Uniform ergodicity of the iterated conditional SMC and geometric ergodicity of particle Gibbs samplers

CHRISTOPHE ANDRIEU[1], ANTHONY LEE[2] and MATTI VIHOLA[3]

[1]*School of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, United Kingdom.*
E-mail: c.andrieu@bristol.ac.uk
[2]*Department of Statistics, University of Warwick, Coventry CV4 7AL, United Kingdom.*
E-mail: anthony.lee@warwick.ac.uk
[3]*Department of Mathematics and Statistics, University of Jyväskylä, P.O. Box 35, Finland.*
E-mail: matti.vihola@iki.fi

We establish quantitative bounds for rates of convergence and asymptotic variances for iterated conditional sequential Monte Carlo (i-cSMC) Markov chains and associated particle Gibbs samplers [*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** (2010) 269–342]. Our main findings are that the essential boundedness of potential functions associated with the i-cSMC algorithm provide necessary and sufficient conditions for the uniform ergodicity of the i-cSMC Markov chain, as well as quantitative bounds on its (uniformly geometric) rate of convergence. Furthermore, we show that the i-cSMC Markov chain cannot even be geometrically ergodic if this essential boundedness does not hold in many applications of interest. Our sufficiency and quantitative bounds rely on a novel non-asymptotic analysis of the expectation of a standard normalizing constant estimate with respect to a "doubly conditional" SMC algorithm. In addition, our results for i-cSMC imply that the rate of convergence can be improved arbitrarily by increasing N, the number of particles in the algorithm, and that in the presence of mixing assumptions, the rate of convergence can be kept constant by increasing N linearly with the time horizon. We translate the sufficiency of the boundedness condition for i-cSMC into sufficient conditions for the particle Gibbs Markov chain to be geometrically ergodic and quantitative bounds on its geometric rate of convergence, which imply convergence of properties of the particle Gibbs Markov chain to those of its corresponding Gibbs sampler. These results complement recently discovered, and related, conditions for the particle marginal Metropolis–Hastings (PMMH) Markov chain.

Keywords: geometric ergodicity; iterated conditional sequential Monte Carlo; Metropolis-within-Gibbs; particle Gibbs; uniform ergodicity

1. Introduction

Particle Markov chain Monte Carlo (P-MCMC) methods are a set of recently proposed sampling techniques particularly well suited to the Bayesian estimation of static parameters in general state–space models [1], although their scope extends beyond this class of models. At an abstract level, once the likelihood function and prior are defined, inference for this class of models relies on a probability distribution $\pi(d\theta \times dx)$, defined on some measurable space $(\Theta \times X, \mathcal{B}(\Theta) \times \mathcal{B}(X))$, where θ is generally a low dimensional static parameter, the static parameter, while x, the hidden state of the system, is a large vector with a non-trivial dependence

structure. Here, $\mathcal{B}(\cdot)$ denotes the σ-algebra related to the corresponding space. In practice, the complexity of such probability distributions requires the use of sampling techniques to effectively carry out inference. When θ is known sequential Monte Carlo methods (SMC), or particle filters, are particularly suitable to carry out inference about x by approximately sampling from the conditional distribution $\pi_\theta(\mathrm{d}x)$. These algorithms rely on interacting particle systems and their performance and accuracy can be improved by increasing the number N of such particles. P-MCMC realises the synthesis between SMC methods and classical Markov chain Monte Carlo (MCMC) methods, that is it allows the construction of Markov transition probabilities leaving $\pi(\mathrm{d}\theta \times \mathrm{d}x)$ at least marginally invariant and from which it is possible to sample realisations $\{(\theta_i, X_i), i \geq 0\}$ with attractive efficiency properties.

The particle marginal Metropolis–Hastings (PMMH) method is one such algorithm, which takes advantage of the availability of unbiased estimators of the likelihood function to provide an exact approximation of an idealized algorithm which computes the likelihood function exactly. The algorithm simply consists of replacing the true value of the likelihood function required to implement the standard Metropolis–Hastings (MH) algorithm with estimators, but is nevertheless guaranteed to be correct in that it leaves the required distribution of interest marginally invariant. In PMMH, the estimator of the likelihood is a byproduct of a sequential Monte Carlo (SMC) algorithm, whose accuracy can be improved by increasing N.

In contrast, the particle Gibbs (PGibbs) sampler [1] involves approximating a Gibbs sampler which consists of constructing a Markov chain $\{(\theta_i, X_i), i \geq 0\}$, by repeatedly sampling from $\pi_\theta(\mathrm{d}x)$ and $\pi_x(\mathrm{d}\theta)$ in turn. In practice, sampling from $\pi_\theta(\mathrm{d}x)$ may be particularly difficult and the conditional SMC (cSMC) [1] update is a Markov transition probability $P_{N,\theta}$ which leaves $\pi_\theta(\mathrm{d}x)$ invariant, therefore allowing the implementation of a Metropolis-within-Gibbs algorithm, that is a Markov transition probability leaving $\pi(\mathrm{d}\theta \times \mathrm{d}x)$ invariant. The cSMC relies for its construction, as suggested by its name, on an SMC-like procedure and it is expected that as N increases $P_{N,\theta}$ approaches $\pi_\theta(\mathrm{d}x)$.

While PMMH methods have been studied in a series of papers [4,5,11,14,24], a theoretical study of the PGibbs is still missing. Indeed it has been shown that as N increases, performance of the PMMH approaches that of the exact MH algorithm but the question of the approximation of the Gibbs sampler by a PGibbs has not been addressed to date. We note however that a study of one of its components, the cSMC update, has recently been undertaken in [8], in which a coupling argument is central to their analysis. We refer to the Markov chain obtained by iterating the cSMC algorithm for a fixed target distribution as iterated i-cSMC here in order to distinguish it from that of the PGibbs. The present manuscript addresses questions concerning the i-cSMC similar to those of [8], but our results differ in many respects and complement their findings in several directions. At a technical level, our approach seems to be more straightforward in the scenario considered, relies on weaker assumptions for uniform convergence which we prove are necessary and sufficient and lead to quantitative bounds on performance measures in terms of the number N of particles involved. We additionally transfer sufficient conditions for uniform ergodicity of the i-cSMC Markov chain into sufficient conditions for geometric ergodicity of the associated PGibbs Markov chain, the main motivation behind our work. This allows us in particular to show that under some conditions PGibbs is asymptotically as efficient as the Gibbs sampler as the number N of particles increases.

Contemporary to the first version of the present manuscript [3,16] have also provided essentially the same sufficient conditions for the uniform convergence of the i-cSMC Markov chain

(Theorem 1, Section 3) using a different proof technique. Here we have further established that the aforementioned conditions are also necessary for uniform convergence in general, but also geometric ergodicity in many realistic scenarios (Section 6). Similarly to us [16] also provide quantitative bounds and associated scaling properties of the i-cSMC, albeit for a different set of specialised conditions (a detailed comparison of the assumptions is provided after Theorem 3 at the end of Section 3). We have also very recently become aware of the contribution [10] to the analysis of the properties of the cSMC, established using the formalism of [9], but their practical implications are unclear. Similarly to [8,16] do not attempt to address the practically important question of how uniform ergodicity of the i-cSMC can be translated into geometric ergodicity of the PGibbs sampler, an issue we address in Section 7. In Section 8, we contrast the results obtained in this paper concerning the i-cSMC and PGibbs algorithm with known results concerned with other particle MCMC methods and draw final conclusions.

Similarly to SMC methods, the cSMC and associated algorithms are complex mathematical objects which require the introduction of sometimes overwhelming notation which may obscure the main ideas. In the next section, we attempt to remedy this by presenting our results in a simplified scenario, which captures our main ideas, before moving on to the general scenario.

2. Statement of our results in a simplified scenario

We first explain our results on a particularly simple instance of the i-cSMC algorithm. This should provide the reader with the essence of the results proved later on in the general scenario, while its simple structure will allow us to outline the main idea behind our proof in the general set-up (in Section 4).

Assume we are interested in sampling from a probability distribution π on some measurable space $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$. We define the probability distribution $\tilde{\pi}$ on $\{1, \ldots, N\} \times \mathsf{X}^N$

$$\tilde{\pi}\left(k, \mathrm{d}z^{1:N}\right) = \frac{1}{N}\pi\left(\mathrm{d}z^k\right) \prod_{j=1, j \neq k}^{N} M\left(\mathrm{d}z^j\right), \tag{1}$$

for some probability distribution M defined on $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ and such that for any $S \in \mathcal{B}(\mathsf{X})$ such that $\pi(S) > 0$ then $M(S) > 0$. As pointed out in the authors' discussion reply of [1], in this simple scenario one can define an MCMC algorithm targeting π by iterating the classical sampling importance resampling (SIR) procedure. More specifically, we sample alternately from (a) $Z^{1:N \setminus k} \mid (K = k, Z^k = z^k) \sim \prod_{i=1, i \neq k}^{N} M(z^i)$ and (b) $K \mid (Z^{1:N} = z^{1:N}) \sim \tilde{\pi}(k|z^{1:N})$, where $Z^{1:N \setminus k} := (Z^1, Z^2, \ldots, Z^{k-1}, Z^{k+1}, \ldots, Z^N)$. Owing to the fact that this algorithm is a Gibbs sampler on the distribution above and from the standard interlacing property of the two stage Gibbs sampler, one can check that the sequence $\{Z_i^{K_i}\}$ defines a Markov chain with invariant distribution π, and that its transition kernel is for any $(x, S) \in \mathsf{X} \times \mathcal{B}(\mathsf{X})$

$$P_N(x, S) = \int_{\mathsf{X}^{N-1}} \sum_{k=1}^{N} \frac{G(z^k)}{\sum_{j=1}^{N} G(z^j)} \mathbb{I}\{z^k \in S\} \prod_{i=2}^{N} M\left(\mathrm{d}z^i\right)$$

with $G(x) := \pi(dx)/M(dx)$ and the convention $z^1 = x$. Our first results are concerned with properties of the homogeneous Markov chain with transition probability P_N, in terms of $\bar{G} := \pi\text{-ess sup}_x G(x)$ and N. We refer to the resulting algorithm as iterated SIR (i-SIR).

We briefly introduce notions that allow us to make quantitative statements about the Markov chains under study. We use classical Hilbert space techniques for the analysis of reversible Markov chains. Letting $\mu(\cdot)$ be a probability distribution defined on some measurable space $(\mathsf{E}, \mathcal{B}(\mathsf{E}))$, we define the function space

$$L^2(\mathsf{E}, \mu) := \{ f : \mathsf{E} \to \mathbb{R} : \mu(f^2) < \infty \},$$

where the functions are taken to be measurable; hereafter all functions considered are assumed to be measurable with respect to an appropriate σ-algebra. Let $\Pi : \mathsf{E} \times \mathcal{B}(\mathsf{E}) \to [0, 1]$ be a μ-reversible Markov transition kernel and let $\{\xi_i, i \geq 0\}$ be the stationary Markov chain with transition kernel Π (such that $\xi_0 \sim \mu$). We will use the standard notation for any probability distribution ν on $(\mathsf{E}, \mathcal{B}(\mathsf{E}))$ and measurable function $f : \mathsf{E} \to \mathbb{R}$,

$$\nu(f) := \int_\mathsf{E} f(x)\nu(dx), \qquad \Pi f(x) := \int_\mathsf{E} f(y)\Pi(x, dy),$$

for $k \geq 2$, by induction,

$$\Pi^k f(x) := \int_\mathsf{E} \Pi(x, dy)\Pi^{k-1} f(y).$$

We denote $\nu\Pi^k f := \nu(\Pi^k f)$ and refer to $\nu\Pi^k$ as either a probability measure or its corresponding operator on $L^2(\mathsf{E}, \mu)$. For $f \in L^2(\mathsf{E}, \mu)$, we define the variance of f under μ as $\text{var}_\mu(f) := \mu(f^2) - \mu(f)^2$ and the "asymptotic variance" of $M^{-1}\sum_{i=1}^M f(\xi_i)$ for stationary realizations $\{\xi_i, i \geq 0\}$ associated to the homogeneous Markov chain with transition Π as

$$\text{var}(f, \Pi) := \lim_{M \to \infty} \text{var}\left(M^{-1/2} \sum_{i=1}^M [f(\xi_i) - \mu(f)] \right).$$

Some of our results involve norms of signed measures. As in, for example, [21], for any signed measure ν on $(\mathsf{E}, \mathcal{B}(\mathsf{E}))$ we let

$$\|\nu\|_{\text{TV}} := \frac{1}{2} \sup_{f:\mathsf{E}\to[-1,1]} \nu(f)$$

denote the total variation distance and for $\nu \ll \mu$,

$$\|\nu\|^2_{L^2(\mathsf{E},\mu)} := \int_\mathsf{E} \left| \frac{d\nu}{d\mu} \right|^2 d\mu = \sup_{f \in L^2(\mathsf{E},\mu), \|f\|_\mu > 0} \frac{|\nu(f)|}{\|f\|_\mu} \tag{2}$$

denote the $L^2(\mathsf{E}, \mu)$ norm.

