CONVERGENCE OF SEQUENTIAL MONTE CARLO-BASED
SAMPLING METHODS

JONATHAN H. HUGGINS AND DANIEL M. ROY

ABSTRACT. Originally designed for state-space models, Sequential Monte Carlo (SMC) methods are now routinely applied in the context of general-purpose Bayesian inference. Traditional analyses of SMC algorithms have focused on their application to estimating expectations with respect to intractable distributions such as those arising in Bayesian analysis. However, these algorithms can also be used to obtain approximate samples from a posterior distribution of interest. We investigate the asymptotic and non-asymptotic convergence rates of SMC from this sampling viewpoint. In particular, we study the expectation of the particle approximation that SMC produces as the number of particles tends to infinity. This “expected approximation” is equivalent to the law of a sample drawn from the SMC approximation. We give convergence rates of the Kullback-Leibler divergence between the target and the expected approximation. Our results apply to both deterministic and adaptive resampling schemes. In the adaptive setting, we introduce a novel notion of effective sample size, the \( \infty \)-ESS, and show that controlling this quantity ensures stability of the SMC sampling algorithm. We also introduce an adaptive version of the conditional SMC proposal, which allows us to prove quantitative bounds for rates of convergence for adaptive versions of iterated conditional sequential Monte Carlo Markov chains and associated adaptive particle Gibbs samplers.

1. Introduction
2. Preliminaries
3. Summary of Results
4. Basic Results
5. SMC with Adaptive Resampling
6. Convergence in the Marginal and Predictive Cases
7. Particle MCMC
Appendix A. Additional Proofs
Acknowledgments
References

Date: Wednesday 4\textsuperscript{th} March, 2015.
1. Introduction

Sequential Monte Carlo (SMC) methods are a widely-used class of algorithms for approximate inference \[11–14, 16, 17\]. In the context of Bayesian inference, SMC produces a particle approximation to the posterior distribution as well as an unbiased estimate of the marginal likelihood. Traditionally, particle approximations were intended to be used to estimate expectations such as the posterior probabilities of events or the expected values of parameters, and the analysis of SMC methods focused on this operator perspective, i.e., the analysis of how well a particle approximation approximates the expectation operator induced by the target distribution.

Increasingly, however, SMC methods are being used to produce approximate samples, usually in the inner loop of another approximate inference algorithm. A key example is the class of particle Markov chain Monte Carlo (PMCMC) methods, which aim to combine the best features of SMC and MCMC approaches by using SMC as a proposal mechanism for a Metropolis-Hastings (“particle MH”) or approximate Gibbs (“particle Gibbs”) sampler \[1, 15\]. Characterizing the efficiency of PMCMC methods is an active area of investigation \[2–5, 18, 19\].

When SMC methods are employed for sampling, convergence guarantees from the operator perspective are not appropriate. In this work, we take up the measure perspective on SMC, i.e., we characterize how well SMC-based methods can approximate the target distribution as a measure by assessing convergence in terms of total variation distance, KL divergence, and other measures of discrepancy between distributions.

Let \( \pi \) be a distribution of interest and let \( \pi^N \) denote an \( N \)-particle SMC approximation to \( \pi \). In this work, the majority of our attention is devoted to investigating the mean of \( \pi^N \), denoted \( \bar{\pi}^N \), which can also be understood as the (marginal) distribution of a single sample drawn from the true posterior \( \pi^N \). We use the Kullback-Leibler (KL) divergence from \( \pi \) to \( \bar{\pi}^N \) to measure performance of the SMC sampler. We begin by giving convergence rates for sequential importance sampling and sampling importance resampling (also known as bootstrap filtering) (Section 4). We provide non-asymptotic bounds under minimal assumptions, which can then be used to obtain quantitative bounds under a number of different conditions. These quantitative bounds are obtained using recent results from \[2\] for bounding the expected value of the partition function estimator with respect to the conditional SMC kernel. We also obtain quantitative bounds for the KL divergence from \( \pi \) to \( \bar{\pi}^N \) for adaptive SMC algorithms (Section 5). In the adaptive case our approach uses the \( \alpha \)SMC framework \[23\], and thus applies to a large class of adaptive algorithms. We introduce a novel notion of effective sample size, \( \infty \)-ESS, the control of which is sufficient to guarantee convergence of \( \alpha \)SMC samplers.

