

## STABILITY OF SEQUENTIAL MARKOV CHAIN MONTE CARLO METHODS \*

ANDREAS EBERLE<sup>1</sup> AND CARLO MARINELLI<sup>2</sup>

**Abstract.** Sequential Monte Carlo Samplers are a class of stochastic algorithms for Monte Carlo integral estimation w.r.t. probability distributions, which combine elements of Markov chain Monte Carlo methods and importance sampling/resampling schemes. We develop a stability analysis by functional inequalities for a nonlinear flow of probability measures describing the limit behavior of the methods as the number of particles tends to infinity. Stability results are derived both under global and local assumptions on the generator of the underlying Metropolis dynamics. This allows us to prove that the combined methods sometimes have good asymptotic stability properties in multimodal setups where traditional MCMC methods mix extremely slowly. For example, this holds for the mean field Ising model at all temperatures.

### 1. INTRODUCTION

Spectral gap estimates, or, equivalently, Poincaré inequalities, as well as other related functional inequalities provide powerful tools for the study of convergence to equilibrium of reversible time-homogeneous Markov processes (see e.g. [8], [9], [10]). In particular, they have been successfully used to analyze convergence properties of Markov Chain Monte Carlo (MCMC) methods based on reversible Markov chains (see e.g. [11]). The idea of MCMC methods is to produce approximate samples from a probability distribution  $\mu$  by simulating for sufficiently long time an ergodic Markov chain having  $\mu$  as invariant measure. MCMC methods have become the standard to carry out Monte Carlo integrations with respect to complex probability distributions in many fields of applications, including in particular Bayesian statistics, statistical physics, and chemistry. We just refer the interested reader to [15] and [19] and references therein as an example of work in this area, as the literature is by now enormous. Since the Markov chain is usually started with an initial distribution that is very different from  $\mu$ , strong convergence properties, such as exponential convergence to equilibrium with a sufficiently large rate, are required to ensure that the corresponding MCMC method produces sufficiently good approximate samples from  $\mu$ . However, these strong convergence properties often do not hold in multimodal, and in particular high-dimensional problems, as they arise in many applications. For example, in statistical mechanics models with phase transitions, the rate of convergence often decays exponentially in the system size within the multi-phase regime.

In [12] and [13] we initiate a study of convergence properties by functional inequalities for a different class of stochastic algorithms that are a combination of sequential Monte Carlo and MCMC methods. Instead of trying to produce constantly improved samples of a fixed distribution  $\mu$ , these *sequential MCMC methods* try

---

\* This work is supported by the Sonderforschungsbereich 611, Bonn. The second author also gratefully acknowledges the hospitality and financial support of the Institute of Mathematics, Polish Academy of Sciences, Warsaw and of the Institute des Hautes Études Scientifiques, Bures-sur-Yvette through an IPDE fellowship.

<sup>1</sup> Institut für Angewandte Mathematik, Universität Bonn, Wegelerstr. 6, 53115 Bonn, Germany

<sup>2</sup> Institut für Angewandte Mathematik, Universität Bonn, Wegelerstr. 6, 53115 Bonn, Germany

to keep track as precisely as possible of an evolving sequence  $(\mu_t)_{0 \leq t \leq \beta}$  of probability distributions. Here  $\mu_0$  is an initial distribution that is easy to simulate, and  $\mu_\beta$  is the target distribution that we would like to simulate. Importance sampling and resampling steps are included to constantly adjust for the change of the underlying measure. Whereas for MCMC methods exponential asymptotic stability is usually required to obtain improved samples, the sequential MCMC method starts with a good estimate of  $\mu_0$ , and one only has to control the growth of the “size” of the error. As a consequence, the method sometimes works surprisingly well in multimodal situations where traditional MCMC methods fail, cf. also the examples below. The price one has to pay is that samples from  $\mu_\beta$  cannot be produced individually. Instead, the corresponding algorithm produces directly a Monte Carlo estimator for  $\mu_\beta$  given by the empirical distribution of a system of interacting particles at the final time. To ensure good approximation properties, a large number  $N$  of particles is required.

Variants of such sequential MCMC methods have recently been proposed at several places in the statistics literature, see in particular Del Moral, Doucet and Jasra [5] and references therein, as well as chapters 7–9 in [3]. However, precise and general mathematical methods for the convergence analysis, in the spirit of those developed for traditional MCMC methods by Diaconis, Saloff-Coste, Jerrum, Sinclair, and many others, seem still to be missing – although very important first steps can be found in the work of Del Moral and coauthors, cf. e.g. [4], [6] and [7]. The classical approach via Dobrushin contraction coefficients is usually limited to very regular situations. Moreover, it rarely yields precise statements on the convergence properties, and it can not be combined easily with decomposition techniques.