Our results can be summarized as follows:

1. P_N is reversible with respect to π and positive, that is the i-SIR Markov chain has non-negative stationary autocorrelations.
2. If $\bar{G} < \infty$, and $N \geq 2$, the i-SIR Markov chain is uniformly ergodic with for any $x \in \mathsf{X}$,

$$\left\| P_N^n(x, \cdot) - \pi(\cdot) \right\|_{\mathrm{TV}} \leq \left(1 - \frac{N-1}{2\bar{G} + N - 2} \right)^n.$$

3. If $\bar{G} < \infty$, then for any $f \in L^2(\mathsf{X}, \pi)$,

$$\mathrm{var}_\pi(f) \leq \mathrm{var}(f, P_N) \leq \left[2\left(1 + \frac{2\bar{G} - 1}{N - 1} \right) - 1 \right] \mathrm{var}_\pi(f).$$

4. If $\bar{G} = \infty$, then the i-SIR Markov chain cannot be geometrically ergodic for any finite N.

The second and third points provide quantitative bounds on standard measures of performance for MCMC algorithms, where the second provides a bound on the uniform (or equivalently uniformly geometric) rate of convergence of the Markov chain. Interest in algorithms such as i-SIR is motivated empirically from observed behaviour in line with the above bounds, as performance improves as N increases, and part of our purpose here is to confirm and quantify theoretically such empirical successes. Moreover, this improvement can often be obtained with little extra computational effort, since on a parallel architecture one can sample from M and evaluate G in parallel, a characteristic of SMC algorithms more generally [15].

While i-SIR can be used alone to sample from fairly general distributions, it can also be used as a constituent element of more elaborate MCMC schemes. Assume now that we wish to sample from a distribution π defined on some measurable space $(\Theta \times \mathsf{X}, \mathcal{B}(\Theta) \times \mathcal{B}(\mathsf{X}))$, often defined for some $S \in \mathcal{B}(\Theta) \times \mathcal{B}(\mathsf{X})$ via (note the different nature of π as compared to earlier)

$$\pi(S) := \frac{\int_S G_\theta(x) M_\theta(\mathrm{d}x) \varpi(\mathrm{d}\theta)}{\int_{\Theta \times \mathsf{X}} G_\theta(x) M_\theta(\mathrm{d}x) \varpi(\mathrm{d}\theta)},$$

where $\{G_\theta, \theta \in \Theta\}$ is a collection of non-negative potential functions and $\{M_\theta, \theta \in \Theta\}$ a collection of probability measures which define for each $\theta \in \Theta$ the conditional distributions $\pi_\theta(\mathrm{d}x) := M_\theta(\mathrm{d}x) G_\theta(x) / \gamma_\theta$ with

$$\gamma_\theta := \int_\mathsf{X} G_\theta(x) M_\theta(\mathrm{d}x).$$

The interpretation in a statistical context is that ϖ is the prior distribution for some parameter θ of interest, whilst γ_θ is the likelihood function associated with some observed data and x corresponds to the so-called latent variable(s). The form of γ_θ is often derived from the data being explained by the latent variable x whose a priori distribution conditional upon θ is M_θ and the likelihood function given the data and x is $G_\theta(x)$. Assume here that we are able to sample from π_x, the conditional distribution of θ given $X = x$. For any $\theta \in \Theta$, one can define the i-SIR kernel for any $(x, S) \in \mathsf{X} \times \mathcal{B}(\mathsf{X})$ via

$$P_{N,\theta}(x, S) = \int_{\mathsf{X}^{N-1}} \sum_{k=1}^N \frac{G_\theta(z^k)}{\sum_{j=1}^N G_\theta(z^j)} \mathbb{I}\{z^k \in S\} \prod_{i=2}^N M_\theta(\mathrm{d}z^i),$$

with $z^1 = x$, so that the invariant distribution associated with $P_{N,\theta}$ is π_θ, the conditional distribution of X given θ. One can sample from $\pi(\mathrm{d}\theta \times \mathrm{d}x)$ with the following Markov transition, defined for any $(\theta_0, x, S) \in \Theta \times X \times (\mathcal{B}(\Theta) \times \mathcal{B}(X))$ via

$$\Phi_N(\theta_0, x; S) := \int_S P_{N,\theta}(x, \mathrm{d}y)\pi_x(\mathrm{d}\theta),$$

which can be viewed as an exact approximation of the Gibbs sampler defined via

$$\Gamma(\theta_0, x; S) := \int_S \pi_\theta(\mathrm{d}y)\pi_x(\mathrm{d}\theta).$$

The term exact approximation refers to the fact that while $P_{N,\theta}$ can be thought of as an approximation of the conditional distribution π_θ the resulting algorithm converges to π and can be made arbitrarily close to Γ as we increase N as explained below – we will refer to this algorithm and its generalisation as the particle Gibbs (PGibbs) sampler. Throughout the paper, we will use the following convention: we will say $f \in L^2(E, \pi)$ with $E = \Theta$ (resp., $E = X$) to mean that $f : E \to \mathbb{R}$ is square integrable under the relevant marginal of π, or $f : \Theta \times X \to \mathbb{R}$ does not depend on x (resp., θ) and is square integrable under the relevant marginal of π. This should not lead to any possible confusion. Letting $\bar{G} := \pi\text{-ess sup}_{\theta,x} \frac{G_\theta(x)}{\gamma_\theta}$, our results for the PGibbs sampler, are as follows:

1. Assume the Γ Markov chain is such that there exists $\beta \in (0, 1]$ such that for any $f : X \to [-1, 1]$ and $\nu \ll \pi$

$$\left|\nu\Gamma^n(f) - \pi(f)\right| \le \|\nu - \pi\|_{L^2(X,\pi)}(1 - \beta)^n.$$

If $\bar{G} < \infty$, and $N \ge 2$, then for any $f : X \to [-1, 1]$ and $\nu \ll \pi$

$$\left|\nu\Phi_N^n(f) - \pi(f)\right| \le \|\nu - \pi\|_{L^2(X,\pi)}\left(1 - \beta_N'\right)^n,$$

where β_N' satisfies

$$\beta_N' \ge \frac{N-1}{2\bar{G} + N - 2}\beta.$$

2. For any $f \in L^2(X, \pi)$ and $N \ge 2$, the asymptotic variance $\mathrm{var}(f, \Phi_N)$ satisfies

$$\mathrm{var}(f, \Gamma) \le \mathrm{var}(f, \Phi_N) \le \frac{2\bar{G} - 1}{N - 1}\mathrm{var}_\pi(f) + \left(1 + \frac{2\bar{G} - 1}{N - 1}\right)\mathrm{var}(f, \Gamma).$$

3. For any $f \in L^2(\Theta, \pi)$ and $N \ge 2$, the asymptotic variance $\mathrm{var}(f, \Phi_N)$ satisfies

$$\mathrm{var}(f, \Gamma) \le \mathrm{var}(f, \Phi_N) \le \left(1 + \frac{2\bar{G} - 1}{N - 1}\right)\mathrm{var}(f, \Gamma) - \left(\frac{2\bar{G} - 1}{N - 1}\right)\mathrm{var}_\pi(f).$$

In the sequel, we prove similar results in the more general (and complex) scenario where $P_{N,\theta}$ is defined by a general cSMC algorithm with multinomial resampling, but the key ideas and results are similar (Section 3). The results concerning the general form of the PGibbs sampler, from which its convergence in the sense of points 1–3 above follows, can be found in Section 7.

3. The i-cSMC and its properties

We mostly follow the notation of [9] and use the following conventions for lists, indices and superscripts. For $N \in \mathbb{N}$, we denote $[N] := \{1, \ldots, N\}$, and for any $p \in \mathbb{N}$, $\mathbf{k}, \mathbf{l} \in [N]^p$ and $u_k^l : \mathbb{N}^2 \to \mathsf{E}$ (for a generic set E dependent on the context) we will use the notation $u_{\mathbf{k}}^{\mathbf{l}}$ to mean $(u_{k_1}^{l_1}, u_{k_2}^{l_2}, \ldots, u_{k_p}^{l_p})$, and whenever there is no dependence on l (resp., k) of u_k^l we simply ignore this superscript (resp., this index). We will also use the notation, for $k, l \in \mathbb{N}$ such that $l \geq k$, $k : l := (k, k+1, \ldots, l)$. Let $(\mathsf{Z}, \mathcal{B}(\mathsf{Z}))$ be a measurable space and for some $T \geq 1$ define a family of Markov transition probabilities on this space $\{M_t(\cdot, \cdot), t \in [T]\}$ with the convention that for $t = 1$ and any $z \in \mathsf{Z}$, $M_1(z, \mathrm{d}u) = M_1(\mathrm{d}u)$ and a family of measurable non-negative functions, the potentials $G_t : \mathsf{Z} \to [0, \infty)$, again for $t \in [T]$. We first define an inhomogeneous Markov chain $\{Z_1, \ldots, Z_T\}$ on $\mathsf{X} := \mathsf{Z}^T$ endowed with the product σ-algebra $\mathcal{B}(\mathsf{X}) = \mathcal{B}(\mathsf{Z})^T$ and with probability distribution $\mathbb{P}(\cdot)$ and associated expectation $\mathbb{E}(\cdot)$ such that for $t = 1$, the initial distribution is $\mathbb{P}(Z_1 \in \mathrm{d}z_1) := M_1(\mathrm{d}z_1)$, and for $t = 2, \ldots, T$ the transition probability is given by M_t, i.e.,

$$\mathbb{P}(Z_t \in \mathrm{d}z_t | Z_{t-1} = z_{t-1}) := M_t(z_{t-1}, \mathrm{d}z_t).$$

We define for $p \in [T]$ and $f_p : \mathsf{Z}^p \to \mathbb{R}$

$$\gamma_p(f_p) := \mathbb{E}\left(f_p(Z_1, \ldots, Z_p) \prod_{t=1}^{p} G_t(Z_t) \right),$$

and can define for any $S \in \mathcal{B}(\mathsf{X})$ the probability distribution π (which will be the target distribution of interest)

$$\pi(S) := \frac{\gamma_T(\mathbb{I}\{\cdot \in S\})}{\gamma_T}, \tag{3}$$

where $\mathbb{I}\{\cdot\}$ denotes the indicator function and $\gamma_T := \gamma_T(1)$. For $l > k \geq 0$, we define

$$M_{k,l}(z_k, \mathrm{d}z_{k+1:l}) := \prod_{t=k+1}^{l} M_t(z_{t-1}, \mathrm{d}z_t).$$

Note in particular that with the convention above, for any $l \geq 2$ and $z_0 \in \mathsf{Z}$, $M_{0,l}(z_0, \mathrm{d}z_{1:l}) := M_1(\mathrm{d}z_1) \times M_{1,l}(z_1, \mathrm{d}z_{2:l})$.

The iterated conditional SMC (i-cSMC) is a family of homogeneous Markov chains, with state–space $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$, indexed by $N \in \mathbb{N}$ (the concrete meaning of N shall become clearer below). We denote by $P_N(\cdot, \cdot) : \mathsf{X} \times \mathcal{B}(\mathsf{X}) \to [0, 1]$ the corresponding Markov transition kernels, which we now define. To that end, we first detail for any $N \in \mathbb{N}$ the probability distribution of the conditional SMC (cSMC) algorithm, which corresponds to a process defined on the extended space $\mathsf{W} := (\mathsf{Z}^N \times [N]^N)^{T-1} \times \mathsf{Z}^N \times [N]$ endowed with the corresponding product σ-algebra $\mathcal{B}(\mathsf{W})$, of which P_N is a simple by-product. Our focus is on a particular implementation of the algorithm corresponding to "multinomial resampling" – other schemes are considered in [8]. For

any $x \in \mathsf{X}$ and with $\mathbf{1} \in \{1\}^T$ we define the process $\{Z_t, A_t, t = 1, \ldots, T\}$ on W through

$$\mathbb{P}^N_{\mathbf{1},x}(Z_1 \in dz_1) := \delta_{x_1}(dz_1^1) \prod_{i=2}^N M_1(dz_1^i) \tag{4}$$

and for $t \in \{2, \ldots, T\}$

$$\mathbb{P}^N_{\mathbf{1},x}(Z_t \in dz_t, A_{t-1} = a_{t-1} | Z_{1:t-1} = z_{1:t-1}, A_{1:t-2} = a_{1:t-2})$$

$$= \mathbb{P}^N_{\mathbf{1},x}(Z_t \in dz_t, A_{t-1} = a_{t-1} | Z_{t-1} = z_{t-1}) \tag{5}$$

$$= \delta_{x_t}(dz_t^1) \mathbb{I}\{a_{t-1}^1 = 1\} \prod_{i=2}^N \left(\sum_{k=1}^N \frac{G_{t-1}(z_{t-1}^k)}{\sum_{j=1}^N G_{t-1}(z_{t-1}^j)} \mathbb{I}\{a_{t-1}^i = k\} M_t(z_{t-1}^k, dz_t^i) \right),$$

where we keep k to emphasize that we are sampling from that mixture. For the last iteration, we only require one index and point out that whereas $A_t \in [N]^N$ for $t = 1, \ldots, T-1$, we have $A_T \in [N]$ following

$$\mathbb{P}^N_{\mathbf{1},x}(A_T = k | Z_T = z_T) = \frac{G_T(z_T^k)}{\sum_{j=1}^N G_T(z_T^j)}.$$

The stochastic process defined by $\mathbb{P}^N_{\mathbf{1},x}$ is referred to as the conditional SMC algorithm because it is closely related to a standard SMC algorithm, but where x is a "fixed path" with lineage $\mathbf{1}$. However, as remarked in [1], $\mathbb{P}^N_{\mathbf{1},x}$ is not a conditional distribution of $\mathbb{P}^N(\cdot)$, the standard SMC algorithm whose definition here is deferred to [2], Appendix F. We note further that in order to simplify presentation we have focused here on the scenario where the lineage of x was $\mathbf{1}$ but that we could also use, as in [2], the cSMC with $\mathbf{k} \in [N]^T$ (with associated symbol $\mathbb{P}^N_{\mathbf{k},x}$ and $\mathbb{E}^N_{\mathbf{k},x}$) corresponding to the process above, but where $\delta_{x_t}(dz_t^1)\mathbb{I}\{a_{t-1}^1 = 1\}$ in (5) is replaced with $\delta_{x_t}(dz_t^{k_t})\mathbb{I}\{a_{t-1}^{k_t} = k_{t-1}\}$ and $\delta_{x_1}(dz_1^1)$ with $\delta_{x_1}(dz_1^{k_1})$ in (4).