We also propose a version of the conditional SMC kernel with adaptive resampling, which we call the conditional \( \alpha \)SMC kernel (Section 7). Mirroring results from \[2\], we give conditions under which iterated conditional \( \alpha \)SMC Markov chains are uniformly ergodic and the associated particle Gibbs sampler with adaptive resampling (the \( \alpha \)PG sampler) is geometrically ergodic. As with the \( \alpha \)SMC samplers, control of the \( \infty \)-ESS is key to proving these ergodicity results.
2. Preliminaries

After fixing some notation, we define the inference problem and present three SMC algorithms: sequential importance sampling, sampling importance resampling, and αSMC. We then introduce the conditional SMC kernel, as well as a novel adaptive version based on αSMC, which allows us to define adaptive versions of the iterated conditional SMC process and an associated particle Gibbs algorithm.

2.1. Notation. For a positive integer $K$, let $[K] \triangleq \{1, 2, \ldots, K\}$. If $x_i, \ldots, x_j$ are elements of a sequence, write $x_{i:j} \triangleq \langle x_i, x_{i+1}, \ldots, x_j \rangle$. We use the following conventions: $\sum_0 = 0$, $\prod_0 = 1$, and $0/0 = 0$.

Let $(S, S)$, $(S', S')$ be measurable spaces. $K : S \times S' \to \mathbb{R}$ is a kernel if $K(\cdot, B)$ is a $(S, S)$-measurable function for all $B \in S'$ and $K(x, \cdot)$ is measure on $(S', S')$ for all $x \in S$. For a measure $\mu$ on $(S, S)$ and kernels $K, K' : S \times S \to \mathbb{R}$, let $\mu K(dy) \triangleq \int \mu(dx) K(x, dy)$ and $K' K(x, dy) \triangleq \int K(x, dy) K'(y, dz)$. We will often use measures and kernels as operators. For a measurable function $\phi : S \to \mathbb{R}$, let $\mu(\phi) \triangleq \mathbb{E}_{\xi \sim \mu}[\phi(\xi)] = \int \phi(x) \mu(dx)$ and $K(x)(\phi) \triangleq \int \phi(y) K(x, dy)$. We will write $\mathbb{V}_{\xi \sim \mu}[\phi(\xi)] \triangleq \mu((\phi - \mu(\phi))^2)$ for the variance of $\phi$ with respect to $\mu$. For measures $\mu, \nu$ on $(S, S)$, we will write $\mu \ll \nu$ to denote that $\mu$ is absolutely continuous with respect to $\nu$, in which case we will write $\frac{d\mu}{d\nu}$ for the $\nu$-almost everywhere ($\nu$-a.e.) unique function $f$ satisfying $\mu(A) = \int_A f d\nu$, for all $A \in S$. When the choice is clear from context, we may write $B(S)$ for the $\sigma$-algebra of the space $S$.

Let $B_b(S)$ be the set of all measurable bounded real functions on $S$ and let $\mathcal{P}(S)$ denote the collection of all probability measures on $(S, S)$. For $\mu, \nu \in \mathcal{P}(S)$, the total variation distance between $\mu$ and $\nu$ is

$$d_{TV}(\mu, \nu) \triangleq \sup_{A \in S} |\mu(A) - \nu(A)|. \quad (1)$$

If $\mu \ll \nu$, then the KL divergence from $\mu$ to $\nu$ is

$$\text{KL}(\mu||\nu) \triangleq \mu(\log d\mu/d\nu) \quad (2)$$

and the $\chi^2$-divergence from $\mu$ to $\nu$ is

$$d_{\chi^2}(\mu, \nu) \triangleq \nu((d\mu/d\nu - 1)^2). \quad (3)$$

Finally, we note that, when there is little at stake, we will ignore measure-theoretic niceties such as the distinction between equality and a.e.-equality.

2.2. Problem Statement. We follow a similar setup and notation to [9]. Let $(X_t)_{t \geq 1}$ be an inhomogeneous Markov chain on $(E, \mathcal{E})$ with transition kernels $(M_t)_{t \geq 2}$ and initial distribution $M_1$. We write $M_1(x, \cdot) = M_1(\cdot)$ when convenient. For all $t \geq 1$ and $x_{t-1} \in E$, assume that $M_t(x_{t-1}, \cdot)$ has a density with respect to some $\sigma$-finite dominating measure (which we denote by $dx$). We abuse notation and write the density of $M_t(x_{t-1}, \cdot)$ as $M_t(x_{t-1}, x_t)$. Denote expectations and variances with respect to the Markov chain by $\mathbb{E}[\cdot]$ and $\mathbb{V}[\cdot]$, respectively. Let $g_t : E \to \mathbb{R}_+$, for $t \geq 1$, be a sequence of $\mathcal{E}$-measurable potential functions on $E$. Denote a product of potentials by $g_{s,t}(x_{s:t}) \triangleq \prod_{r=s}^t g_r(x_r)$ and let $g_0 \equiv 1$.

