynman-Kac formulae*, Springer-Verlag, New York, 2004. MR MR2044973 (2005f:60003)
10. P. Del Moral, A. Doucet, and A. Jasra, *Sequential Monte Carlo samplers*, J. R. Statist. Soc. B **68** (2006), no. 3, 411–436. MR MR1819122 (2002k:60013)
11. ———, *On adaptive resampling procedures for sequential Monte Carlo methods*, Bernoulli (to appear).
12. P. Del Moral and A. Guionnet, *On the stability of interacting processes with applications to filtering and genetic algorithms*, Ann. Inst. H. Poincaré Probab. Statist. **37** (2001), no. 2, 155–194. MR MR1819122 (2002k:60013)
13. P. Del Moral and L. Miclo, *On the convergence and applications of generalized simulated annealing*, SIAM J. Control Optim. **37** (1999), no. 4, 1222–1250 (electronic). MR 1691939 (2000d:90125)
14. ———, *Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering*, Séminaire de Probabilités, XXXIV, Lecture Notes in Math., vol. 1729, Springer, Berlin, 2000, pp. 1–145. MR MR1768060 (2001g:60091)
15. P. Diaconis, *The cutoff phenomenon in finite Markov chains*, Proc. Nat. Acad. Sci. U.S.A. **93** (1996), no. 4, 1659–1664. MR 1374011 (97b:60112)
16. J. Ding, E. Lubetzky, and Y. Peres, *Total variation cutoff in birth-and-death chains*, Probab. Theory Related Fields **146** (2010), no. 1-2, 61–85. MR 2550359 (2010m:60011)
17. R. Douc, A. Guillin, J.-M. Marin, and C. P. Robert, *Minimum variance importance sampling via population Monte Carlo*, ESAIM Probab. Stat. **11** (2007), 427–447 (electronic). MR 2339302 (2008m:62121)
18. R. Douc and E. Moulines, *Limit theorems for weighted samples with applications to sequential Monte Carlo methods*, Ann. Statist. **36** (2008), no. 5, 2344–2376. MR 2458190 (2009k:60053)
19. A. Doucet, N. de Freitas, and N. Gordon (eds.), *Sequential Monte Carlo methods in practice*, Springer-Verlag, New York, 2001. MR 1847783 (2003h:65007)
20. A. Eberle and C. Marinelli, *Stability of sequential Markov chain Monte Carlo methods*, Conference Oxford sur les méthodes de Monte Carlo séquentielles, ESAIM Proc., vol. 19, EDP Sci., Les Ulis, 2007, pp. 22–31. MR 2405646 (2009e:82063)
21. ———, *L^p estimates for Feynman-Kac propagators with time-dependent reference measures*, J. Math. Anal. Appl. **365** (2010), no. 1, 120–134. MR 2585083
22. C. J. Geyer, *Markov chain Monte Carlo maximum likelihood*, Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface, 1991, pp. 156–163.
23. Ĭ. Ī. Gīhman and A. V. Skorohod, *The theory of stochastic processes. II*, Springer-Verlag, New York, 1975. MR 0375463 (51 #11656)
24. A. Gulisashvili and J. A. van Casteren, *Non-autonomous Kato classes and Feynman-Kac propagators*, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2006. MR 2253111 (2008b:60161)

25. K. Hukushima and K. Nemoto, *Exchange Monte Carlo method and application to spin glass simulations*, J. Phys. Soc. Japan **65** (1996), no. 6, 1604–1608.
26. C. Jarzynski, *Nonequilibrium equality for free energy differences*, Phys. Rev. Lett. **78** (1997), no. 14, 2690–2693.
27. A. Jasra and A. Doucet, *Stability of sequential Monte Carlo samplers via the Foster-Lyapunov condition*, Statist. Probab. Lett. **78** (2008), no. 17, 3062–3069. MR 2474398 (2010b:60066)
28. C. Kipnis and C. Landim, *Scaling limits of interacting particle systems*, Springer-Verlag, Berlin, 1999. MR MR1707314 (2000i:60001)
29. S. C. Kou, Qing Zhou, and Wing Hung Wong, *Equi-energy sampler with applications in statistical inference and statistical mechanics*, Ann. Statist. **34** (2006), no. 4, 1581–1652. MR 2283711
30. J. S. Liu, *Monte Carlo strategies in scientific computing*, Springer-Verlag, New York, 2001. MR MR1842342 (2002i:65006)
31. N. Madras and Z. Zheng, *On the swapping algorithm*, Random Structures Algorithms **22** (2003), no. 1, 66–97. MR MR1943860 (2004c:82117)
32. L. Miclo, *An example of application of discrete Hardy's inequalities*, Markov Process. Related Fields **5** (1999), no. 3, 319–330. MR 1710983 (2000h:60081)
33. R. M. Neal, *Annealed importance sampling*, Stat. Comput. **11** (2001), no. 2, 125–139. MR 1837132
34. C. P. Robert and G. Casella, *Monte Carlo statistical methods*, second ed., Springer-Verlag, New York, 2004. MR MR2080278 (2005d:62006)
35. M. Rousset, *Continuous time population Monte Carlo and computational physics*, Ph.D. thesis, Université Paul Sabatier, Toulouse, 2006.
36. ———, *On the control of an interacting particle estimation of Schrödinger ground states*, SIAM J. Math. Anal. **38** (2006), no. 3, 824–844 (electronic). MR 2262944 (2007m:60300)
37. L. Saloff-Coste, *Lectures on finite Markov chains*, Lectures on probability theory and statistics (Saint-Flour, 1996), Lecture Notes in Math., vol. 1665, Springer, Berlin, 1997, pp. 301–413. MR MR1490046 (99b:60119)
38. N. Schweizer, *Non-asymptotic error bounds for Sequential MCMC*, Ph.D. thesis, Universität Bonn, 2011.
39. W. Stannat, *On the convergence of genetic algorithms—a variational approach*, Probab. Theory Related Fields **129** (2004), no. 1, 113–132. MR MR2052865 (2005d:35040)
40. N. Whiteley, *Sequential Monte Carlo samplers: error bounds and insensitivity to initial conditions*, arXiv:1103.3970v1.

(Andreas Eberle) Institut für Angewandte Mathematik, Universität Bonn, Endenicher Allee 60, D-53115 Bonn, Germany

E-mail address: `eberle@uni-bonn.de`

URL: `http://www.uni-bonn.de/~eberle`

(Carlo Marinelli) Institut für Angewandte Mathematik, Universität Bonn, Endenicher Allee 60, D-53115 Bonn, Germany, and Facoltà di Economia, Università di Bolzano, Piazza Università 1, I-39100 Bolzano, Italy

URL: `http://www.uni-bonn.de/~cm788`