For any $\mathbf{i} := (i_1, i_2, \ldots, i_T) \in [N]^T$, $z_{1:T} \in (\mathsf{Z}^N)^T$, $a_{1:T} := (a_1, \ldots, a_T) \in ([N]^N)^{T-1} \times [N]$ and $S \in \mathcal{B}(\mathsf{X})$ define

$$I_{\mathbf{i}}(z_{1:T}, a_{1:T}, S) := \mathbb{I}\{z_{1:T}^{\mathbf{i}} \in S, i_T = a_T\} \prod_{t=1}^{T-1} \mathbb{I}\{i_t = a_t^{i_{t+1}}\}. \tag{6}$$

Then the transition kernel of the iterated conditional SMC (i-cSMC), in the multinomial sampling scenario, is given for any $x \in \mathsf{X}$ and $S \in \mathcal{B}(\mathsf{X})$ by

$$P_N(x, S) := \mathbb{E}^N_{\mathbf{1},x}\left[\sum_{\mathbf{i} \in [N]^T} I_{\mathbf{i}}(Z_{1:T}, A_{1:T}, S) \right], \tag{7}$$

that is, conditional upon x we consider the probability distribution of those trajectories $Z_{1:T}^{\mathbf{i}}$ generated by the cSMC which form a lineage compatible with the lineages defined by the random

variables $A_{1:T}$. Our main results concerning the i-cSMC algorithm are the following (our results concerning the particle Gibbs sampler are provided in Section 7). We will denote by π_t the corresponding marginal distribution of π (see (8) for a precise definition).

Theorem 1. *For $N \geq 2$, the i-cSMC algorithm with kernel P_N*

(a) *is reversible with respect to π and defines a positive operator,*

(b) *if for all $t \in \{1, \ldots, T\}$ π_t-ess $\sup_{z_t} G_t(z_t) < \infty$ then there exists $\epsilon_N > 0$ such that*

   (i) *for any $(x, S) \in X \times \mathcal{B}(X)$,*

$$P_N(x, S) \geq \epsilon_N \pi(S),$$

    where $1 - \epsilon_N = O(1/N)$,

  (ii) *for any probability distribution $\nu \ll \pi$ on $(X, \mathcal{B}(X))$ and $k \geq 1$*

$$\left\| \nu P_N^k(\cdot) - \pi(\cdot) \right\|_{L^2(X, \pi)} \leq \|\nu - \pi\|_{L^2(X, \pi)} (1 - \epsilon_N)^k,$$

  (iii) *for any $x \in X$*

$$\left\| \delta_x P_N^k(\cdot) - \pi(\cdot) \right\|_{TV} \leq (1 - \epsilon_N)^k,$$

  (iv) *for any $f \in L^2(X, \pi)$*

$$\operatorname{var}_\pi(f) \leq \operatorname{var}(f, P_N) \leq \left[2\epsilon_N^{-1} - 1 \right] \operatorname{var}_\pi(f),$$

(c) *if π_t-ess $\sup_{z_t} G_t(z_t) = \infty$ for some $t \in [T]$, then, the i-cSMC kernel P_N is not uniformly ergodic for any $N \in \mathbb{N}$,*

(d) *if π_t-ess $\sup_{z_t} G_t(z_t) = \infty$ for some $t \in [T]$ then, the i-cSMC kernel P_N cannot be geometrically ergodic for any $N \in \mathbb{N}$ if π is equivalent to a Lebesgue or counting measure on X.*

Remark 2. From Lemma 22, statement (d) holds under a more abstract assumption, but we have chosen this explicit simplified statement for clarity at this point. In fact we suspect that (d) holds under the assumption π_t-ess $\sup_{z_t} G_t(z_t) = \infty$ for some $t \in [T]$ only, that is essential boundedness is a necessary condition for geometric ergodicity; see Conjecture 24.

With additional conditions on $\{M_t, G_t, t = 1, \ldots\}$ one can characterize ϵ_N in Theorem 1(b) further, and in particular characterize the rate at which N should grow in terms of T in order to maintain a set level of performance. This also requires additional notation and following [9] we define for any $z \in Z$, $p, q \in \mathbb{N}$, $p \leq q$ and $f_q : Z \to \mathbb{R}$,

$$Q_{p,q}(f_q)(z) := \mathbb{E}\left[f_q(Z_q) \prod_{k=p}^{q-1} G_k(Z_k) \,\middle|\, Z_p = z \right],$$

and with the convention $Q_{0,p}(f_p)(x) = M_1 Q_{1,p}(f_p)$ for any $f_p : Z \to \mathbb{R}$, and

$$\eta_p(f_p) := \frac{Q_{0,p}(f_p)}{Q_{0,p}(1)}.$$

and $\bar{M}_{p,p+1}(z,\cdot) = M_{p+1}(z,\cdot)$ and for $q > p \geq 0$ we have the recursive definition, for any $z_p \in \mathsf{Z}$,

$$\bar{M}_{p,q}(z_p,\cdot) = \int M_{p+1}(z_p, \mathrm{d}z_{p+1})\bar{M}_{p+1,q}(z_{p+1},\cdot).$$

The first condition is rather abstract, and can be viewed as a condition on the h-functions investigated in [25] in the context of stability properties of standard SMC algorithms.

(A1). *There exists a constant $\alpha > 0$ such that for any $p, k \in \mathbb{N}$,*

$$\sup_{z \in \mathsf{Z}} \frac{Q_{p,p+k}(1)(z)}{\eta_p Q_{p,p+k}(1)} \leq \alpha.$$

One can however show that (A1) is implied by the following stronger assumption (see Lemma 16).

(A2) (*Strong mixing conditions*). *There exists $m \in \mathbb{Z}_+$ such that*

(a) *There exists a constant $1 \leq \beta < \infty$ such that for any $p \geq 1$ and any $(z, z') \in \mathsf{Z}$ and $S \in \mathcal{B}(\mathsf{Z})$,*

$$\bar{M}_{p,p+m}(z, S) \leq \beta \bar{M}_{p,p+m}(z', S).$$

(b) *The potential functions G_p satisfy, for some $\delta < \infty$,*

$$1 \leq \sup_{z,z' \in \mathsf{Z}^2, p \in \{1,\dots,T\}} \frac{G_p(z)}{G_p(z')} \leq \delta^{1/m}.$$

Theorem 3. *Assume that for all $t \in \mathbb{N}$ π_t-ess $\sup_{z_t} G_t(z_t) < \infty$ and (A1) (or the stronger assumption (A2)) holds. Then with ϵ_N as in Theorem 1(b) for any $N \geq 2$, there exists $C, \varepsilon > 0$ such that with $N = C \times T$, then for any $T \geq 1$, $\epsilon_N \geq \varepsilon > 0$.*

Remark 4. Similar results for the PGibbs sampler are provided in Section 7.

Proof of Theorem 1. The proofs of the various results are the subject of the following sections. More specifically, statement

(a) follows from Lemma 8 (the latter property was established in [8] and the former noted/proved in [1,8]),

(b) all parts follow from Corollary 12 and [2], Proposition 32, which gathers generic results on π-invariant Markov chains satisfying (b)(i),

(c) follows from Proposition 17,

(d) follows from Proposition 20 and Lemma 22; Remark 23. □

Proof of Theorem 3. Follows from Proposition 13, Corollary 14 and Lemma 16. □

As pointed out in the Introduction, soon after completing this work we have become aware of [16], where a subset of our results have also been independently discovered. This motivates the following comparison. Result (b)(i) of Theorem 1 is identical to Theorem 1 of [16], but relies on a different proof. Results (b)(ii)–(iv) rely on standard arguments, although (iv) does not seem to be well known and establishes informative quantitative bounds. The study of the necessity of our conditions to imply uniform or geometric ergodicity is not addressed in [16]. The result of Theorem 3 corresponds to Proposition 5 of [16]. The conditions under which Theorem 3 holds are rather stringent for some applications, in particular in the state–space model scenario. As discussed by [16] in that scenario (A2) will essentially only hold in the case where X is compact. The condition (A1) is weaker and more natural in our analysis, but is not currently easy to verify in applications except through (A2).

In an attempt to relax (A2), the authors of [16] investigate another set of specialised assumptions guaranteeing that the result of Theorem 3 holds even in some non-compact scenarios provided the number of particles N grows at a rate $T^{1/\gamma}$ for any $\gamma \in (0, 1)$, a result in line with what is obtained with the stronger assumption (A2), for which $\gamma = 1$ is permissible. This requires the specification of a "moment assumption" which aims at controlling the variations of the various quantities involved under the law of the observation process $\{Y_t, t \geq 0\}$. Their approach, however, does not seem to allow one to consider the scaling properties of the PGibbs sampler (i.e., not just the i-cSMC); see their Theorem 6 and Remark 7. More importantly, we note that their results require the law of the data to coincide with that of the specified model for some $\theta^\star \in \Theta$ which, although suggestive of what may happen in practice, is always an idealization. This delicate work is the main focus of the remainder of their investigation while here, in addition to establishing the necessity of some of the conditions, we have focused on the transference of the results obtained for the i-cSMC to the PGibbs sampler (Section 7) with the aim of showing that the PGibbs has performance inferior to that of the Gibbs sampler, but arbitrarily close if we increase N.

4. Establishing the uniform minorization condition

Before proceeding, we turn to the i-SIR which is particularly simple to analyze. The reason for detailing the short analysis of this simple scenario is to provide the reader with an overview of the developments which are to follow – the remainder of the paper essentially replicates the key steps of the argument below, albeit in the more complex SMC framework. Notice that in this scenario $X = Z$ since $T = 1$. We let $G(x) := \pi(\mathrm{d}x)/M(\mathrm{d}x)$ for any $x \in X$ and assume that $\bar{G} := \sup_{x \in X} G(x) < \infty$. Then for $(x, S) \in X \times \mathcal{B}(X)$ we can rewrite

$$P_N(x, S) = \sum_{k=1}^{N} \int_{X^N} \frac{\pi(\mathrm{d}z^k)/M(\mathrm{d}z^k)}{\sum_{j=1}^{N} G(z^j)} \mathbb{I}\{z^k \in S\} \left(\delta_x(\mathrm{d}z^1) \prod_{i=2}^{N} M(\mathrm{d}z^i) \right)$$

$$= \int_{X^N} \frac{1}{\sum_{j=1}^{N} G(z^j)} \frac{\pi(\mathrm{d}z^1)}{M(\mathrm{d}z^1)} \mathbb{I}\{z^1 \in S\} \left(\delta_x(\mathrm{d}z^1) \prod_{i=2}^{N} M(\mathrm{d}z^i) \right)$$

$$+ \sum_{k=2}^{N} \int_{X^N} \frac{1}{\sum_{j=1}^{N} G(z^j)} \mathbb{I}\{z^k \in S\} \pi\left(dz^k\right) \left(\delta_x\left(dz^1\right) \prod_{i=2, i \neq k}^{N} M\left(dz^i\right)\right)$$

$$= \sum_{k=1}^{N} \int_{X} \mathbb{E}_{1,x,k,y} \left[\frac{\mathbb{I}\{y \in S\}}{\sum_{j=1}^{N} G(Z^j)}\right] \left(\mathbb{I}\{k = 1\} \frac{\pi(dx)}{M(dx)} \delta_x(dy) + \mathbb{I}\{k \neq 1\} \pi(dy)\right),$$

where $\mathbb{E}_{1,x,k,y}(\cdot)$ defines an expectation for the random variables $Z^1, \ldots, Z^N$ associated to the probability distribution

$$\delta_x\left(dz^1\right) \prod_{i=2}^{N} M\left(dz^i\right)$$

for $k = 1$ and $x = y \in X$, and

$$\delta_{x,y}\left(dz^1 \times dz^k\right) \prod_{i=2, i \neq k}^{N} M\left(dz^i\right)$$

for $k \neq 1$ and $x, y \in X$. This auxiliary process turns out to be central to our analysis, and will be generalised to the general scenario and called "doubly" cSMC (c^2SMC). Indeed, omitting the term $k = 1$ in the representation of P_N and by application of Jensen's inequality to the convex mapping $x \mapsto (x + a)^{-1}$ for $x, a \in \mathbb{R}_+$ we obtain

$$P_N(x, S) \geq \sum_{k=2}^{N} \int_{X} \mathbb{I}\{y \in S\} \mathbb{E}_{1,x,k,y} \left[\frac{1}{G(x) + G(y) + \sum_{j=2, j \neq k}^{N} G(Z^j)}\right] \pi(dy)$$

$$\geq \sum_{k=2}^{N} \int_{X} \frac{\mathbb{I}\{y \in S\}}{G(x) + G(y) + N - 2} \pi(dy)$$

$$\geq \frac{N - 1}{2\bar{G} + N - 2} \pi(S).$$

This is a uniform minorization condition which immediately implies uniform geometric convergence (see the outline of our results in Section 1), but in the present situation the result is even stronger in that, in particular, it provides us with quantitative bounds on the dependence of the performance of the algorithm on N. Indeed it is a standard result that the minorization constant

$$\epsilon_N = \frac{N - 1}{2\bar{G} + N - 2} = 1 - \frac{2\bar{G} - 1}{2\bar{G} + N - 2},$$

provides the upper bound $1 - \epsilon_N$ on the (geometric) rate of convergence of the algorithm, which here vanishes at an asymptotic rate N^{-1} as N increases. As we shall see the fact that the minorization measure is the invariant distribution leads to a direct lower bound on associated Dirichlet forms associated to P_N which in turn provide quantitative bounds on the spectral gap and the

associated asymptotic variance. In the remainder of the section, we generalize the representation of P_N in terms of the c²SMC algorithm and "the estimator of the normalizing constant" which suggests applying Jensen's inequality as above. This requires us to consider estimates of the resulting expectation in Section 5.