\footnote{In the state-space setting the potential $g_t$ would be the likelihood from the observation at time $t$: $g_t(x_t) = p_t(y_t | x_t)$.}
The unnormalized predictive and updated measures are defined, respectively, to be
\[ \gamma_{1,t}(\phi_{1,t}) \triangleq \mathbb{E}[\phi_{1,t}(X_{1,t})g_{1,t-1}(X_{1,t-1})] = Q_{0,t}(\phi_{1,t}) \] (4)
and
\[ \gamma_{1,t}(\phi_{1,t}) \triangleq \mathbb{E}[\phi_{1,t}(X_{1,t})g_{1,t}(X_{1,t})] \] (5)
with corresponding marginal versions
\[ \gamma_{t}(\phi_{t}) \triangleq \mathbb{E}[\phi_{t}(X_{t})g_{1,t-1}(X_{1,t-1})] = Q_{0,t}(\phi_{t}) \] (6)
and
\[ \gamma_{t}(\phi_{t}) \triangleq \mathbb{E}[\phi_{t}(X_{t})g_{1,t}(X_{1,t})] . \] (7)

Our goal is to approximate, in a sense to be discussed later, the normalized predictive and updated measures, and their marginal versions, which are defined to be
\[ \pi_{1,t}(\phi_{1,t}) \triangleq \frac{\gamma_{1,t}(\phi_{1,t})}{Z_t}, \quad \eta_{1,t}(\phi_{1,t}) \triangleq \frac{\gamma_{1,t}(\phi_{1,t})}{Z'_t}, \] (8)
\[ \pi_{t}(\phi_{t}) \triangleq \frac{\gamma_{t}(\phi_{t})}{Z_t}, \quad \eta_{t}(\phi_{t}) \triangleq \frac{\gamma_{t}(\phi_{t})}{Z'_t}, \] (9)
where \( Z_t \triangleq \gamma_{t}(1) \) and \( Z'_t \triangleq \gamma'_{t}(1) \) are normalization constants.

For \( t \geq 1 \), let \( Q_{s,t}(x_{t-1}, dx_t) \triangleq g_{t-1}(x_{t-1})M_t(x_{t-1}, dx_t) \), and for \( 0 \leq s < t \), let
\[ Q_{s,t} \triangleq Q_{s+1}Q_{s+2} \cdots Q_t, \] (10)
so \( Q_{s,t+1} = Q_t \). By convention \( Q_{s,t}(x_t, dy_t) = \delta_{x_t}(dy_t) \) and \( Q_{0,t}(dx_t) \) is a measure, not a probability kernel. Notice that for \( s \in [t] \), \( x_s \in E \), and \( \phi_t : E \to \mathbb{R} \),
\[ Q_{s,t}(x_s)(\phi_t) = \mathbb{E}[\phi_t(X_t)g_{s,t-1}(X_{s,t-1}) | X_s = x_s] \] (11)
and \( Q_{0,t}(\phi_t) = M_t Q_{1,t}(\phi_t) \). Generalizing these identities, we will abuse notation and write, for \( s \in [t] \), \( x_s \in E \), and \( \phi_{s,t} : E^{t-s} \to \mathbb{R} \),
\[ Q_{s,t}(x_s)(\phi_{s,t}) \triangleq \mathbb{E}[\phi_{s,t}(X_{s,t})g_{s,t-1}(X_{s,t-1}) | X_s = x_s] \] (12)
and \( Q_{0,t}(\phi_{s,t}) \triangleq M_t Q_{1,t}(\phi_{s,t}) \). Let \( G_{s,t}(y) \triangleq Q_{s,t}(y)(1) \) for \( s \in [t-1] \) and \( G_{0,t} \triangleq Q_{0,t}(1) \).