Our aim is to make variants of the powerful techniques of the spectral gap/Dirichlet form approach to convergence rates of time-homogeneous Markov chains (e.g. canonical paths, comparison and decomposition results) available in the different context of sequential MCMC methods. Mathematically, this means at first to study a class of nonlinear evolutions of probability measures by functional inequalities. Such a study has been initiated in a related context by Stannat [22]. In this work, we restrict ourselves to a very simple and natural variant of sequential MCMC, where importance sampling/resampling is *only* used to adjust constantly for the change of the underlying distribution, and MCMC steps at time  $t$  are always carried out such that detailed balance holds w.r.t. the measure  $\mu_t$  (and not w.r.t.  $\mu_0$ !). This seems crucial for establishing good stability properties. The type of sequential Monte Carlo samplers studied here is different from those analyzed by Del Moral and Doucet in [4]. An algorithmic realization has been applied to simulations in Bayesian mixture models by Del Moral, Doucet and Jasra in [5], who observed substantial benefits compared to other methods.

In a first step, we study the stability properties of nonlinear flows of probability measures describing the limit as the number  $N$  of particles goes to infinity. In the follow-up work [13] we will apply these results to control the asymptotic variances of the Monte Carlo estimators as  $N \rightarrow \infty$ . The functional inequality approach enables us to prove stability properties not only under global but also under local conditions, i.e., if good estimates hold on each set of a decomposition of the state space. As a consequence, we obtain a procedure for analyzing the asymptotics of sequential MCMC methods applied to multimodal distributions. For example, in the spirit of previous results for tempering algorithms by Madras and Zheng [18] and others, we can prove good (polynomial in the system size and the inverse temperature) stability properties in the case of the mean field Ising model.

Let us now become more precise. Let  $S$  denote a finite state space, and  $\mu$  a probability distribution on  $S$  with full support, i.e.  $\mu(x) > 0$  for all  $x \in S$ . The finiteness of the state space is only assumed to keep the presentation as simple and non-technical as possible. Most results of this paper extend to continuous state spaces under standard regularity assumptions. By  $\mathcal{M}_1(S)$  we denote the space of probability measures on  $S$ . As usual,

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

denotes the expectation of a function  $f : S \rightarrow \mathbb{R}$  w.r.t. a measure  $\nu \in \mathcal{M}_1(S)$ . We consider methods for Monte Carlo integration with respect to the probability distributions of an exponential family

$$\mu_t(x) = \frac{1}{Z_t} e^{-tH(x)} \mu(x), \quad 0 \leq t < \infty, \quad (1)$$

where  $H : S \rightarrow [0, \infty)$  is a given function, and  $Z_t := \sum_{x \in S} e^{-tH(x)} \mu(x)$  is a normalization constant. Below,  $t$  will play the rôle of a time parameter for a particle system approximation.

Note that for a fixed  $\beta > 0$ , *any* given probability distribution  $\nu$  on  $S$  that is mutually absolutely continuous with respect to  $\mu$  can be written in the form (1) with  $t = \beta$  by setting  $H(x) = \frac{1}{\beta} \log \frac{\mu(x)}{\nu(x)}$ . One should then think of the family  $(\mu_t)_{0 \leq t \leq \beta}$  of probability distributions as a particular way to interpolate between the target distribution  $\mu_\beta$  that we would like to simulate, and the reference distribution  $\mu_0 = \mu$  that can be simulated more easily. Although we restrict our attention here to this simple way of interpolating between two probability distributions, other interpolations can be treated by similar methods. In fact, an arbitrary family  $(\mu_t)_{0 \leq t \leq \beta}$  of mutually absolutely continuous probability distributions on  $S$  with smooth dependence on  $t$  can be written in the form

$$\mu_t(x) = \frac{1}{Z_t} e^{-\int_0^t U_s(x) ds} \mu(x), \quad 0 \leq t \leq \beta, \quad (2)$$

where  $Z_t$  is a normalization constant, and  $(s, x) \mapsto U_s(x)$  is a continuous non-negative function on  $[0, \beta] \times S$ . Our results below extend to this more general case.

One way to obtain sequential methods for Monte Carlo estimation of expectation values with respect to the measures  $\mu_t$  is to proceed as follows:

- a) Construct a semigroup  $(\Phi_{s,t})_{0 \leq s \leq t < \infty}$  of nonlinear transformations on the space of probability measures on  $S$ , such that

$$\Phi_{s,t} \mu_s = \mu_t \quad \text{for all } 0 \leq s \leq t. \quad (3)$$

- b) *Spatial discretization by interacting particle system*: Construct an appropriate Markov process  $(X_t^1, \dots, X_t^N)$  on  $S^N$  ( $N \in \mathbb{N}$ ) related to the nonlinear semigroup  $\Phi_{s,t}$ , and estimate  $\mu_t = \Phi_{0,t} \mu$  by the empirical distributions

$$\hat{\mu}_t^{(N)} := \frac{1}{N} \sum_{i=1}^N \delta_{X_t^i}, \quad t \geq 0,$$

of the process with initial distribution  $\mu^N$ .