In order to proceed further it is required to define the c²SMC process, which is essentially similar to the cSMC process but where conditioning is now upon two trajectories $x, y \in \mathsf{X}$. The definition is therefore similar, but for reasons which will become clearer below the second fixed trajectory is set to have a lineage of the general form $\mathbf{k} := k_{1:T} \in [N]^T$. We will use below the convention that $\delta_{a,b}(dz^1 \times dz^k)$ reduces to $\delta_a(dz^1)$ whenever $k = 1$. The definition of this process is similar to that of the cSMC algorithm and the distributions involved are defined for $x, y \in \mathsf{X}$ and $\mathbf{k} \in [N]^T$ as follows

$$\mathbb{P}^N_{\mathbf{1},x,\mathbf{k},y}(Z_1 \in dz_1) = \delta_{x_1,y_1}(dz_1^1 \times dz_1^{k_1}) \prod_{i=2,i\neq k_1}^{N} M_1(dz_1^i),$$

and for $t = 2, \ldots, T - 1$ (with the convention $a_{t-1}^{k,l} := (a_{t-1}^k, a_{t-1}^l)$)

$$\mathbb{P}^N_{\mathbf{1},x,\mathbf{k},y}(Z_t \in dz_t, A_{t-1} = a_{t-1}|Z_{t-1} = z_{t-1})$$

$$= \delta_{x_t,y_t}(dz_t^1 \times dz_t^{k_t})$$

$$\times \mathbb{I}\{a_{t-1}^{1,k_t} = (1, k_{t-1})\} \prod_{i=2,i\neq k_t}^{N} \left(\sum_{l=1}^{N} \frac{G_{t-1}(z_{t-1}^l)}{\sum_{j=1}^{N} G_{t-1}(z_{t-1}^j)} \mathbb{I}\{a_{t-1}^i = l\} M_t(z_{t-1}^l, dz_t^i) \right)$$

and

$$\mathbb{P}^N_{\mathbf{1},x,\mathbf{k},y}(A_T = l|Z_T = z_T) = \frac{G_T(z_T^l)}{\sum_{j=1}^{N} G_T(z_T^j)}.$$

We note that although the transitions and the initial distributions are, by the convention, well defined for $k_t = 1$ and $x_t \neq y_t$ the distribution above will never be used in such a context. Just as $\mathbb{P}^N_{\mathbf{1},x}$ is not a conditional distribution of $\mathbb{P}^N(\cdot)$, the law of the SMC algorithm, the same holds between $\mathbb{P}^N_{\mathbf{1},x,\mathbf{k},y}(\cdot)$ and $\mathbb{P}^N_{\mathbf{1},x}(\cdot)$. However we now provide an important property relating these two probability distributions, which together with (7) will allow us to decompose this transition into key quantities and establish the sought minorization condition. The proof of the following Lemma is in [2], Appendix A.

Lemma 5. *For* $\mathbf{i} \in \{2, \ldots, N\}^T$ *and* $x \in \mathsf{X}$,

$$\mathbb{E}^N_{\mathbf{1},x}[I_{\mathbf{i}}(Z_{1:T}, A_{1:T}, S)] = \frac{\gamma_T}{N^T} \int_{\mathsf{X}} \pi(dy) \times \mathbb{I}\{y \in S\} \times \mathbb{E}^N_{\mathbf{1},x,\mathbf{i},y}\left[\frac{1}{\prod_{t=1}^{T}(1/N)\sum_{j=1}^{N} G_t(Z_t^j)} \right].$$

As we shall see, the concentration properties of the "estimator of the normalizing constant" plays a central role for any $z_{1:T} \in (\mathsf{Z}^N)^T$

$$\hat{\gamma}_T^N(z_{1:T}) := \prod_{t=1}^T \frac{1}{N} \sum_{j=1}^N G_t(z_t^j).$$

We first obtain a uniform minorization condition for the cSMC transition probability. This simple result establishes the expectation of $\hat{\gamma}_T^N(Z_{1:T})$ with respect to a c^2SMC algorithm as a key quantity of interest, and motivates the non-asymptotic analysis and bounds of Section 5.

Proposition 6. *For any $(x, S) \in \mathsf{X} \times \mathcal{B}(\mathsf{X})$ and $N \geq 2$, we have*

$$P_N(x, S) \geq \int_S \frac{\gamma_T \times (1 - 1/N)^T}{\mathbb{E}_{\mathbf{1}, x, \mathbf{2}, y}^N[\hat{\gamma}_T^N(Z_{1:T})]} \pi(\mathrm{d}y).$$

Proof. Using (7), we only keep the trajectories for which there is no coalescence with the first trajectory, that is, we exclude terms such that $i_t = 1$ for some $t \in [T]$ and obtain

$$P_N(x, S) \geq \sum_{\mathbf{i} \in \{2, \ldots, N\}^T} \mathbb{E}_{\mathbf{1}, x}^N[I_{\mathbf{i}}(Z_{1:T}, A_{1:T}, S)].$$

Consequently, using Lemma 5,

$$P_N(x, S) \geq \sum_{i_{1:T} \in [2:N]^T} \frac{\gamma_T}{N^T} \int_{\mathsf{X}} \pi(\mathrm{d}y) \times \mathbb{I}\{y \in S\} \times \mathbb{E}_{\mathbf{1}, x, \mathbf{i}, y}^N \left[\frac{1}{\prod_{t=1}^T (1/N) \sum_{j=1}^N G_t(Z_t^j)} \right]$$

$$= \frac{\gamma_T (N - 1)^T}{N^T} \int_S \mathbb{E}_{\mathbf{1}, x, \mathbf{2}, y}^N \left[\frac{1}{\prod_{t=1}^T (1/N) \sum_{j=1}^N G_t(Z_t^j)} \right] \pi(\mathrm{d}y_{1:T}),$$

using invariance by permutation of $i_1, \ldots, i_T$ of the expectations. We conclude by application of Jensen's inequality for the convex function $u \mapsto 1/u$ for $u \in \mathbb{R}_+$. $\qquad \square$

Corollary 7. *Let $N \geq 2$ and assume that*

$$\epsilon_N := \frac{\gamma_T \times (1 - 1/N)^T}{\sup_{x, y \in \mathsf{X}} \mathbb{E}_{\mathbf{1}, x, \mathbf{2}, y}^N[\hat{\gamma}_T^N(Z_{1:T})]} > 0,$$

then for any $(x, S) \in \mathsf{X} \times \mathcal{B}(\mathsf{X})$, $P_N(x, S) \geq \epsilon_N \pi(S)$ and from Proposition 6 all the properties of [2], Proposition 32, apply to the i-cSMC with $\varepsilon = \epsilon_N$.

The next section is dedicated to finding a useful expression for the expectation $\mathbb{E}_{\mathbf{1}, x, \mathbf{2}, y}^N[\hat{\gamma}_T^N(Z_{1:T})]$ and establishing explicit bounds on this quantity, and therefore ϵ_N in Corollary 7, under additional assumptions.

Before proceeding to novel analysis, for completeness we gather two known properties of the i-cSMC (in the general set-up) in the following lemma which will be exploited throughout the remainder of the paper. Both results are immediate upon noticing that the i-cSMC is a two stage Gibbs sampler on an artificial joint distribution (see (B.1) in [2], Appendix B, which is a generalization of (1)). The results have also been shown in detail in [8]. A proof is included in [2], Appendix B, for completeness.

Lemma 8. P_N, *viewed as an operator on $L^2(\mathsf{X}, \pi)$, is self-adjoint and positive.*

5. Quantitative bounds for the doubly conditional i-cSMC expectation

In this section, we first find an exact expression for $\mathbb{E}^N_{\mathbf{1},x,\mathbf{2},y}[\hat{\gamma}^N_T(Z_{1:T})]$ in terms of quantities underpinning the definition of π given in Section 3 and then move on to provide various estimates of the conditional expectation involved in the minorization established in Proposition 6, under various assumptions on the aforementioned quantities. Throughout, we use the usual convention that $\sum_\varnothing = 0$ and $\prod_\varnothing = 1$. We let $G_{p,q}(z) := Q_{p,q}(1)(z)$ and $G^{1+2}_{p,q} := G_{p,q}(x_p) + G_{p,q}(y_p)$. We note that $G_{p,p+1}(z) = G_p(z)$ for $p \in [T]$ and we use the convention throughout that for any $z \in \mathsf{Z}$, $G_0(z) = 1$ and $Q_{0,p}(f_p)(z) := M_1(Q_{1,p}(f_q))$. We write $G_{0,p} := M_1(Q_{1,p}(1))$ since $G_{0,p}(z)$ is independent of z. Our first result, whose proof can be found in [2], Appendix D, is:

Proposition 9. *Let $x, y \in \mathsf{X}$ and $N \geq 2$. Then,*

$$\mathbb{E}^N_{\mathbf{1},x,\mathbf{2},y}\big[\hat{\gamma}^N_T(Z_{1:T})\big] = \frac{1}{N^T} \sum_{s=1}^{T+1} (N-2)^{T+1-s} \sum_{\mathbf{i} \in \mathcal{I}_{T+1,s}} G_{0,i_1} C_{T,s}(\mathbf{i}, x, y),$$

where for any $s = 1, \ldots, k$,

$$\mathcal{I}_{k,s} := \big\{i_1, \ldots, i_s \in \mathbb{N}^s : T - k + 1 < i_1 < \cdots < i_s = T + 1\big\},$$

and for $\mathbf{i} \in \mathcal{I}_{k,s}$

$$C_{k,s}(\mathbf{i}, x, y) := \prod_{m=1}^{s-1} \big[G_{i_m,i_{m+1}}(x_{i_m}) + G_{i_m,i_{m+1}}(y_{i_m})\big].$$

Remark 10. While the expectation of interest here has been hitherto uninvestigated, the form of Proposition 9 is reminiscent of non-asymptotic results in [7], in which second moments of $\hat{\gamma}^N_T(Z_{1:T})$ are analyzed with respect to the law of a standard SMC algorithm.

We now turn to estimates of the expectation above, starting with very minimal assumptions which allow us to establish the minorization condition required to apply [2], Proposition 32, and deduce most of our results, without the need for assumptions on the dynamic of the system – the number of particles is however required to grow exponentially in order to maintain a set

level of performance. We show subsequently that with stronger assumptions on $\{M_t, G_t\}_{t=1}^T$ it is possible to show that N should grow linearly with T to ensure that a set level of performance is maintained.

Proposition 11. *Assume that for all $t \in \{1, \ldots, T\}$, $\bar{G}_t := \sup_{z \in Z} G_t(z) < \infty$, then for any $N \geq 2$*

$$\mathbb{E}_{\mathbf{1},x,\mathbf{2},y}^N[\hat{\gamma}_T^N(Z_{1:T})] \leq \gamma_T \left\{ 1 + \left[1 - \left(1 - \frac{2}{N} \right)^T \right] \left[\frac{\prod_{t=1}^T \bar{G}_t}{\gamma_T} - 1 \right] \right\}.$$

Proof. The assumption on the potentials implies that for any $p, q \in \mathbb{N}$ with $p < q$ we have $G_{p,q} \leq \prod_{k=p}^{q-1} \bar{G}_k$, and from Proposition 9 we have

$$\mathbb{E}_{\mathbf{1},x,\mathbf{2},y}^N[\hat{\gamma}_T^N(Z_{1:T})] = \sum_{s=1}^{T+1} \left(\frac{N-2}{N} \right)^{T+1-s} \frac{2^{s-1}}{N^{s-1}} \sum_{\mathcal{I}_{T+1,s}} G_{0,i_1} \prod_{m=1}^{s-1} \frac{1}{2} G_{i_m,i_{m+1}}^{1+2}$$

$$\leq \gamma_T \left(\frac{N-2}{N} \right)^T + \prod_{k=1}^T \bar{G}_k \times \sum_{s=2}^{T+1} \binom{T}{s-1} \left(\frac{N-2}{N} \right)^{T+1-s} \frac{2^{s-1}}{N^{s-1}}$$

$$= \gamma_T \left(\frac{N-2}{N} \right)^T + \left[1 - \left(\frac{N-2}{N} \right)^T \right] \prod_{k=1}^T \bar{G}_k,$$

and the result follows. □

Corollary 12. *Propositions 6 and 11 together imply that for any $x, S \in \mathsf{X} \times \mathcal{B}(\mathsf{X})$,*

$$P_N(x, S) \geq \epsilon_N \pi(S) \qquad with \ \epsilon_N = \frac{(1 - 1/N)^T}{1 + [1 - (1 - 2/N)^T][(\prod_{t=1}^T \bar{G}_t)/\gamma_T - 1]}$$

and $\lim_{N \to \infty} \epsilon_N = 1$.

It should be clear that despite Corollary 12, the term $\prod_{t=1}^T \bar{G}_t / \gamma_T$ typically grows exponentially fast with T whenever the potentials are not constant functions. Therefore, Proposition 11 suggests that the number of particles N should grow exponentially with T in general. However, stronger assumptions on the system under consideration will allow us to maintain a given lower bound on ϵ_N by increasing N only linearly with T. We first state our main result using the abstract condition (A1) and then show that classical strong mixing conditions (A2) imply (A1).