Observe that \( \gamma'_{1,t}(\phi_{1,t}) = Q_{0,t}(\phi_{1,t}) \), \( \gamma'_{t}(\phi_{t}) = Q_{0,t}(\phi_{t}) \), and \( Z'_t = Z_{t-1} = G_{0,t} \), so the one-step predictive distribution and its marginal version can be written as
\[ \eta_{1,t}(\phi_{1,t}) = \frac{Q_{0,t}(\phi_{1,t})}{G_{0,t}} \quad \text{and} \quad \eta_{t}(\phi_{t}) = \frac{Q_{0,t}(\phi_{t})}{G_{0,t}} . \] (13)

2.3. SMC Algorithms. Sequential Monte Carlo algorithms construct approximations to the distributions \( \pi_{1,t}, \pi_t, \eta_{1,t}, \) and \( \eta_t \), for each \( t = 1, 2, \ldots \) in turn, and use earlier approximations to produce later ones.
Algorithm 1 Sequential Importance Sampling

\begin{align*}
\text{for } n = 1, \ldots, N \text{ do} \\
\quad \text{Sample } X^n_t \text{ from } M_1 \\
\quad \text{Set } X^n_{1,t} \leftarrow X^n_1 \\
\text{end for} \\
\text{for } t = 2, 3, \ldots \text{ do} \\
\quad \text{for } n = 1, \ldots, N \text{ do} \\
\quad \quad \text{Sample } X^n_t \text{ from } M_t(X^n_{t-1,\cdot}) \\
\quad \quad \text{Set } X^n_{1,t} \leftarrow \langle X^n_{1,t-1}, X^n_t \rangle \\
\quad \text{end for} \\
\text{end for} \\
\text{Form the SIS joint and marginal estimators } \pi_{S,N}^{t,1}, \pi_{S,N}^t, \eta_{S,N}^{t,1}, \eta_{S,N}^t.
\end{align*}

2.3.1. Sequential Importance Sampling. The SIS algorithm, given as Algorithm 1, operates by propagating a collection of \( N \) particles \( X_{1,t} \triangleq \{X^n_{1,t}\}_{n=1}^N \) with corresponding nonnegative weights \( W^n_t \triangleq \{W^n_t\}_{n=1}^N \), where

\[ W^n_t \triangleq \frac{g_t(X^n_t)}{\sum_{k=1}^N g_t(X^n_t)} \quad (14) \]

The updated distributions \( \pi_{1,t}^{S,N} \) and \( \pi_t^{S,N} \) are then approximated by

\[ \pi_{1,t}^{S,N} \triangleq \sum_{n=1}^N W^n_t \delta_{X^n_t} \quad \text{and} \quad \pi_t^{S,N} \triangleq \sum_{n=1}^N W^n_t \delta_{X^n_t} \quad (15) \]

and the predictive distributions \( \eta_{1,t}^{S,N} \) and \( \eta_t^{S,N} \) are approximated by

\[ \eta_{1,t}^{S,N} \triangleq \sum_{n=1}^N W^n_{t-1} \delta_{X^n_t} \quad \text{and} \quad \eta_t^{S,N} \triangleq \sum_{n=1}^N W^n_t \delta_{X^n_t}. \quad (16) \]

The estimators of the normalization constants \( Z_t \) and \( Z'_t \) are \( \hat{Z}_t \triangleq \frac{1}{N} \sum_{k=1}^N g_t(X^n_t) \) and \( \hat{Z}'_t \triangleq \frac{1}{N} \sum_{k=1}^N g_{t-1}(X^n_t) \). Expectations with respect the law of the SIS algorithm are written as \( E^{S,N}[\cdot] \).

Remark 2.1. Since all four measures of interest take very similar forms, going forward we will only explicitly define quantities related to them — such as estimators — for the marginal measures \( \pi_t^{S,N} \) and \( \eta_t^{S,N} \).