- c) *Time-discretization*: Approximate the continuous time Markov process by a time-discrete Markov chain on  $S^N$  (which can then be simulated).

To define the nonlinear semigroup  $\Phi_{s,t}$  we have in mind, we consider the generators ( $Q$ -matrices)  $\mathcal{L}_t$  at time  $t \geq 0$  of a time-inhomogeneous Markov chain on  $S$  satisfying the detailed balance condition

$$\mu_t(x) \mathcal{L}_t(x, y) = \mu_t(y) \mathcal{L}_t(y, x) \quad \forall t \geq 0, x, y \in S. \quad (4)$$

The generators  $\mathcal{L}_t$  are related to the MCMC steps in a corresponding sequential MCMC method. We assume that  $\mathcal{L}_t(x, y)$  depends continuously on  $t$ . To compare algorithmic performance in a reasonable way, one might also assume

$$\sum_{y \neq x} \mathcal{L}_t(x, y) \leq 1 \quad \forall x \in S, \quad (5)$$

although this is not necessary for the considerations below. For example,  $\mathcal{L}_t$  could be the generator of a Metropolis dynamics w.r.t.  $\mu_t$ , i.e.,

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

$\mathcal{L}_t(x, x) = -\sum_{y \neq x} \mathcal{L}_t(x, y)$ , where the proposal matrix  $K_t$  is a given symmetric transition matrix on  $S$ . By (4),  $\mathcal{L}_t$  defines a symmetric linear operator on  $L^2(S, \mu_t)$ . The associated Dirichlet form on functions  $f, g : S \rightarrow \mathbb{R}$  is

$$\mathcal{E}_t(f, g) := -\mathbb{E}_t[f \mathcal{L}_t g] = \frac{1}{2} \sum_{x, y \in S} (f(y) - f(x))(g(y) - g(x)) \mathcal{L}_t(x, y) \mu_t(x),$$

where  $\mathbb{E}_t$  stands for expectation w.r.t.  $\mu_t$ , and  $(\mathcal{L}_t g)(x) := \sum_y \mathcal{L}_t(x, y)g(y)$ . We shall often use the abbreviated notation  $\mathcal{E}_t(f) := \mathcal{E}_t(f, f)$ .

We also fix non-negative constants  $M_t$  ( $t \geq 0$ ) that will represent the relative frequency of MCMC moves compared to importance sampling/resampling steps. Again, we assume that  $t \mapsto M_t$  is continuous.

Let  $p_{s,t}(x, y)$  and  $q_{s,t}(x, y)$  ( $x, y \in S$ ) be the unique solutions of the forward equations

$$\frac{\partial}{\partial t} p_{s,t} f = p_{s,t}(M_t \mathcal{L}_t f - H f), \quad p_{s,s} f = f, \quad (6)$$

$$\frac{\partial}{\partial t} q_{s,t} f = q_{s,t}(M_t \mathcal{L}_t f - H_t f), \quad q_{s,s} f = f, \quad (7)$$

where

$$H_t := H - \mathbb{E}_t[H].$$

The linear semigroups  $p_{s,t}(x, y)$  and  $q_{s,t}(x, y)$  admit the Feynman-Kac representations

$$\begin{aligned} p_{s,t} f(x) &= \mathbb{E}_{s,x} \left[ e^{-\int_s^t H(X_r) dr} f(X_t) \right], \\ q_{s,t} f(x) &= \mathbb{E}_{s,x} \left[ e^{-\int_s^t H_r(X_r) dr} f(X_t) \right], \end{aligned}$$

where  $(X_t, \mathbb{P}_{t,x})$  is a time-inhomogeneous Markov process with generator  $M_t \cdot \mathcal{L}_t$ , and  $\mathbb{E}_{s,x}$  denotes expectation with respect to  $\mathbb{P}_{s,x}$ . In particular one has

$$q_{s,t} f = \exp \left( \int_s^t \mathbb{E}_r[H] dr \right) p_{s,t} f.$$

We consider the nonlinear semigroup

$$\Phi_{s,t} : \nu \mapsto \frac{\nu p_{s,t}}{(\nu p_{s,t})(S)} = \frac{\nu q_{s,t}}{(\nu q_{s,t})(S)}, \quad 0 \leq s \leq t,$$

on the space  $\mathcal{M}_1(S)$  of probability measures on  $S$ . Here

$$\nu p(y) = \sum_{x \in S} \nu(x) p(x, y).$$

The semigroup  $\Phi_{s,t}$  describes the time evolution of the law of an inhomogeneous Markov chain with generator  $M_t \cdot \mathcal{L}_t$  and absorption rate  $H$ , conditioned to be alive at time  $t$  (see e.g. [2]). It is not difficult to verify that (3) holds, cf. Remark 2 below.