Proposition 13. *Assume (A1), then for any $N \geq 2$*

$$\mathbb{E}_{\mathbf{1},x,\mathbf{2},y}^N[\hat{\gamma}_T^N(Z_{1:T})] \leq \gamma_T \left(1 + \frac{2(\alpha - 1)}{N} \right)^T.$$

Proof. First, notice that for any $1 \leq k \leq n$

$$Q_{0,n}(1) = Q_{0,k}(1) \frac{Q_{0,n}(1)}{Q_{0,k}(1)} = Q_{0,k}(1) \eta_k Q_{k,n}(1),$$

and therefore for any $s \in \{1, \ldots, T\}$ and $0 < i_1 < \cdots < i_{s-1} < i_s = T + 1$ with the notation defined earlier,

$$Q_{0,T}(1) = Q_{0,i_1}(1) \prod_{k=1}^{s-1} \eta_{i_k} Q_{i_k, i_{k+1}}(1) = G_{0,i_1} \prod_{k=1}^{s-1} \eta_{i_k} G_{i_k, i_{k+1}},$$

and from (A1), with $\bar{G}_{p,q} := \sup_{z \in \mathsf{Z}} G_{p,q}(z)$, and applying Proposition 9 yields the following upper bound for $\mathbb{E}_{\mathbf{1},x,\mathbf{2},y}^N[\hat{\gamma}_T^N(Z_{1:T})]$:

$$\sum_{s=1}^{T+1} \left(\frac{N-2}{N} \right)^{T+1-s} \frac{2^{s-1}}{N^{s-1}} \sum_{0 < i_1 < \cdots < i_{s-1} < i_s = T+1} G_{0,i_1} \prod_{m=1}^{s-1} \frac{1}{2} G_{i_m, i_{m+1}}^{1+2}$$

$$\leq \gamma_T \left(\frac{N-2}{N} \right)^T + \gamma_T \sum_{s=2}^{T+1} \left(\frac{N-2}{N} \right)^{T+1-s} \frac{2^{s-1}}{N^{s-1}} \sum_{\mathcal{I}_{T+1,s}} \frac{G_{0,i_1}}{G_{0,i_1}} \prod_{m=1}^{s-1} \frac{\bar{G}_{i_m, i_{m+1}}}{\eta_{i_k} G_{i_k, i_{k+1}}}$$

$$\leq \gamma_T \sum_{s=1}^{T+1} \binom{T}{s-1} \left(\frac{N-2}{N} \right)^{T+1-s} \frac{2^{s-1}}{N^{s-1}} \alpha^{s-1},$$

and we conclude by an application of the binomial theorem. $\qquad\square$

Corollary 14. *Propositions 6 and 11 together imply that for any $(x, S) \in \mathsf{X} \times \mathcal{B}(\mathsf{X})$,*

$$P_N(x, S) \geq \epsilon_N \pi(S) \qquad \text{with } \epsilon_N = \left(\frac{1 - 1/N}{1 + 2(\alpha - 1)/N} \right)^T.$$

Now, let $N - 1 \geq CT$ for some $C > 0$. Then $\epsilon_N \geq \exp(-\frac{2\alpha - 1}{C})$.

Proof. Propositions 6 and 13 together imply that

$$\epsilon_N \geq \left(1 + \frac{2\alpha - 1}{N - 1} \right)^{-T}.$$

Since $(N - 1) \geq CT$ for some $C > 0$, and $\log(1 + x) \leq x$ for all $x \geq 0$,

$$\left(1 + \frac{2\alpha - 1}{N - 1} \right)^T \leq \left(1 + \frac{2\alpha - 1}{CT} \right)^T \leq \exp\left(\frac{2\alpha - 1}{C} \right). \qquad\square$$

Remark 15. The combination of the upper bound of $\text{var}(f, P_N)$ in Theorem 1 with Corollary 14 suggests a rough rule of thumb to select N for the i-cSMC Markov kernel. In particular, there is generally a tradeoff between iterating a less computationally intensive Markov kernel more times and iterating a more computationally intensive expensive fewer times. This suggests that one should minimize the function $f(N) := N \text{var}(f, P_N)$. While an analytic expression for $\text{var}(f, P_N)$ is not available, we can minimize its upper bound

$$(CT + 1)\left\{2 \exp\left(\frac{2\alpha - 1}{C}\right) - 1\right\},$$

with respect to C. Assuming that we are in the scenario where $N \gg 1$ and therefore $CT + 1 \approx CT$, one then finds the unique minimum

$$C^* = \frac{2\alpha - 1}{\text{Lambert}_W(-1/(2\exp(1))) + 1} \approx 1.302(2\alpha - 1),$$

(where Lambert_W is the principal branch of the Lambert W function) or correspondingly

$$\epsilon_N^* \approx 0.464.$$

Hence, under (A1) it is only required for N to scale linearly with T in order to maintain a non-vanishing ergodicity rate. Following, for example, [7,9] we make the following assumptions on $\{M_t\}$ and the potentials $\{G_t\}$ which combined define an m-step "strong mixing" condition which automatically implies (A1). The following result relies on classical arguments [7], Lemma 4.3, [9],

Lemma 16. *Assume* (A2). *Then for any $k \in \mathbb{Z}_+$ we have*

$$\sup_{z,z' \in \mathsf{Z}^2} \frac{Q_{p,p+k}(1)(z)}{Q_{p,p+k}(1)(z')} \leq \beta\delta,$$

i.e., (A1) *is satisfied.*

6. Necessity of the boundedness assumption and a conjecture

Proposition 11 showed that the i-cSMC kernel is uniformly ergodic if the potentials are bounded. We study here the opposite case, where at least one of the potentials is unbounded. We discover that then the algorithm cannot be uniformly ergodic (Proposition 17), and in many cases the algorithm cannot be geometrically ergodic (Proposition 20 and Lemma 22; Remark 23). We believe that the latter holds in general (Conjecture 24), but a proof has remained elusive. This dichotomy of algorithms which are uniformly ergodic and sub-geometrically ergodic would be in perfect analogy with the behaviour of the independent Metropolis–Hastings [17], Theorem 2.1.

We will denote hereafter the marginal densities of π by

$$\pi_{t:u}(A) := \pi\left(\mathsf{Z}^{t-1} \times A \times \mathsf{Z}^{T-u}\right) \qquad \text{for } A \in \mathcal{B}\left(\mathsf{Z}^{u-t+1}\right), \tag{8}$$

where $1 \le t \le u \le T$ and we use the shorthand $\pi_t(A) := \pi_{t:t}(A)$.

In this section, we will assume that $\mathsf{S} \in \mathcal{B}(\mathsf{Z})^T$ is a fixed set such that for all $x \in \mathsf{S}$, $\prod_{t=1}^{T} G_t(x_t) > 0$ and $\pi(\mathsf{S}) = 1$. Further, S contains all possible starting points of the algorithm, that is, we assume that the state space of the i-cSMC is S. In the discrete case, the minimal S consists of the points of positive π-measure, and in the continuous case where π admits a density, the set S can be taken as the set where the density is positive.

Further, we will assume that π_1 is not concentrated on a single point. We can do this without loss of generality, because if $\pi_1, \ldots, \pi_t$ were concentrated on single points of the state space, the algorithm would be deterministic until π_{t+1} and we could consider the i-cSMC for $\pi' = \pi_{t+1:T}$.

Proposition 17. *Suppose π_t-ess $\sup_{x_t} G_t(x_t) = \infty$ for some $t \in [T]$. Then, the i-cSMC kernel P_N is not uniformly ergodic for any $N \in \mathbb{N}$.*

Proof. If the i-cSMC kernel is uniformly ergodic, then there exist $K < \infty$ and $\rho \in (0, 1)$ such that

$$\sup_{x \in \mathsf{S}} \left\| P_N^n(x, \cdot) - \pi(\cdot) \right\|_{\mathrm{TV}} \le K\rho^n \qquad \text{for all } n \in \mathbb{N}.$$

Fix $\epsilon' > 0$ and let $n \in \mathbb{N}$ be such that $K\rho^n \le \epsilon'$. We will prove that there exists a set $B_{\epsilon'} \in \mathcal{B}(\mathsf{Z})$ such that $\pi_1(B_{\epsilon'}) > 0$ and $\inf_{x \in B_{\epsilon'}} P_N^n(x, \{x_1\} \times \mathsf{Z}^{T-1}) \ge 1 - \epsilon'$. For all $x \in B_{\epsilon'}$, we have $|P_N^n(x, \{x_1\} \times \mathsf{Z}^{T-1}) - \pi_1(\{x_1\})| \le K\rho^n \le \epsilon'$. This, with $\epsilon' > 0$ small enough, will contradict $\pi_1(\{x_1\}) < 1$.

Lemma 18 shows that there exists $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\lim_{g \to \infty} \phi(g) = 0$, and

$$P_N\left(x, \{x_1\}^{\complement} \times \mathsf{Z}^{T-1}\right) \le \phi\left(G(x_t)\right).$$

Denote the level set $L_t(\underline{G}) := \{x_t \in \mathsf{Z} : G_t(x_t) \le \underline{G}\}$. Lemma 18 shows that there exists $c_2 = c_2(N) \in [1, \infty)$ such that for $G_t(x_t) \ge \underline{G}$

$$P_N\left(x, \mathsf{Z}^{t-1} \times L_t(\underline{G}) \times \mathsf{Z}^{T-t}\right) \le c_2 \underline{G}/G_t(x_t).$$

Let $\epsilon \in (0, 1)$ and define $\delta := \epsilon/c_2$ and let G_* be large enough so that $\phi(\delta^n G_*) \le \epsilon$. Define the (sub-probability) kernels $\mu_{\bar{G}}(x, \mathrm{d}y) := P_N(x, \mathrm{d}y)\delta_{x_1}(y_1)\mathbb{I}\{G_t(y_t) \ge \bar{G}\}$ on $(\mathsf{S}, \mathcal{B}(\mathsf{S}))$ for any $\bar{G} > 0$ and observe that we may estimate

$$\mathbb{I}\{G_t(x_t) \ge G_*\} P_N^n\left(x, \{x_1\} \times \mathsf{Z}^{T-1}\right)$$

$$\ge \mathbb{I}\{G_t(x_t) \ge G_*\} \int \mu_{\delta G_*}\left(x, \mathrm{d}y^{(2)}\right) \int \mu_{\delta^2 G_*}\left(y^{(2)}, \mathrm{d}y^{(3)}\right) \cdots \int \mu_{\delta^{n-1} G_*}\left(y^{(n-1)}, \mathrm{d}y^{(n)}\right).$$

We may estimate for any $i \in [n]$ and all $x \in \mathsf{S}$ such that $G_t(x_t) \ge \delta^{i-1} G_*$,

$$\int \mu_{\delta^i G_*}(x, \mathrm{d}y) \ge 1 - P_N\left(x, \{x_1\}^{\complement} \times \mathsf{Z}^{T-1}\right) - P_N\left(x, \mathsf{Z}^{t-1} \times L_t\left(\delta^i G_*\right) \times \mathsf{Z}^{T-t}\right)$$

$$\ge 1 - 2\epsilon.$$

We conclude that for $x \in \mathsf{S}$ such that $G_t(x_t) \geq G_*$,

$$P_N^n\big(x, \{x_1\} \times \mathsf{Z}^{T-1}\big) \geq (1 - 2\epsilon)^n.$$

This proves the claim, as $\epsilon > 0$ was arbitrary. $\qquad\qquad\square$

Lemma 18. *For all $x \in \mathsf{S}$ and all $\underline{G} \in \mathbb{R}_+$,*

1. $P_N(x, \{x_1\}^{\complement} \times \mathsf{Z}^{T-1}) \leq \phi(G(x_t))$,
2. $P_N(x, \mathsf{Z}^{t-1} \times L_t(\underline{G}) \times \mathsf{Z}^{T-t}) \leq (N-1)^2 \underline{G}/G_t(x_t)$ *whenever* $G_t(x_t) \geq \underline{G}$,

where $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ is a function such that $\lim_{g \to \infty} \phi(g) = 0$.

Proof. In both cases, we consider the case $t < T$; the special case $t = T$ can be treated similarly. In order to facilitate the theoretical analysis, we introduce a non-standard implementation of the cSMC which relies on the remark that at any time instant a given particle can only have a maximum number N of children. Hence, when implementing the cSMC it is always possible to draw N children first and then decide who is carried forward according to the standard selection mechanism. It is in fact possible to push this idea further and, given a fixed $x \in \mathsf{S}$, to sample the following N-ary tree of random variables first

$$\hat{Z}_1^1 = x_1, \qquad \hat{Z}_1^i \sim M_1(\cdot), \qquad i \in [N] \setminus \{1\}$$
$$\hat{Z}_2^{1,1} = x_2, \qquad \hat{Z}_2^{i,j} \sim M_2(\hat{Z}_1^i, \cdot), \qquad (i, j) \in [N]^2 \setminus \{(1, 1)\}$$
$$\vdots$$
$$\hat{Z}_T^{\mathbf{1}} = x_T, \qquad \hat{Z}_T^{i_1, \dots, i_T} \sim M_T\big(\hat{Z}_{T-1}^{i_1, \dots, i_{T-1}}, \cdot\big), \qquad (i_1, \dots, i_T) \in [N]^T \setminus \{\mathbf{1}\},$$

and then prune the tree using the selection mechanism of the cSMC algorithm with fixed path $x \in \mathsf{S}$. As a result, each Z_t^j in the cSMC is associated with some $\hat{Z}_t^i$. The construction above permits the bound

$$U := \sum_{\mathbf{i} \in [N]^t} G_t\big(Z_t^{i_t}\big) \mathbb{I}\{i_1 \neq 1\} \prod_{p=2}^{t} \mathbb{I}\{i_{p-1} = A_{p-1}^{i_p}\} \leq \sum_{\mathbf{i} \in \{2, \dots, N-1\}^t} G_t\big(\hat{Z}_t^{\mathbf{i}}\big) =: V,$$

where U corresponds to the sum of potentials associated with those Z_t^j whose ancestral lineage does not contain the value 1. It therefore follows that

$$P_N\big(x, \{x_1\}^{\complement} \times \mathsf{Z}^{T-1}\big) = \mathbb{E}_{\mathbf{1},x}^N\left[\frac{U}{G_t(x_t) + \sum_{j=2}^{N} G_t(Z_t^j)}\right] \leq \mathbb{E}_{\mathbf{1},x}^N\left[\frac{U}{G_t(x_t) + U}\right]$$
$$\leq \mathbb{E}_{\mathbf{1},x}^N\left[\frac{V}{G_t(x_t) + V}\right],$$

because $u \mapsto u/(g + u)$ is increasing. Now, V is a finite non-negative random variable independent of x. We may define

$$\phi(g) := \mathbb{E}^N_{\mathbf{1},x}\left[\frac{V}{g + V}\right],$$

which satisfies $\lim_{g \to \infty} \phi(g) = 0$ by the monotone convergence theorem.