2.3.2. Sampling Importance Resampling. A more practical algorithm, which does not suffer from the weight degeneracy problem of SIS, is sampling importance resampling (SIR) [14, 16]. The SIR algorithm, given as Algorithm 2, is identical to SIS except for a resampling step performed after each iteration. Let \( \tilde{W}_t \triangleq \{\tilde{W}_t^n\}_{n=1}^N \) denote weights for the particles \( \tilde{X}_{1,t} \triangleq \{\tilde{X}_{1,t}^n\}_{n=1}^N \) at time \( t \), where

\[ \tilde{W}_t^n \triangleq \frac{g_t(\tilde{X}_t^n)}{N} \frac{1}{\sum_{k=1}^N g_t(\tilde{X}_t^k)}. \]

The SIR marginal estimators are

\[ \pi_t^{R,N} \triangleq \sum_{n=1}^N \tilde{W}_t^n \delta_{\tilde{X}_t^n} \quad \text{and} \quad \eta_t^{R,N} \triangleq \sum_{n=1}^N \frac{1}{N} \delta_{\tilde{X}_t^n}. \quad (17) \]
Algorithm 2  Sampling Importance Resampling

\[\text{for } n = 1, \ldots, N \text{ do} \]
\[\quad \text{Sample } X^n_1 \text{ from } M_1 \]
\[\quad \text{Set } X^n_{1,1} \leftarrow X^n_1 \]
\[\text{end for} \]
\[\text{for } t = 2, 3, \ldots \text{ do} \]
\[\quad \text{Sample } A_{t-1} \text{ from } r(\cdot | \tilde{W}_{t-1}) \]
\[\quad \text{for } n = 1, \ldots, N \text{ do} \]
\[\quad \quad \text{Sample } X^n_t \text{ from } M_t(A^n_{t-1}, \cdot) \]
\[\quad \quad \text{Set } X^n_{1,t} \leftarrow \langle X^n_{1,1}, X^n_t \rangle \]
\[\text{end for} \]
\[\text{end for} \]

Form the SIR joint and marginal estimators \(\pi^{R,N}_{1,t}, \pi^{R,N}_t, \eta^{R,N}_{1,t}, \eta^{R,N}_t\).

and the estimators of the normalization constants \(Z_t\) and \(Z'_t\) are \(\tilde{Z}_t \triangleq \frac{1}{N} \prod_{k=1}^t \sum_{k=1}^N g_k(X^k_t)\) and \(\tilde{Z}'_t \triangleq \tilde{Z}_{t-1}\). Expectations with respect the law of the SIR algorithm are written as \(E^{R,N}[\cdot]\).

The SIR algorithm makes use of a resampling kernel with density \(r(a|w)\), where \(w \in \Delta_N \triangleq \{v \in [0,1]^N | \sum_{n=1}^N v_n = 1\} \text{ and } a \in [N]^N\). A minimal assumption required for the algorithm to be correct is that \(r(a_n = k | w) = w_k\). However, we will focus on the simplest case where \(r(a | w) = r_{\text{multi}}(a | w) \triangleq \prod_{n=1}^N w_{a_n}\), i.e., \(a_n | w \overset{\text{iid}}{\sim} \text{Multi}(w)\), for \(n = 1, \ldots, N\).

2.3.3. Adaptive Resampling. The SIR algorithm uses a deterministic resampling scheme. However, it is common for practitioners to use adaptive resampling algorithms to choose when to resample based on the realized particle weights. The most popular adaptive scheme relies on the effective sample size (ESS) criterion to determine when a resampling step is performed \([12, 14, 20, 21]\). Specifically, resampling occurs when the ESS is below some fixed threshold (e.g., \(N/2\)), where the ESS for unnormalized weights \(w = \langle w_1, \ldots, w_N \rangle \in \mathbb{R}_+^N\) is

\[\mathcal{E}^2(w) \triangleq \frac{(\sum_{n=1}^N w_n)^2}{\sum_{n=1}^N w_n^2}\tag{18}\]

and takes on values between 1 to \(N\). The nomenclature arises from interpreting \(\mathcal{E}^2(w)\) as the effective number of particles the algorithm is using if the particles have weights \(w\).

Whiteley, Lee, and Heine \([23]\) recently introduced a very general adaptive algorithm they call \(\alpha\)SMC, which includes SIS, SIR, and numerous other SMC variants as special cases. The \(\alpha\)SMC algorithm, which is given as Algorithm 3, provides a flexible resampling mechanism: at each time \(t\), a stochastic matrix \(\alpha_{t-1}\) is chosen from a set \(\mathbb{A}_N\) of \(N \times N\) matrices. We denote the value in the \(n\)-th row and \(k\)-th column of \(\alpha_{t-1}\) by \(\alpha_{t-1}^{nk}\). The \(\alpha\)SMC marginal estimators are