The quality of Monte Carlo estimates of  $\mu_t(f) = \int f d\mu_t$  for some function  $f : S \rightarrow \mathbb{R}$  can be measured by the bias and the (asymptotic) variance of the corresponding estimators. The theoretical analysis of sequential MCMC methods can be subdivided into several steps as above:

- a) Stability properties of the semigroup  $\Phi_{s,t}$ .
- b) Bias and asymptotic variance of the estimators  $\hat{\mu}_t^N(f) = \frac{1}{N} \sum_{i=1}^N f(X_t^i)$ .
- c) Effect of the discretization in time.

In this note, we will focus exclusively on the first step, that is we develop a stability analysis for  $\Phi_{s,t}$  based on functional inequalities. Detailed proofs are given in [12]. A follow-up paper [13] will be devoted to the time dependence of the asymptotic (as  $N \rightarrow \infty$ ) mean square error of the particle system based estimators  $\hat{\mu}_t^N(f)$ . Let us remark for the moment, that significant work in this direction has already been done, e.g., by del Moral and Miclo in [7]. The results clearly indicate that techniques very close to those developed here can also be applied to control the asymptotic variances of the approximating particle systems. This will be made precise in [13].

We also point out that usually the time discretization is carried out before the spatial discretization, i.e. one usually directly considers semigroups and particle systems in discrete time. Even though this is closer to the algorithmic realization, the convergence analysis becomes more transparent in continuous time due to the infinitesimal description (at least from an analytic perspective). Moreover, the continuous time setup allows us to see more clearly how frequently different types of moves of the particle systems should be carried out.

## 2. RESULTS

### 2.1. Time evolution of the mean square error

Let  $\nu_t := \Phi_{0,t}\nu$  for some given initial distribution  $\nu \in \mathcal{M}_1(S)$ , and let

$$g_t(y) := \frac{\nu_t(y)}{\mu_t(y)}, \quad t \geq 0,$$

denote the relative density of  $\nu_t$  w.r.t. the measure  $\mu_t$  defined by (1). Moreover, let

$$\varepsilon_t := \mathbb{E}_t[(g_t - 1)^2]$$

denote the mean square error ( $\chi^2$ -contrast) of  $\nu_t$  w.r.t.  $\mu_t$ . Our first result shows that  $\mu_t = \Phi_{0,t}\mu_0$ , and it gives a general method to analyze the stability of this evolution in an  $L^2$  sense:

**Theorem 1.** (i)  $\nu_t = \Phi_{0,t}\nu$  is the unique solution of the nonlinear evolution equation

$$\frac{\partial}{\partial t}\nu_t = M_t\nu_t\mathcal{L}_t - H\nu_t + \nu_t(H)\nu_t, \quad t \geq 0. \quad (8)$$

with initial condition  $\nu_0 = \nu$ .

(ii) The densities  $g_t$  solve

$$\frac{\partial}{\partial t}g_t = M_t\mathcal{L}_t g_t + \mathbb{E}_t[H(g_t - 1)]g_t. \quad (9)$$

(iii) The time evolution of the mean square error is given by

$$\frac{1}{2}\frac{d}{dt}\varepsilon_t = -M_t\mathcal{E}_t(g_t - 1) - \frac{1}{2}\mathbb{E}_t[H_t(g_t - 1)^2] + \mathbb{E}_t[H_t(g_t - 1)]\varepsilon_t. \quad (10)$$

**Remark 2.** (8) is the forward equation for the nonlinear semigroup  $\Phi_{s,t}$  (for  $s = 0$ ). The corresponding assertion holds for  $\nu_t := \Phi_{s,t}\nu$  for  $t \geq s > 0$ . Since  $\mu_t$  solves (8), we obtain in particular

$$\mu_t = \Phi_{s,t}\mu_s \quad \text{for all } t \geq s \geq 0.$$

Similar equations have been derived in a more general setup by Stannat [22].

Our main objective is to develop efficient tools to bound the growth of  $\varepsilon_t$  based on Theorem 1. To estimate the right-hand side of (10) we have to control the two terms involving  $H_t$  (which correspond to importance sampling/resampling) by the Dirichlet form  $\mathcal{E}_t$  (which corresponds to MCMC moves). We first discuss how this can be achieved in the presence of a good global spectral gap estimate. Afterwards, we give results based on local Poincaré-type inequalities, which can sometimes be used to control the error growth in multimodal setups where good global mixing properties of the underlying Markov chain do not hold.