For the second inequality, we can show similarly that for $G_t(x_t) \geq \underline{G}$

$$\mathbb{P}^N_{\mathbf{1},x}\left[G_t\left(Z_t^{A_t^i}\right) \leq \underline{G}\right] = \mathbb{E}^N_{\mathbf{1},x}\left[\frac{\sum_{k=2}^N G_t(Z_t^k)\mathbb{I}\{G_t(Z_t^k) \leq \underline{G}\}}{G_t(x_t) + \sum_{k=2}^N G_t(Z_t^k)}\right] \leq \frac{(N-1)\underline{G}}{G_t(x_t)}$$

and so

$$P_N\left(x, \mathsf{Z}^{t-1} \times L_t(\underline{G}) \times \mathsf{Z}^{T-t}\right) \leq \sum_{i=2}^N \mathbb{P}^N_{\mathbf{1},x}\left[G_t\left(Z_t^{A_t^i}\right) \leq \underline{G}\right] = (N-1)\mathbb{P}^N_{\mathbf{1},x}\left[G_t\left(Z_t^{A_t^i}\right) \leq \underline{G}\right]. \qquad \square$$

To establish that P_N cannot be even geometrically ergodic whenever $\pi_t\text{-ess sup}_{x_t} G_t(x_t) = \infty$ for some $t \in [T]$ in many settings, we use Proposition 19. This allows for the developments of Proposition 20 and Lemma 22, leading to the desired result under assumptions satisfied in many applications; see Remark 23.

Proposition 19. *Suppose P is an ergodic Markov kernel on a state space $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ with invariant distribution π. Suppose that for any $\epsilon, \delta > 0$ there exists a set $A \in \mathcal{B}(\mathsf{X})$ such that $\pi(A) \in (0, \delta)$ and $\inf_{x \in A} P(x, A) \geq 1 - \epsilon$. Then P is not geometrically ergodic.*

Proof. The result follows directly by following the proof of [23], Theorem 3.1, or by a conductance argument [14], Theorem 1. $\qquad \square$

Proposition 20. *Assume that for at least one $t \in [T]$*

$$\pi\text{-ess sup}_x \mathbb{E}^N_{\mathbf{1},x}\left[\frac{G_t(x_t)}{\sum_{k=1}^N G_t(Z_t^k)}\right] = 1. \qquad (9)$$

Then P_N cannot be geometrically ergodic.

Proof. Because of Proposition 19 it suffices to establish that

$$\pi_{1:t}\text{-ess sup}_{x_{1:t} \in \mathsf{S}}\left\{\inf_{x_{t+1:T}} P_N\left(x_{1:T}; \{x_{1:t}\} \times \mathsf{Z}^{T-t}\right)\right\} = 1. \qquad (10)$$

We note that

$$\inf_{x_{t+1:T}} P_N\left(x_{1:T}, \{x_{1:t}\} \times \mathsf{Z}^{T-t}\right) \geq \mathbb{P}^N_{\mathbf{1},x}\left(A_t^{1:N} = 1\right)$$

$$\geq 1 - \sum_{i=2}^N \mathbb{P}^N_{\mathbf{1},x}\left(A_t^i \neq 1\right),$$

because $A_t^1 = 1$ by construction. We emphasize that A_t^i are independent of $x_{t+1:T}$. Now (10) follows directly from (9) because for $i \in \{2, \ldots, N\}$,

$$\mathbb{P}_{\mathbf{1},x}^N \left(A_t^i = 1 \right) = \mathbb{E}_{\mathbf{1},x}^N \left[\frac{G_t(x_t)}{\sum_{j=1}^N G_t(Z_t^j)} \right].$$

$\square$

Lemma 21. *Assume that for any $\epsilon > 0$*

$$\pi\text{-ess}\inf_x \mathbb{P}_{\mathbf{1},x}^N \left(\frac{G_t(Z_t^2)}{G_t(x_t)} \geq \epsilon \right) = 0.$$

Then, (9) *holds*.

Proof. For any $\epsilon, \delta > 0$ there exists $A_{\epsilon,\delta}$ such that $\pi(A_{\epsilon,\delta}) > 0$ and for $x \in A_{\epsilon,\delta}$

$$\mathbb{P}_{\mathbf{1},x}^N \left(\frac{G_t(Z_t^2)}{G_t(x_t)} \geq \epsilon \right) < \delta.$$

Because of exchangeability, for any x and $2 \leq k \leq N$,

$$\mathbb{P}_{\mathbf{1},x}^N \left(\frac{G_t(Z_t^k)}{G_t(x_t)} \geq \epsilon \right) = \mathbb{P}_{\mathbf{1},x}^N \left(\frac{G_t(Z_t^2)}{G_t(x_t)} \geq \epsilon \right).$$

Denote $B = \{ \frac{\sum_{k=2}^N G_t(Z_t^k)}{G_t(x_t)} \geq (N-1)\epsilon \}$, then for $x \in A_{\epsilon,\delta}$ also

$$\mathbb{P}_{\mathbf{1},x}^N(B) \leq \sum_{k=2}^N \mathbb{P}_{\mathbf{1},x}^N \left(\frac{G_t(Z_t^k)}{G_t(x_t)} \geq \epsilon \right) < (N-1)\delta.$$

We may bound for any $x \in A_{\epsilon,\delta}$,

$$\mathbb{E}_{\mathbf{1},x}^N \left[\frac{G_t(x_t)}{\sum_{k=1}^N G_t(Z_t^k)} \right] \geq \mathbb{E}_{\mathbf{1},x}^N \left[\mathbb{I}\{B^{\complement}\} \frac{G_t(x_t)}{\sum_{k=1}^N G_t(Z_t^k)} \right]$$

$$\geq \mathbb{E}_{\mathbf{1},x}^N \left[\{B^{\complement}\} \frac{1}{1 + (N-1)\epsilon} \right]$$

$$\geq \frac{1 - (N-1)\delta}{1 + (N-1)\epsilon}.$$

Letting $\epsilon, \delta \to 0$ completes the proof. $\square$

Lemma 22. *Assume that there exists $t \in [T]$ such that $\pi_t\text{-ess}\sup_{x_t} G_t(x_t) = \infty$, and if $t \geq 2$, suppose also that for any $A \in \mathcal{B}(\mathsf{Z}^{1:t-1})$ and $B \in \mathcal{B}(\mathsf{Z})$,*

$$\pi_{1:t-1}(A) > 0 \quad \text{and} \quad \pi_t(B) > 0 \quad \Rightarrow \quad \pi_{1:t}(A \times B) > 0.$$

Then, *the assumption of Lemma* 21 *and consequently* (9) *holds for t*.

Proof. Assume that $t \in \{2, \ldots, T\}$, and for any $x_{1:t-1} \in Z^{t-1}$ let $\mu_{x_{1:t-1}}$ denote the distribution of $G_t(Z_t^2)$ under $\mathbb{P}_{1,x_{1:t-1}}^N$. By [2], Lemma 35, there exists $A \in \mathcal{B}(Z^{t-1})$ such that $\pi_{1:t-1}(A) \geq 1/2$ and the family $\{\mu_{x_{1:t-1}}\}_{x_{1:t-1} \in A}$ is tight. Therefore, for any $\epsilon, \delta > 0$ there exists $\bar{G}_t < \infty$ such that $\mathbb{P}_{1,x}^N(G_t(Z_t^2)/\bar{G}_t \geq \epsilon) < \delta$ for all $x_{1:t-1} \in A$. Because $\pi_t\text{-ess sup}_{x_t} G_t(x_t) = \infty$, the set $A \times \{x_t : G_t(x_t) \geq \bar{G}_t\} \times Z^{T-t-1}$ is of positive π-measure. The case $t = 1$ follows similarly because the distribution of $G_1(Z_1^2)$ is independent of x. $\qquad \square$

Remark 23. An immediate implication of Propositions 20 and 11 and Lemma 22 is that if π is equivalent to a Lebesgue or counting measure on X then P_N is geometrically ergodic for any $N \geq 2$ if and only if $\pi_t\text{-ess sup}_{x_t} G_t(x_t) < \infty$ for all $t \in [T]$. This covers many applications in statistics, where often the potentials G_t are strictly positive and for any $x_t \in Z$, the Markov kernel $M_t(x_t, \cdot)$ is equivalent to a Lebesgue or counting measure on Z.

Proposition 20 does not characterize all situations in which P_N fails to be geometrically ergodic. Indeed, in the following example (9) does not hold, and P_N still fails to be geometrically ergodic.

Example. Let $Z = \mathbb{N}$, $T = 2$, $G_1(z) \equiv 1$ and $M_1(z_1)$ be any probability distribution supported on $\mathbb{N}$ (e.g., a Poisson distribution). Define $M_2(z_1, z_2) = \frac{1}{2}\delta_{2z_1}(z_2) + \frac{1}{2}\delta_{2z_1+1}(z_2)$ and $G_2(z_2) = z_2$. It is not difficult to see that this example does not satisfy (9), but $\pi_2\text{-ess sup}_{z_2} G_2(z_2) = \infty$. It is easy to observe as well that the sets $A_n := \{(n, 2n), (n, 2n + 1)\}$ satisfy $\pi(A_n) > 0$ and that $\inf_{x \in A} P_N(x, A_n) \geq 1 - \delta_n$ where $\delta_n \to 0$ as $n \to \infty$.

Our findings above suggest that the essential boundedness of the potentials could in fact be a necessary condition for geometric ergodicity. We have considered also various other examples, and it seems that in any specific scenario it is easy to identify "sticky" sets and conclude by Lemma 19. However, we have yet to identify such sets in general, and so have resorted to stating the following.

Conjecture 24. *Suppose $\pi_t\text{-ess sup}_{x_t} G_t(x_t) = \infty$ for some $t \in [T]$. Then, the i-cSMC kernel is not geometrically ergodic for any $N \in \mathbb{N}$.*

7. The particle Gibbs sampler

In numerous situations of practical interest, one is interested in sampling from a probability distribution $\pi(\mathrm{d}\theta \times \mathrm{d}x)$ defined on some measurable space $(\Theta \times X, \mathcal{B}(\Theta) \times \mathcal{B}(X))$ for which direct sampling is difficult, but sampling from the associated conditional probability distributions $\pi_\theta(\mathrm{d}x)$ and $\pi_x(\mathrm{d}\theta)$ for any $(\theta, x) \in \Theta \times X$ turns out to be easier. In fact, when sampling exactly from these conditionals is possible one can define the two stage Gibbs sampler [20] which alternately samples from these conditional distributions. More precisely, let us define, for any $(\theta, x) \in \Theta \times X$ and $S \in \mathcal{B}(\Theta) \times \mathcal{B}(X)$,

$$\Gamma(\theta, x; S) := \int_S \pi_x(\mathrm{d}\vartheta)\pi_\vartheta(\mathrm{d}y). \tag{11}$$

This can be interpreted as a Markov transition probability, and is precisely the Markov kernel underpinning the standard two stage Gibbs sampler. The corresponding Markov chain $\{(\theta_i, X_i), i \geq 0\}$ on $\Theta \times \mathsf{X}$ leaves π invariant and is ergodic under fairly general and natural conditions. In fact, it can be shown that $\{X_i, i \geq 0\}$ and $\{\theta_i, i \geq 0\}$ are themselves Markov chains leaving the marginals $\pi(\mathrm{d}x)$ and $\pi(\mathrm{d}\theta)$ invariant respectively. For reasons which will appear clearer below, we define for any $(x_0, S) \in \mathsf{X} \times \mathcal{B}(\mathsf{X})$ the Markov transition probability $\Gamma_x(x_0, S) := \Gamma(x_0, \Theta \times S)$ corresponding to the Markov chain $\{X_i, i \geq 0\}$ (we point out that the index x in this notation is a name, not a variable). In some situations, however, while sampling from the conditional distribution $\pi_x(\mathrm{d}\theta)$ may be routine, sampling from $\pi_\theta(\mathrm{d}x)$ may be difficult and this step is instead replaced by a Markov transition probability $\Pi_\theta(x, \mathrm{d}y)$ leaving $\pi_\theta(\mathrm{d}x)$ invariant for any $\theta \in \Theta$. The resulting algorithm, whose transition kernel Φ is given below, is often referred to as "Metropolis-within-Gibbs" in the common situation where Π_θ is a Metropolis–Hastings transition kernel – we will however use this name in order to refer to the general scenario. In the particular situation where Π_θ is a cSMC transition kernel, the resulting algorithm is known as the particle Gibbs (PGibbs) sampler [1]. We note that in the general scenario, for any $(\theta_0, x, S) \in \Theta \times \mathsf{X} \times (\mathcal{B}(\Theta) \times \mathcal{B}(\mathsf{X}))$

$$\Phi(x, S) = \Phi(\theta_0, x; S) := \int_S \pi_x(\mathrm{d}\theta) \Pi_\theta(x, \mathrm{d}y). \tag{12}$$