\[\pi^{\alpha,N}_t = \sum_{n=1}^N \tilde{W}^n_t g_t(X^n_t) / \sum_{k=1}^N W^n_t g_t(X^n_t) \delta X^n_t \quad \text{and} \quad \eta^{\alpha,N}_t = \sum_{n=1}^N \tilde{W}^n_t / \sum_{k=1}^N W^n_t \delta X^n_t\tag{19}\]
Algorithm 3 $\alpha$SMC

for $n = 1, \ldots, N$ do
    Sample $\hat{X}_t^n$ from $M_1$
    Set $\hat{X}_{t,1}^n \leftarrow \hat{X}_t^n$
    Set $\hat{W}_t^n \leftarrow 1$
end for

for $t = 2, 3, \ldots$ do
    Select $\alpha_{t-1}$ from $A_N$ according to some functional of $\hat{X}_{1,t-1}$

    for $n = 1, \ldots, N$ do
        Set $\hat{W}_t^n \leftarrow \sum_{k=1}^N \alpha_{t-1}^{nk} \hat{W}_{t-1}^k g_t(\hat{X}_t^k) / \hat{W}_t^n$
        Sample $A_{t-1}^n$ from Multi \( \frac{\alpha_{t-1}^{nk} \hat{W}_{t-1}^k g_t(\hat{X}_t^k)}{\hat{W}_t^n} \) \( k = 1 \)^N

        Sample $\hat{X}_t^n$ from $M_t(\hat{X}_{1,t-1}^n, \cdot)$
        Set $\hat{X}_{t,1}^n, \hat{X}_t^n \leftarrow \langle \hat{X}_{1,t-1}^n, \hat{X}_t^n \rangle$
    end for
end for

and the estimators of the normalization constants $Z_t$ and $Z'_t$ are

$$
\hat{Z}_t \triangleq \frac{1}{N} \sum_{k=1}^N \hat{W}_t^k g_t(\hat{X}_t^k) \quad \text{and} \quad \hat{Z}'_t \triangleq \frac{1}{N} \sum_{k=1}^N \hat{W}_t^k.
$$

Expectations with respect the law of the SIR algorithm are written as $\mathbb{E}^{n,N}[\cdot]$.

Not only does the $\alpha$SMC formulation aid in analyzing adaptive resampling strategies, it provides a useful framework for devising novel adaptive schemes with attractive computational properties, such as admitting parallelization even on resampling steps. SIS, SIR, and the standard adaptive algorithm can all be obtained as special cases of $\alpha$SMC as follows. SIS is recovered when $\alpha_{t-1} = I_N$, the $N \times N$ identity matrix, while SIR is recovered when $\alpha_{t-1} = 1_{1/N}$, the $N \times N$ matrix with all entries equal to $1/N$. The so-called adaptive particle filter (APF) algorithm is obtained by setting $\alpha_{t-1}$ to $1_{1/N}$ if $\mathcal{E}^2(\hat{W}_{t-1}) < \zeta N$ and to $I_N$ otherwise, where $\zeta \in (0,1]$ is fixed.

Remark 2.2. We will write $\pi$ and $\pi^N$ to denote a generic updated/predictive marginal/full distribution and its associated SIS/SIR/$\alpha$SMC estimator. Expectations with respect to the estimator’s law are written $\mathbb{E}^N[\cdot]$.

2.4. Conditional SMC Processes. We will need to consider variants of SIR and $\alpha$SMC in which one or more particle paths is fixed ahead of time. In the case of SIR, these algorithms are known as conditional SMC algorithms \cite{1, 2}. In the case of $\alpha$SMC we will refer to the algorithms as conditional $\alpha$SMC algorithms. Throughout the section, fix $t \geq 1$, $i \geq 1$, and $N \geq i$.

The $i$-times conditional SMC ($c^i$SMC) process is defined on the space $(E^N \times [N]^t)^{t-1} \times E^N \times [N]$, and is essentially SIR in which the first $i$ particle trajectories $y_{1,t}^1, \ldots, y_{1,t}^i \in E^t$ are a priori fixed, with lineages $k^1, \ldots, k^i \in [N]^t$. For $x_{1,t} \in$
for \( \alpha_1, t-1 \in ([N]^N)^{t-1} \), and \( a_t \in [N] \), the law of the \( c^i \)SMC process is given by