## 2.2. Stability based on global estimates

For  $t \geq 0$  let

$$C_t := \sup \left\{ \mathbb{E}_t[f^2] / \mathcal{E}_t(f, f) \mid f : S \rightarrow \mathbb{R} \text{ s.t. } \mathbb{E}_t[f] = 0, f \neq 0 \right\}$$

denote the (possibly infinite) inverse spectral gap of  $\mathcal{L}_t$ , and let

$$A_t := \sup \left\{ \mathbb{E}_t[H_t^- f^2] / \mathcal{E}_t(f, f) \mid f : S \rightarrow \mathbb{R} \text{ s.t. } \mathbb{E}_t[f] = 0, f \neq 0 \right\}.$$

Thus  $C_t$  and  $A_t$  are the optimal constants in the global Poincaré inequalities

$$\text{Var}_t(f) \leq C_t \cdot \mathcal{E}_t(f, f) \quad \forall f : S \rightarrow \mathbb{R}, \quad \text{and} \quad (11)$$

$$\mathbb{E}_t[H_t^-(f - \mathbb{E}_t[f])^2] \leq A_t \cdot \mathcal{E}_t(f, f) \quad \forall f : S \rightarrow \mathbb{R}. \quad (12)$$

Here  $\text{Var}_t$  denotes the variance w.r.t.  $\mu_t$ .

**Remark 3.** (i) There exist efficient techniques to obtain upper bounds for  $C_t$ , for example the method of canonical paths, comparison methods (see e.g. [20]), as well as decomposition methods (see e.g. [14]). Variants of these techniques can be applied to estimate  $A_t$  as well.

(ii) Clearly, one has

$$A_t \leq C_t \cdot \sup_{x \in S} H_t^-(x), \quad (13)$$

so an upper bound on  $C_t$  yields a trivial (and usually far from optimal) upper bound on  $A_t$ .

Let

$$\sigma_t(H) := \text{Var}_t(H)^{1/2} = \mathbb{E}_t[H_t^2]^{1/2}$$

denote the standard deviation of  $H$  w.r.t.  $\mu_t$ . The next result bounds the error growth in terms of  $C_t$  and  $A_t$ :

**Theorem 4.** *If  $M_t \geq A_t/2$  for all  $t \geq 0$ , then*

$$\frac{d}{dt} \log \varepsilon_t \leq -\frac{2M_t - A_t}{C_t} + 2\sigma_t(H)\varepsilon_t^{1/2}. \quad (14)$$

As an immediate consequence of the theorem we obtain estimates on the average relative frequency  $M_t$  of MCMC moves that is sufficient to guarantee stability of the corresponding nonlinear flow of probability measures:

**Corollary 5.** *Let  $0 \leq \beta_0 < \beta_1$ , and assume that for all  $t \in (\beta_0, \beta_1)$ ,*

$$M_t > \frac{A_t}{2} + C_t \sigma_t(H) \varepsilon_{\beta_0}^{1/2}. \quad (15)$$

*Then  $t \mapsto \varepsilon_t$  is decreasing on the interval  $[\beta_0, \beta_1]$ .*

**Remark 6.** (i) On the finite state spaces considered here, the constants  $C_t$  and  $A_t$  are finite if  $\mathcal{L}_t$  is irreducible. However, in multimodal situations, the numerical values of these constants are often extremely large. Alternative estimates based on local Poincaré-type inequalities are given below.

(ii) The case  $H \equiv 0$  corresponds to classical MCMC. Here  $A_t = 0$  for all  $t$ , so  $\partial \varepsilon_t / \partial t \leq -2\frac{M_t}{C_t} \varepsilon_t$ . This yields the classical exponential decay with rate  $2\gamma$  of the mean square error in the presence of the global spectral gap  $M_t/C_t \geq \gamma$  of the generator  $M_t \cdot \mathcal{L}_t$ . For  $H \neq 0$ , additional MCMC moves are required to make up for the error growth due to importance sampling/resampling.

Roughly, the corollary tells us that for small initial error, the stabilizing effects of the MCMC dynamics make up for the error growth due to importance sampling/resampling provided  $M_t \geq A_t/2$ .

More drastic improvements due to sequential MCMC appear when good global spectral gap estimates do not hold, as we shall now demonstrate.

### 2.3. Error control based on local estimates

Madras and Randall [17] and Jerrum, Son, Tetali and Vigoda [14] have shown how to derive estimates for spectral gaps and logarithmic Sobolev constants of the generator of a Markov chain from corresponding local estimates on the sets of a decomposition of the state space combined with estimates for the projected chain. This has been applied to tempering algorithms in [18], [1] and [21]. We now develop related decomposition techniques for sequential MCMC. However, in this case, we will assume *only* local estimates for the generators  $\mathcal{L}_t$ , and no mixing properties for the projections – whence there will be an unavoidable error growth due to importance sampling/resampling between the components. The resulting convergence estimates can nevertheless improve those obtained for tempering algorithms, as the example in Section 2.4 below demonstrates. Since mixing properties for the projections do not have to be taken into account, the analysis of the decomposition simplifies considerably.