Similarly to above, one can show that $\{X_i, i \geq 1\}$ defines a Markov chain, with transition kernel, for $(x_0, S) \in \mathsf{X} \times \mathcal{B}(\mathsf{X})$, $\Phi_x(x_0, S) := \Phi(x_0, \Theta \times S)$ which is $\pi(\mathrm{d}x)$-reversible, and positive as soon as Π_θ defines a positive operator for any $\theta \in \Theta$. Indeed since for any $f, g \in L^2(\mathsf{X}, \pi)$,

$$\int_\mathsf{X} f(x)\pi(\mathrm{d}x) \int_{\Theta \times \mathsf{X}} \pi_x(\mathrm{d}\theta) \Pi_\theta(x, \mathrm{d}y) g(y) = \int_\Theta \pi(\mathrm{d}\theta) \int_{\mathsf{X}^2} f(x)g(y)\pi_\theta(\mathrm{d}x)\Pi_\theta(x, \mathrm{d}y)$$

$$= \int_\Theta \pi(\mathrm{d}\theta) \int_{\mathsf{X}^2} f(x)g(y)\pi_\theta(\mathrm{d}y)\Pi_\theta(y, \mathrm{d}x),$$

we deduce the reversibility from the choice $f(x) = \mathbb{I}\{x \in S_1\}$ and $g(x) = \mathbb{I}\{x \in S_2\}$ for $S_1, S_2 \in \mathcal{B}(\mathsf{X})$ and the positivity by letting $g = f$. This motivates the following simple result, which again draws on the standard Hilbert space techniques outlined in [2], Appendix C, and is to the best of our knowledge not available in the literature. We naturally remark that Γ is a particular instance of Φ corresponding to the case where for any $(\theta, x) \in \Theta \times \mathsf{X}$, $\Pi_\theta(x, \cdot) = \pi_\theta(\cdot)$, therefore also implying that Γ_x is self-adjoint. Our first result, Theorem 25, takes advantage of the fact that Γ_x is reversible, and therefore focuses on the asymptotic variance of functions $f \in L^2(\mathsf{X}, \pi)$. Corollary 26 follows from this result, providing a sufficient condition for geometric ergodicity of the PGibbs Markov chain. Our second result, Theorem 27, focuses on functions $g \in L^2(\Theta, \pi)$, but the same technique is not directly applicable in this scenario. Some of our results concern Dirichlet forms: for a generic μ-reversible Markov kernel and a function $f \in L^2(\mathsf{E}, \mu)$ we define the Dirichlet form $\mathcal{E}_\Pi(f) := \langle f, (I - \Pi)f \rangle_\mu$.

Theorem 25. *Let π be a probability distribution defined on $(\Theta \times \mathsf{X}, \mathcal{B}(\Theta) \times \mathcal{B}(\mathsf{X}))$ and let $\{\Pi_\theta, \theta \in \Theta\}$ be a family of Markov transition probabilities $\{\Pi_\theta, \theta \in \Theta\}$ such that for any $\theta \in \Theta$*

the Markov kernel Π_θ is reversible with respect to π_θ, and let Γ and Φ be as in (11) and (12). *Define*

$$\varrho := \inf_{f \in L^2(\mathsf{X}, \pi)} \frac{\int_\Theta \pi(\mathrm{d}\theta) \, \mathrm{var}_{\pi_\theta}(f) \mathrm{Gap}(\Pi_\theta)}{\int_\Theta \pi(\mathrm{d}\theta) \, \mathrm{var}_{\pi_\theta}(f)}. \tag{13}$$

Then, for any $f \in L^2(\mathsf{X}, \pi)$ we have the following inequalities,

   (a) *for the Dirichlet forms,*

$$2\mathcal{E}_{\Gamma_x}(f) \geq \mathcal{E}_{\Phi_x}(f) \geq \varrho \times \mathcal{E}_{\Gamma_x}(f),$$

   (b) *for the right spectral gaps*

$$2\mathrm{Gap}(\Gamma_x) \geq \mathrm{Gap}(\Phi_x) \geq \varrho \times \mathrm{Gap}(\Gamma_x),$$

   (c) *if the asymptotic variances,*

$$0 \leq \frac{\mathrm{var}(f, \Gamma_x) - \mathrm{var}_\pi(f)}{2} \leq \mathrm{var}(f, \Phi_x) \leq \left(\varrho^{-1} - 1\right) \mathrm{var}_\pi(f) + \varrho^{-1} \mathrm{var}(f, \Gamma_x),$$

   where the latter inequality holds for $\varrho > 0$.
   (d) *In addition if*
      (i) *there exist $\epsilon > 0$ such that for all $\theta \in \Theta$ and all $(x, B) \in \mathsf{X} \times \mathcal{B}(\mathsf{X})$, the minorisation inequality $\Pi_\theta(x, B) \geq \epsilon \pi_\theta(B)$ holds, then for any $f \in L^2(\mathsf{X}, \pi)$*

$$\frac{\mathrm{var}(f, \Gamma_x) - (1 - \epsilon) \mathrm{var}_\pi(f)}{(2 - \epsilon)} \leq \mathrm{var}(f, \Phi_x),$$

      (ii) *for all $\theta \in \Theta$, Π_θ is a positive operator then for any $f \in L^2(\mathsf{X}, \pi)$*

$$\mathrm{var}(f, \Gamma_x) \leq \mathrm{var}(f, \Phi_x).$$

Proof. We prove the first point. Without loss of generality, we consider any $f \in L_0^2(\mathsf{X}, \pi)$ and notice that

$$\mathcal{E}_{\Phi_x}(f) = \int_\Theta \pi(\mathrm{d}\theta) \mathcal{E}_{\Pi_\theta}(f),$$

since

$$\int_{\Theta \times \mathsf{X}^2} \pi(\mathrm{d}x) \pi_x(\mathrm{d}\theta) \Pi_\theta(x, \mathrm{d}y) \big[f(x) - f(y) \big]^2$$

$$= \int_\Theta \pi(\mathrm{d}\theta) \int_{\mathsf{X}^2} \pi_\theta(\mathrm{d}x) \Pi_\theta(x, \mathrm{d}y) \big[f(x) - f(y) \big]^2.$$

Now using that $\mathcal{E}_{\Gamma_x}(f) = \frac{1}{2}\int_{\Theta\times X^2}\pi(dx)\pi_x(d\theta)\pi_\theta(dy)[f(x)-f(y)]^2 = \int_\Theta\pi(d\theta)\,\mathrm{var}_{\pi_\theta}(f)$ and letting $\bar{f}_\theta := f - \pi_\theta(f)$ for any $\theta\in\Theta$, we obtain

$$\mathcal{E}_{\Phi_x}(f) = \int_\Theta \pi(d\theta)\mathrm{var}_{\pi_\theta}(f)\frac{\int_\Theta\pi(d\theta)\mathcal{E}_{\Pi_\theta}(f)}{\int_\Theta\pi(d\theta)\,\mathrm{var}_{\pi_\theta}(f)}$$

$$= \mathcal{E}_{\Gamma_x}(f)\times \frac{\int_\Theta\pi(d\theta)\mathbb{I}\{\mathrm{var}_{\pi_\theta}(f)>0\}\,\mathrm{var}_{\pi_\theta}(f)\frac{\mathcal{E}_{\Pi_\theta}(\bar{f}_\theta)}{\mathrm{var}_{\pi_\theta}(\bar{f}_\theta)}}{\int_\Theta\pi(d\theta)\,\mathrm{var}_{\pi_\theta}(f)}$$

$$\geq \mathcal{E}_{\Gamma_x}(f)\times \frac{\int_\Theta\pi(d\theta)\,\mathrm{var}_{\pi_\theta}(f)\mathrm{Gap}(\Pi_\theta)}{\int_\Theta\pi(d\theta)\,\mathrm{var}_{\pi_\theta}(f)}$$

$$\geq \mathcal{E}_{\Gamma_x}(f)\times \inf_{g\in L_0^2(X,\pi)}\frac{\int_\Theta\pi(d\theta)\,\mathrm{var}_{\pi_\theta}(g)\mathrm{Gap}(\Pi_\theta)}{\int_\Theta\pi(d\theta)\,\mathrm{var}_{\pi_\theta}(g)},$$

where we have used that for any $g\in L_0^2(X,\pi)$, $\mathcal{E}_{\Pi_\theta}(g)\leq 2\,\mathrm{var}_{\pi_\theta}(g)$ and that the set $A := \{\theta\in\Theta : \mathrm{var}_{\pi_\theta}(\bar{f}_\theta)=\infty\}$ satisfies $\pi(A\times X)=0$. The latter result follows from $\mathrm{var}_\pi(f)<\infty$ and the variance decomposition identity: $\|f\|_\pi^2 = \|f-\bar{f}_\theta\|_\pi^2 + \|\bar{f}_\theta\|_\pi^2$. We deduce (a) from the last inequality. Points (b) and (c) then follow from [2], Lemma 33.

We next turn into (d). As above, we find that

$$\mathcal{E}_{\Phi_x}(f) \leq \mathcal{E}_{\Gamma_x}(f)\times \frac{\int_\Theta\pi(d\theta)\mathbb{I}\{\mathrm{var}_{\pi_\theta}(f)>0\}\,\mathrm{var}_{\pi_\theta}(f)\sup_{g\in L_0^2(X,\pi_\theta)}(\mathcal{E}_{\Pi_\theta}(g)/\mathrm{var}_{\pi_\theta}(g))}{\int_\Theta\pi(d\theta)\,\mathrm{var}_{\pi_\theta}(f)}.$$

Under the uniform minorisation condition, we have $\mathcal{E}_{\Pi_\theta}(g)\leq(2-\epsilon)\,\mathrm{var}_{\pi_\theta}(g)$ [2], Proposition 32, and consequently $\mathcal{E}_{\Phi_x}(f)\leq(2-\varepsilon)\mathcal{E}_{\Gamma_x}(f)$. When Π_θ is a positive operator for any $\theta\in\Theta$, we have $\mathcal{E}_{\Pi_\theta}(g)\leq\mathrm{var}_{\pi_\theta}(g)$ and consequently $\mathcal{E}_{\Phi_x}(f)\leq\mathcal{E}_{\Gamma_x}(f)$. □

Remark 26. In relation to Theorem 25:

(a) it may be easier in practice to use the lower bound $\underline{\varrho} := \inf_{\theta\in\Theta}\mathrm{Gap}(\Pi_\theta)\leq\varrho$ which leads to $\mathrm{Gap}(\Phi_x)\geq\underline{\varrho}\times\mathrm{Gap}(\Gamma_x)$ and $\mathrm{var}(f,\Phi_x)\leq(\underline{\varrho}^{-1}-1)\,\overline{\mathrm{var}}_\pi(f)+\underline{\varrho}^{-1}\mathrm{var}(f,\Gamma_x)$ when $\underline{\varrho}>0$,

(b) one could suggest iterating Π_θ sufficiently many times, say $\bar{k}_\theta$ times, in order to ensure that $\Pi_\theta^{k_\theta}$ satisfies the uniform in θ properties of the type suggested above. This would require however a computable quantitative bound on the spectral gap of Π_θ,

(c) the lower bound in (c) is motivated by the fact that $\{\Pi_\theta, \theta\in\Theta\}$ may be a family with non-positive elements, which may introduce negative correlations. On the contrary in the situation where $\{\Pi_\theta, \theta\in\Theta\}$ is a collection of positive operators (e.g., cSMC kernels) then (b) implies that Φ_x is geometrically ergodic as soon as Γ_x is geometrically ergodic and $\varrho>0$ (and of course Γ_x is always positive) and (d)(ii) that Φ_x is always inferior to Γ_x in terms of asymptotic variance. In the context of the PGibbs sampler the latter result parallels what is known for pseudo-marginal algorithms [5],

(d) we note that from [22], Theorem 1; Proposition 1, Φ is geometrically ergodic as soon as Φ_x is geometrically ergodic.

Now we show how these results can be transferred to the $\{\theta_i\}$ chain.