\[
\mathbb{P}_{y_{1,t}, k_{1,t}}(X_1 \in dx_1) \triangleq \prod_{j=1}^{i} \delta_{y^j_1}(dx_1^j) \prod_{n=1}^{N} M_1(dx_n^1), \tag{21}
\]

for \( s = 2, \ldots, t, \)

\[
\mathbb{P}_{y_{1,t}, k_{1,t}}(X_s \in dx_s, A_{s-1} = a_{s-1} | X_{s-1} = x_{s-1}) \]

\[
\triangleq \prod_{j=1}^{i} \delta_{y^j_s}(dx_s^j) \mathbb{I}(a^j_{s-1} = k^j_s) \prod_{n=1}^{N} \frac{g_{s-1}(x^a_{s-1})}{\sum_{i=1}^{N} g_{s-1}(x^i_{s-1})} M_s(x^a_{s-1}, dx^a_s), \tag{22}
\]

and

\[
\mathbb{P}_{y_{1,t}, k_{1,t}}(A_t = a_t | X_t = x_t) \triangleq \frac{g_t(x^{a_t}_t)}{\sum_{n=1}^{N} g_t(x^{n}_t)}. \tag{23}
\]

If \( i = 1 \), we write \( \text{cSMC} \) instead of \( \text{c}^1 \text{SMC} \). Also, note that the \( \text{c}^i \)SMC process is not the same as conditioning on \( i \) trajectories and lineages of the SIR algorithm.

The \( i \)-times conditional \( \alpha \text{SMC (c}^i \alpha \text{SMC) process} \) is defined on the space \((E^N \times [N]^N \times [N]^i)^{t-1} \times E^N \times [N]^N \times [N]^i \times [N]^i \), and is essentially \( \alpha \text{SMC} \) in which the first \( i \) particle trajectories, but not their lineages, are a priori fixed. If \( f \in [N]^i \) are indices of the first \( i \) particles, let \( D(f) \triangleq \prod_{j \neq j'} \mathbb{I}(f_j \neq f_{j'}) \) be the function that indicates whether the indices are distinct. Given trajectories \( y_{1,t}, \ldots, y_{1,t} \in E^t \), for \( x_{1,t} \in (E^N)^t, f_{1,t} \in ([N]^i)^t, a_{1,t-1} \in ([N]^N)^{t-1}, \) and \( a_t \in [N] \), the law of the \( c^i \alpha \text{SMC} \) process is given by

\[
\mathbb{P}_{y_{1,t}}(X_1 \in dx_1, F_1 = f_1) \triangleq C_1 D(f_1) \prod_{j=1}^{i} \frac{1}{N} \delta_{y^j_1}(dx_1^j) \prod_{n=1}^{N} M_1(dx_n^1), \tag{24}
\]

for \( s = 2, \ldots, t, \)

\[
\mathbb{P}_{y_{1,t}}(X_s \in dx_s, A_{s-1} = a_{s-1}, F_s = f_s | X_{s-1} = x_{s-1}) \]

\[
\triangleq C_s D(f_s) \prod_{j=1}^{i} \alpha^{f_j}_{s-1} \mathbb{I}(a^j_{s-1} = f^j_s) \delta_{y^j_s}(dx^j_s) \mathbb{I}(a^j_{s-1} = f^j_s) \]

\[
\times \prod_{n \notin f^j_s} \mathbb{I}(w^n_{s-1} | x^n_{s-1}, a^n_{s-1}, x^n_{s-1}) M_s(x^n_{s-1}, x^n_s) \tag{25}
\]

and

\[
\mathbb{P}_{y_{1,t}}(A_t = a_t | X_{1,t} = x_{1,t}, A_{1,t-1} = a_{1,t-1}) \triangleq \frac{w^n_t g_t(x^{a_t}_t)}{\sum_{n=1}^{N} w^n_t g_t(x^{n}_t)}, \tag{26}
\]

where the \( C_s \) are normalization constants that ensure the expressions are valid probabilities. We have used the notation

\[
w^1_t \triangleq 1, \quad w^n_t \triangleq \sum_{k=1}^{N} \alpha^n_{t-1} w^k_{t-1} g_{t-1}(x^k_{t-1}), \tag{27}
\]
and
\[ r_n(k|w_{s-1}, x_{1:s-1}) \triangleq \frac{\alpha_{s-1}^{nk} w_{s-1}^{k} g_{s-1}(x_{s-1}^{k})}{w_s^{nk}}. \] (28)

If \( i = 1 \), we write \( \alpha \text{SMC} \) instead of \( c^{1} \alpha \text{SMC} \).