Let  $0 \leq \beta_0 < \beta_1 \leq \infty$ . We assume that for every  $t \in (\beta_0, \beta_1)$ , there exist a decomposition

$$S = \bigcup_{i \in I} S_t^i$$

into finitely many disjoint sets with  $\mu_t(S_t^i) > 0$ , as well as non-negative definite quadratic forms  $\mathcal{E}_t^i$  ( $i \in I$ ) on functions on  $S$  such that

$$\sum_i \mu_t(S_t^i) \mathcal{E}_t^i(f, f) \leq K \cdot \mathcal{E}_t(f, f) \quad \forall t \in (\beta_0, \beta_1), f : S \rightarrow \mathbb{R} \quad (16)$$

for some fixed finite constant  $K$ . For example, one might choose  $\mathcal{E}_t^i$  as the Dirichlet form of the Markov chain corresponding to  $\mathcal{L}_t$  restricted to  $S_t^i$ , i.e.,

$$\mathcal{E}_t^i(f, f) = \frac{1}{2} \sum_{x, y \in S_t^i} (f(y) - f(x))^2 \mathcal{L}_t(x, y) \mu_t(x | S_t^i). \quad (17)$$

In this case, (16) holds with  $K = 1$ .

Let us denote by  $\mathbb{E}_t^i$  and  $\text{Var}_t^i$ , respectively, the expectation and variance w.r.t. the conditional measure

$$\mu_t^i(A) := \mu_t(A | S_t^i),$$

and by  $\pi : S \rightarrow I$  the natural projection. In particular, for any function  $f$ , one has

$$\mathbb{E}_t[f | \pi] = \sum_{i \in S} \mathbb{E}_t^i[f] \cdot \chi_{S_t^i}.$$

We set

$$\tilde{H}_t := H - \mathbb{E}_t[H | \pi].$$

Assume that the following local Poincaré type inequalities hold for all  $t \in (\beta_0, \beta_1)$  and  $i \in I$  with constants  $A_t^i, B_t^i \in (0, \infty)$  :

$$\mathbb{E}_t^i[-\tilde{H}_t f^2] \leq A_t^i \cdot \mathcal{E}_t^i(f, f) \quad \forall f : S \rightarrow \mathbb{R} : \mathbb{E}_t[f | \pi] = 0, \quad (18)$$

$$|\mathbb{E}_t^i[\tilde{H}_t f]|^2 \leq B_t^i \cdot \mathcal{E}_t^i(f, f) \quad \forall f : S \rightarrow \mathbb{R} : \mathbb{E}_t[f | \pi] = 0. \quad (19)$$

**Remark 7.** (i) Note that to verify (18) it is enough to estimate  $\mathbb{E}_t^i[\tilde{H}_t^- f^2]$ , while for (19) one has to take into account the positive part of  $\tilde{H}_t$  as well. In particular, (18) can not be used to derive an estimate of type (19). However, if (18) holds with  $-\tilde{H}_t$  replaced by  $|\tilde{H}_t|$ , then (19) holds with  $B_t^i = \mathbb{E}_t^i[|\tilde{H}_t|] \cdot A_t^i$ .

(ii) If local Poincaré inequalities of the type

$$\text{Var}_t^i(f) \leq C_t^i \cdot \mathcal{E}_t^i(f, f) \quad \forall f : S \rightarrow \mathbb{R}, \quad i \in I, \quad (20)$$

hold, then (18) and (19) hold with  $A_t^i = C_t^i \cdot \max_{S_i} \tilde{H}_t^-$  and  $B_t^i = C_t^i \cdot \text{Var}_t^i(H)$ .

Let

$$\hat{A}_t := K \cdot \max_i A_t^i \quad \text{and} \quad \hat{B}_t := K \cdot \max_i B_t^i.$$

The central result of this section is the following estimate on the evolution of the error  $\varepsilon_t$  :

**Theorem 8.** *If  $M_t > \hat{A}_t/2$  for all  $t \in (\beta_0, \beta_1)$  then*

$$\frac{d}{dt} \log \varepsilon_t \leq \frac{\hat{B}_t}{2M_t - \hat{A}_t} \cdot (1 + \varepsilon_t) + (1 + \sqrt{\varepsilon_t})^2 \cdot \max_{i \in I} h_t^-(i) \quad (21)$$

where

$$h_t(i) := \mathbb{E}_t^i[H] - \mathbb{E}_t[H] = - \left. \frac{\partial}{\partial s} \log \mu_s(S_t^i) \right|_{s=t} \quad (i \in I). \quad (22)$$