Theorem 27. *Let the notation be as in Theorem 25. Then,*

(a) *assume that for some class of functions $\mathcal{G} \subset \{g : \mathsf{X} \to \mathbb{R} : \pi(|g|) < \infty\}$ there exists a function $|\cdot|_{\mathcal{G}} : \mathcal{G} \to [0, \infty]$ and $\rho \in [0, 1)$ such that for any probability distribution v on $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ there exist $W_v \in [0, \infty]$ such that for all $g \in \mathcal{G}$ and any $k \geq 1$*

$$\left| v\Phi_x^k(g) - \pi(g) \right| \leq |g|_{\mathcal{G}} W_v \rho^k,$$

then for any $f : \Theta \to \mathbb{R}$ such that $\bar{f}(x) := \pi_x(f) \in \mathcal{G}$ and any $k \geq 2$

$$\left| v\Phi^k(f) - \pi(f) \right| \leq |\bar{f}|_{\mathcal{G}} W_v \rho^{k-1},$$

(b) *for any $f \in L^2(\Theta, \pi)$, letting for any $x \in \mathsf{X}$ $\bar{f}(x) := \pi_x(f) \in L^2(\mathsf{X}, \pi)$, we have for any $k \geq 1$*

$$\left\langle f, \Phi^k f \right\rangle_\pi = \left\langle \bar{f}, \Phi_x^{k-1} \bar{f} \right\rangle_\pi$$

and

$$\mathrm{var}(f, \Phi) = \mathrm{var}_\pi(f) + \mathrm{var}_\pi(\bar{f}) + \mathrm{var}(\bar{f}, \Phi_x),$$

(c) *if $\varrho > 0$ defined in (13), then for $f \in L^2(\Theta, \pi)$*

$$\mathrm{var}(f, \Phi) \leq \mathrm{var}_\pi(f) + \varrho^{-1} \mathrm{var}_\pi(\bar{f}) + \varrho^{-1} \mathrm{var}(\bar{f}, \Gamma_x)$$
$$\leq \left(1 - \varrho^{-1}\right) \mathrm{var}_\pi(f) + \varrho^{-1} \mathrm{var}(f, \Gamma),$$

(d) *if for all $\theta \in \Theta$, Π_θ is a positive operator, then for $f \in L^2(\Theta, \pi)$ $\mathrm{var}(f, \Phi) \geq \mathrm{var}(f, \Gamma)$.*

Proof. We remark that without loss of generality we can let $f \in L_0^2(\Theta, \pi)$ throughout. First note that for $f \in L_0^2(\Theta, \pi)$ and any $(\theta, x_0) \in \Theta \times \mathsf{X}$

$$\Phi(\theta, x_0; f) = \Phi(x_0; f) = \pi_{x_0}(f) = \bar{f}(x_0),$$

and for $g \in L^2(\mathsf{X}, \pi)$ and any $p \geq 1$, $\Phi^p(\theta, x_0; g) = \Phi_x^p(x_0; g)$. The first result is straightforward upon remarking that for $k \geq 1$

$$\Phi^{k+1}(x_0, f) - \pi(f) = \Phi_x^k(x_0, \bar{f}) - \pi(\bar{f}).$$

For the second and third point, using the remarks above, for $f \in L_0^2(\Theta, \pi)$ and $k \geq 1$

$$\left\langle f, \Phi^k f \right\rangle_\pi = \left\langle f, \Phi^{k-1} \bar{f} \right\rangle_\pi = \left\langle \bar{f}, \Phi_x^{k-1} \bar{f} \right\rangle_\pi.$$

Now $\|f\|_\pi^2 = \langle f - \bar{f} + \bar{f}, f - \bar{f} + \bar{f} \rangle_\pi = \|\bar{f}\|_\pi^2 + \|f - \bar{f}\|_\pi^2$, which is the variance decomposition identity and by noting that $\pi(\bar{f}) = 0$ lets us deduce that $f \in L_0^2(\Theta, \pi)$ implies that

$\bar{f} \in L^2_0(X, \pi)$. Now,

$$\text{var}(f, \Phi) = \|f\|^2_\pi + 2\sum_{k=1}^\infty \langle f, \Phi^k f \rangle_\pi = \|f\|^2_\pi + 2\sum_{k=1}^\infty \langle \bar{f}, \Phi_x^{k-1} \bar{f} \rangle_\pi$$

$$= \|f\|^2_\pi + 2\|\bar{f}\|^2_\pi + 2\sum_{k=1}^\infty \langle \bar{f}, \Phi_x^k \bar{f} \rangle_\pi = \|f\|^2_\pi + \|\bar{f}\|^2_\pi + \text{var}(\bar{f}, \Phi_x).$$

We conclude by noting that for $f \in L^2(X, \pi)$ then $\text{var}_\pi(f) = \|f - \pi(f)\|^2_\pi$ and $\text{var}_\pi(\bar{f}) = \|\bar{f} - \pi(f)\|^2_\pi = \|\overline{f - \pi(f)}\|^2_\pi$. We will also use the equality above for Γ and Γ_x, since again the latter corresponds to a particular instance of the above. We can now use the bound from Theorem 25, which leads, for $f \in L^2_0(\Theta, \pi)$, to

$$\text{var}(f, \Phi) \leq \|f\|^2_\pi + \|\bar{f}\|^2_\pi + (\varrho^{-1} - 1)\|\bar{f}\|^2_\pi + \varrho^{-1}\text{var}(\bar{f}, \Gamma_x)$$

$$= \|f\|^2_\pi + \varrho^{-1}\|\bar{f}\|^2_\pi + \varrho^{-1}\text{var}(\bar{f}, \Gamma_x).$$

From the remark above, we deduce that

$$\|f\|^2_\pi + \varrho^{-1}\|\bar{f}\|^2_\pi + \varrho^{-1}\text{var}(\bar{f}, \Gamma_x) \leq \|f\|^2_\pi + \varrho^{-1}\|\bar{f}\|^2_\pi + \varrho^{-1}\big[\text{var}(f, \Gamma) - \|f\|^2_\pi - \|\bar{f}\|^2_\pi\big]$$

$$= (1 - \varrho^{-1})\|f\|^2_\pi + \varrho^{-1}\text{var}(f, \Gamma).$$

We conclude as above. The final statement follows from $\text{var}(\bar{f}, \Phi_x) \geq \text{var}(\bar{f}, \Gamma_x)$ (see Theorem 25) and the equality established above for Φ and Φ_x and Γ and Γ_x. $\square$

Corollary 28. *Consider the PGibbs sampler with $N \geq 2$ particles with kernel Φ_N defined as in* (12) *such that for any $\theta \in \Theta$, $\Pi_\theta = P_{\theta, N}$ is the i-cSMC kernel as defined in Section 3 for the families $\{M_{\theta, t}\}$ and $\{G_{\theta, t}\}$ of kernels and potentials on $Z \times B(Z)$ and Z respectively. For any $\theta \in \Theta$ we let $\gamma_{\theta, T}$ be the corresponding normalizing constant as defined below* (3). *Then, the results of Theorems 25 and 27 hold as follows:*

(a) *if*

$$\pi\text{-ess sup}_\theta \frac{\prod_{t=1}^T \bar{G}_{\theta, t}}{\gamma_{\theta, T}} < \infty,$$

then $\varrho \geq \epsilon_N$ as defined in Corollary 12,

(b) *or we have the uniform mixing condition, for some $0 \leq \alpha < \infty$,*

$$\pi\text{-ess sup}_{\theta, z} \frac{Q_{\theta, p, p+k}(1)(z)}{\eta_{\theta, p} Q_{\theta, p, p+k}(1)} \leq \alpha,$$

then $\varrho \geq \epsilon_N$ as defined in Corollary 14.

In particular, in both cases ϱ convergences to one as $N \to \infty$, implying that the spectral gaps and the asymptotic variances associated with the PGibbs sampler converge to those of the related Gibbs sampler.

Remark 29. It is worth noting that terms related to $\gamma_{\theta,T}$ appear in all these bounds. So, for example, in the first part it is not sufficient that our potentials $\{G_{\theta,t}\}$ are essentially bounded, but it is sufficient if, for all $t \in [T]$, π_t-ess $\sup_{\theta,x_t} G_{\theta,t}(x_t)/\eta_{\theta,t}(G_t)$ is bounded.

8. Discussion

The developments above go some way in characterizing the behaviour of i-cSMC and associated PGibbs Markov chains, and raise a number of possible future directions for research. We have already embarked upon investigating some potentially practical uses of the minorization conditions and spectral properties for these chains. Of particular interest in practice is how to choose N in the i-cSMC algorithm so as to balance the trade off between mixing properties of P_N and the total number of iterations that can be performed with limited computational resources. Remark 15, for example, can be used to find approximately good values of N in this spirit, but can only serve as a heuristic. In particular, while Proposition 6 may provide a fairly accurate bound in the large N regime, it is unclear how much is lost in applying Jensen's inequality, and consequently how accurate estimates such as those in Remark 15 can be. It is possible that results such as those in [6] may provide a way to exploit additional structure often found in statistical applications.

The results for the i-cSMC and PGibbs Markov chains developed here can be compared and contrasted with similar results for the Particle Independent Metropolis–Hastings (PIMH) and PMMH Markov chains [1]. We summarize here the detailed comparison provided in [2], Appendix F. Like i-cSMC, PIMH is an exact approximation of an independent sampler but PMMH is an exact approximation of an idealized Metropolis–Hastings kernel, rather than a Gibbs sampler. Just as i-cSMC can be viewed as a constituent element of PGibbs, PIMH can be viewed as playing the same role within PMMH. Central to the analysis of PIMH is the essential supremum of the normalizing constant estimate $\hat{\gamma}_T^N(Z_{1:T})$ introduced in Section 4 with respect to the law of a standard SMC algorithm and indeed the PIMH Markov chain is (uniformly) geometrically ergodic if and only if this supremum is finite as a consequence of the characterisation of independent Metropolis–Hastings chains in [17]. However, it can also be seen that the rate of convergence of PIMH will typically not improve as N increases, in contrast with the convergence for the i-cSMC (see Propositions 11 and 13).

For PMMH, [5] show that if the essential supremum of the relative normalizing constant estimate $\hat{\gamma}_{\theta,T}^N(Z_{1:T})/\gamma_{\theta,T}$ is moreover bounded essentially uniformly in θ then the existence of a spectral gap of the idealized Metropolis–Hastings Markov kernel it approximates is inherited by PMMH. However, the rate of convergence of the PMMH Markov chain when this occurs does not improve in general as N increases, in contrast to our results for PGibbs Markov chains. In this context, weak convergence in N of the asymptotic variance of estimates of $\pi(f)$ to the corresponding asymptotic variance of the Metropolis–Hasting kernel is nevertheless provided by [5], Proposition 19, for all $f \in L^2(\Theta, \pi)$ but this can be contrasted with quantitative bounds obtained in Theorem 27.

The one step uniform minorization condition in Corollary 7, where the minorization measure is the invariant distribution of the Markov chain, suggests that it may be possible to apply coupling from the past techniques (see, e.g., [12,18,19]) in order to produce samples from exactly this

distribution. It is, however, not clear how to implement such an algorithm in general, although [13] provides a perfect simulation algorithm motivated by Theorem 1. Finally, our analysis has focused mainly on the case where the essential boundedness condition holds. However, a refined analysis may permit characterization of the i-cSMC and hence the PGibbs Markov chains even in the absence of this condition, with parallels to [5].

Acknowledgements

Supplementary Material

Proofs (DOI: 10.3150/15-BEJ785SUPP; .pdf). The supplementary material contains proofs, intermediate results and a more detailed comparison of the i-cSMC and PGibbs with the PIMH and PMMH.

References

[1] Andrieu, C., Doucet, A. and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 269–342. MR2758115

[2] Andrieu, C., Lee, A. and Vihola, M. Supplement to "Uniform ergodicity of the iterated conditional SMC and geometric ergodicity of particle Gibbs samplers." DOI:10.3150/15-BEJ785SUPP.

[3] Andrieu, C., Lee, A. and Vihola, M. (2013). Uniform ergodicity of the iterated conditional SMC and geometric ergodicity of particle Gibbs samplers. Preprint. Available at arXiv:1312.6432.

[4] Andrieu, C. and Roberts, G.O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.* **37** 697–725. MR2502648

[5] Andrieu, C. and Vihola, M. (2015). Convergence properties of pseudo-marginal Markov chain Monte Carlo algorithms. *Ann. Appl. Probab.* **25** 1030–1077. MR3313762

[6] Bérard, J., Del-Moral, P. and Doucet, A. (2013). A lognormal central limit theorem for particle approximations of normalizing constants. *Electron. J. Probab.* **19**.

[7] Cérou, F., Del Moral, P. and Guyader, A. (2011). A nonasymptotic theorem for unnormalized Feynman–Kac particle models. *Ann. Inst. Henri Poincaré Probab. Stat.* **47** 629–649. MR2841068

[8] Chopin, N. and Singh, S.S. (2013). On the particle Gibbs sampler. *Bernoulli* **21** 1855–1883.

[9] Del Moral, P. (2004). *Feynman–Kac Formulae. Probability and Its Applications* (*New York*). New York: Springer. MR2044973

[10] Del Moral, P., Kohn, R. and Patras, F. (2014). On Feynman–Kac and particle Markov chain Monte Carlo models. Preprint. Available at arXiv:1404.5733.

[11] Doucet, A., Pitt, M.K., Deligiannidis, G. and Kohn, R. (2015). Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator. *Biometrika* **102** 295–313. MR3371005

[12] Hobert, J.P. and Robert, C.P. (2004). A mixture representation of π with applications in Markov chain Monte Carlo and perfect sampling. *Ann. Appl. Probab.* **14** 1295–1305. MR2071424

[13] Lee, A., Doucet, A. and Łatuszyński, K. (2014). Perfect simulation using atomic regeneration with application to sequential Monte Carlo. Available at arXiv:1407.5770.

[14] Lee, A. and Łatuszyński, K. (2014). Variance bounding and geometric ergodicity of Markov chain Monte Carlo kernels for approximate Bayesian computation. *Biometrika* **101** 655–671. MR3254907

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

[16] Lindsten, F., Douc, R. and Moulines, E. (2015). Uniform ergodicity of the particle Gibbs sampler. *Scand. J. Stat.* **42** 775–797. MR3391692

[17] Mengersen, K.L. and Tweedie, R.L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24** 101–121. MR1389882

[18] Murdoch, D.J. and Green, P.J. (1998). Exact sampling from a continuous state space. *Scand. J. Stat.* **25** 483–502. MR1650023

[19] Propp, J.G. and Wilson, D.B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. In *Proceedings of the Seventh International Conference on Random Structures and Algorithms* (*Atlanta*, *GA*, 1995) **9** 223–252. Wiley. MR1611693

[20] Robert, C.P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. *Springer Texts in Statistics*. New York: Springer. MR1707311

[21] Roberts, G.O. and Rosenthal, J.S. (1997). Geometric ergodicity and hybrid Markov chains. *Electron. Commun. Probab.* **2** 13–25 (electronic). MR1448322

[22] Roberts, G.O. and Rosenthal, J.S. (2001). Markov chains and de-initializing processes. *Scand. J. Stat.* **28** 489–504. MR1858413

[23] Roberts, G.O. and Tweedie, R.L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83** 95–110. MR1399158

[24] Sherlock, C., Thiery, A.H., Roberts, G.O. and Rosenthal, J.S. (2015). On the efficiency of pseudo-marginal random walk Metropolis algorithms. *Ann. Statist.* **43** 238–275. MR3285606

[25] Whiteley, N. (2013). Stability properties of some particle filters. *Ann. Appl. Probab.* **23** 2500–2537. MR3127943