To understand the consequences of (21), let us first consider the asymptotics as  $M_t$  tends to infinity. In this case, (21) reduces to

$$\frac{d}{dt} \log \varepsilon_t \leq (1 + \sqrt{\varepsilon_t})^2 \cdot \max h_t^-.$$

In order to ensure that for  $t > \beta_0$  the error  $\varepsilon_t$  remains below a given threshold  $\delta > 0$ , note that as long as  $\varepsilon_t \leq \delta$ , we have

$$\frac{d}{dt} \log \varepsilon_t \leq (1 + \sqrt{\delta})^2 \cdot \max h_t^-.$$

Thus

$$\min(\varepsilon_t, \delta) \leq \varepsilon_{\beta_0} \cdot G_t^{(1+\sqrt{\delta})^2} \quad \forall t \in [\beta_0, \beta_1], \quad (23)$$

where

$$G_t := \exp \left( \int_{\beta_0}^t \max h_r^- dr \right) = \exp \left( \int_{\beta_0}^t \max_i \left. \frac{\partial}{\partial s} \log \mu_s(S_r^i) \right|_{s=r} dr \right).$$

**Remark 9.** The term  $G_t^{(1+\sqrt{\delta})^2}$  in (23) accounts for the maximum error growth due to importance sampling between the components. If  $S_t^i = S^i$  is independent of  $t$  for every  $i$ , and there is an  $i_0 \in I$  such that  $\left. \frac{\partial}{\partial s} \log \mu_s(S^i) \right|_{s=t}$  is maximized by  $S^{i_0}$  for all  $s \in (\beta_0, \beta_1)$ , then

$$G_t = \exp \left( \int_{\beta_0}^t \max_i \frac{d}{ds} \log \mu_s(S^i) ds \right) = \frac{\mu_t(S^{i_0})}{\mu_{\beta_0}(S^{i_0})} \quad \forall t \geq \beta_0$$

is just the growth rate of this strongest growing component. In general, things are more complicated, but a similar interpretation is at least possible on appropriate subintervals of  $[\beta_0, \beta_1]$

Now we return to the case when  $M_t$  is finite. The next corollary tells us how many MCMC moves are sufficient to obtain an estimate on the error growth that is not much worse than (23).



**Corollary 10.** *Let  $\beta \in (\beta_0, \beta_1]$  and  $\delta > 0$ . If*

$$M_t \geq \frac{\hat{A}_t}{2} + (\beta - \beta_0) \cdot \hat{B}_t \quad \forall t \in (\beta_0, \beta), \quad (24)$$

then

$$\min(\varepsilon_\beta, \delta) \leq \varepsilon_{\beta_0} \cdot G_\beta^{(1+\sqrt{\delta})^2} \cdot e^{\frac{1+\delta}{2}}. \quad (25)$$

#### 2.4. Example: The mean field Ising model

As a simple example of a bimodal distribution, we now consider the mean field Ising (Curie–Weiss) model, i.e.  $\mu_\beta$  is of type (1) where  $\mu_0 = \mu$  is the uniform distribution on the hypercube

$$S = \{-1, +1\}^N,$$

and

$$H(\sigma) = -\frac{1}{2N} \sum_{i,j=1}^N \sigma_i \sigma_j \quad (26)$$

for some  $N \in \mathbb{N}$ . Let  $\mathcal{L}_\beta$  be the generator of the (time–continuous) Metropolis chain w.r.t.  $\mu_\beta$  based on the nearest neighbour random walk on  $S$  as proposal matrix. It is well known that this chain is rapidly mixing (i.e. the spectral gap decays polynomially in  $N$ ) for  $\beta < 1$ , but torpid mixing holds (i.e. the spectral gap decays exponentially fast in  $N$ ) for  $\beta > 1$ . Thus in the multi-phase regime  $\beta > 1$ , the classical Metropolis algorithm converges to equilibrium extremely slowly for large  $N$ .

Now assume for simplicity that  $N$  is odd, and decompose  $S$  into the two components

$$\begin{aligned} S^+ &:= \left\{ \sigma \in S \mid \sum_{i=1}^N \sigma_i > 0 \right\} \quad \text{and} \\ S^- &:= \left\{ \sigma \in S \mid \sum_{i=1}^N \sigma_i < 0 \right\}. \end{aligned}$$

Improving on previous results (e.g. of Madras and Zheng [18]), Schweizer [21] showed recently that the spectral gaps of the restricted Metropolis chains on both  $S^+$  and  $S^-$  are bounded from below by  $\frac{1}{9}N^{-2}$  for every  $\beta \geq 0$ . Applying the results above to the error growth for sequential MCMC in this situation, we obtain :

**Theorem 11.** *For every  $\beta > 0$  and  $N \in \mathbb{N}$ ,*

$$\sup_{0 \leq t \leq \beta} \varepsilon_t \leq e \cdot \varepsilon_0$$

holds whenever  $\varepsilon_0 \leq 1$  and

$$M_t \geq \frac{9}{4}N^3 + \frac{9}{8}\beta N^4 \quad \forall t \in (0, \beta). \quad (27)$$

**Remark 12.** Let  $K_\beta = \int_0^\beta M_t dt$ . Note that  $K_\beta$  is a measure for the total number of MCMC steps that a corresponding sequential MCMC algorithm will perform on average. The theorem implies that choosing  $M_t$  constant on  $[0, \beta]$  with  $K_\beta$  of order  $O(\beta^2 N^4)$  is sufficient to guarantee that the nonlinear flow of measures has good stability properties on  $[0, \beta]$ , and can thus be used to efficiently approximate  $\mu_\beta$ . In contrast to this situation, the flow of measures corresponding to the standard simulated annealing algorithm has good stability properties only if  $K_\beta$  grows exponentially both in  $\beta$  and  $N$ .

## REFERENCES

- [1] N. Bhatnagar and D. Randall, *Torpid mixing of simulated tempering on the Potts model*, SODA '04: Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms (Philadelphia, PA, USA), Society for Industrial and Applied Mathematics, 2004, pp. 478–487.
- [2] R. M. Blumenthal and R. K. Gettoor, *Markov processes and potential theory*, Pure and Applied Mathematics, Vol. 29, Academic Press, New York, 1968. MR MR0264757 (41 #9348)
- [3] O. Cappé, E. Moulines, and T. Rydén, *Inference in hidden Markov models*, Springer Series in Statistics, Springer, New York, 2005. MR MR2159833 (2006e:60002)
- [4] P. Del Moral and A. Doucet, *On a class of genealogical and interacting Metropolis models*, Séminaire de Probabilités XXXVII, Lecture Notes in Math., vol. 1832, Springer, Berlin, 2003, pp. 415–446. MR MR2053058 (2005g:65013)
- [5] P. Del Moral, A. Doucet, and A. Jasra, *Sequential Monte Carlo samplers*, J. R. Statist. Soc. B **68** (2006), no. 3, 411–436.
- [6] 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)
- [7] P. Del Moral and L. Miclo, *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)
- [8] P. Diaconis and L. Saloff-Coste, *Comparison theorems for reversible Markov chains*, Ann. Appl. Probab. **3** (1993), no. 3, 696–730. MR MR1233621 (94i:60074)
- [9] ———, *Logarithmic Sobolev inequalities for finite Markov chains*, Ann. Appl. Probab. **6** (1996), no. 3, 695–750. MR MR1410112 (97k:60176)
- [10] ———, *Nash inequalities for finite Markov chains*, J. Theoret. Probab. **9** (1996), no. 2, 459–510. MR MR1385408 (97d:60114)
- [11] ———, *What do we know about the Metropolis algorithm?*, J. Comput. System Sci. **57** (1998), no. 1, 20–36, 27th Annual ACM Symposium on the Theory of Computing (STOC'95) (Las Vegas, NV). MR MR1649805 (2000b:68094)
- [12] A. Eberle and C. Marinelli, *Convergence of sequential Markov chain Monte Carlo methods: I. Nonlinear flow of probability measures*, Preprint.
- [13] A. Eberle and C. Marinelli, *Convergence of sequential Markov chain Monte Carlo methods: II. Asymptotic analysis of interacting particle systems*, In preparation.
- [14] M. Jerrum, J.-B. Son, P. Tetali, and E. Vigoda, *Elementary bounds on Poincaré and log-Sobolev constants for decomposable Markov chains*, Ann. Appl. Probab. **14** (2004), no. 4, 1741–1765. MR MR2099650 (2005i:60139)
- [15] J. S. Liu, *Monte Carlo strategies in scientific computing*, Springer-Verlag, New York, 2001. MR MR1842342 (2002i:65006)
- [16] N. Madras and M. Piccioni, *Importance sampling for families of distributions*, Ann. Appl. Probab. **9** (1999), no. 4, 1202–1225. MR MR1728560 (2001e:60139)
- [17] N. Madras and D. Randall, *Markov chain decomposition for convergence rate analysis*, Ann. Appl. Probab. **12** (2002), no. 2, 581–606. MR MR1910641 (2003d:60135)
- [18] N. Madras and Z. Zheng, *On the swapping algorithm*, Random Structures Algorithms **22** (2003), no. 1, 66–97. MR MR1943860 (2004c:82117)
- [19] C. P. Robert and G. Casella, *Monte Carlo statistical methods*, second ed., Springer-Verlag, New York, 2004. MR MR2080278 (2005d:62006)
- [20] 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)
- [21] N. Schweizer, *Mixing rates of stochastic dynamics on mean field Ising models and multi-level MCMC methods*, Diploma thesis, Universität Bonn, 2006.
- [22] 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)