2.5. Particle Gibbs and Iterated Conditional SMC. In the language of state-space models, the algorithms so far have been designed to approximate the posterior distribution of a Markov chain given indirect stochastic observations of the chain’s values. However, it is often the case that the chain and the potentials are controlled by a global parameter \( \theta \in \Theta \) for which there is a prior distribution \( \varpi(d\theta) \). We replace \( M_s \) by \( M^\theta_s \) and \( g_s \) by \( g^\theta_s \), then parameterize the other quantities defined previously in terms of \( M_s \) and \( g_s \) by the additional parameter \( \theta \). Throughout this section we fix \( t \) and let \( (Y,Y) \triangleq (E^t, B(E^t)) \). Since \( t \) is fixed we will suppress much of the time notation when possible in order to make the notation less cluttered.

The target distribution on the product space \((\Theta \times Y, B(\Theta \times Y))\) is
\[ \pi(d\theta \times dy) \triangleq \gamma(d\theta \times dy)/Z, \] (29)
where
\[ \gamma(d\theta \times dy) \triangleq \prod_{s=1}^{t} g^\theta_s(y_s) M^\theta_s(y_{s-1}, dy_s) \varpi(d\theta) \quad \text{and} \quad Z \triangleq \gamma(1). \] (30)

We denote the conditional distributions given \( \theta \in \Theta \) or \( y \in Y \) by, respectively, \( \pi_\theta(dy) \) or \( \pi_y(d\theta) \).

Particle Markov chain Monte Carlo (PMCMC) methods use SMC as a component within an MCMC algorithm to obtain approximate samples from \( \pi(d\theta \times dy) \).

We now introduce the particle Gibbs (PG) sampler, which approximates the two-stage Gibbs kernel
\[ \Pi(\theta, y, d\theta \times dz) \triangleq \pi_y(d\theta) \pi_\theta(dz). \] (31)

In many settings, such as non-linear or non-Gaussian state-space models, it is possible to sample from \( \pi_\theta(d\theta) \), but difficult to sample from \( \pi_y(dz) \). The idea is to replace \( \pi_x(dz) \) with an SMC-based approximation \( \Pi_x(y, dz) \) that leaves \( \pi_x(dz) \) invariant, leading to a kernel of the form \( \pi_y(d\theta) \pi_\theta(y, dz) \).

The standard PG sampler employs the iterated conditional SMC (i-cSMC) kernel \( P^{\text{cSMC}}_{\theta} \) to approximate the conditional distribution: \( \Pi_\theta = P^{\text{cSMC}}_{\theta} \). For \( y \in Y \), \( \theta \in \Theta \), and with \( 1 \in \{1\}^N \), the i-cSMC kernel is given by
\[ P^{\text{cSMC}}_{\theta}(y, dz) \triangleq \mathbb{E}_{y,1,\theta}^{\text{cSMC}} \left[ \delta_{x^1_1}(dz) \right]. \] (32)

The invariant distribution of \( P^{\text{cSMC}}_{\theta} \) is \( \pi_\theta \). The family of Markov chains with transition kernels of the form \( P^{\text{cSMC}}_{\theta}(y, dz) \) are called i-cSMC processes.

We now introduce the \( \alpha \) particle Gibbs (\( \alpha \text{PG} \)) sampler, which employs what we will call the iterated conditional \( \alpha \)SMC (i-\( \alpha \text{SMC} \)) kernel \( P^{\alpha}_{\theta} \) to approximate the conditional distribution: \( \Pi_\theta = P^{\alpha}_{\theta} \). For \( y \in Y \), \( \theta \in \Theta \), the i-\( \alpha \)SMC kernel, which has invariant distribution \( \pi_\theta \), is given by
\[ P^{\alpha}_{\theta}(y, dz) \triangleq \mathbb{E}_{y,1,\theta}^{\alpha} \left[ \delta_{x^1_1}(dz) \right]. \] (33)
The family of Markov chains with transition kernels of the form \( P_\theta^{(N)}(y, dz) \) will be called i-c\( \alpha \)SMC processes.

3. Summary of Results

We now provide a summary of our results concerning SMC for sampling directly and as a component of an MCMC algorithm.

3.1. SMC for Sampling. The primary focus of our study will be on the expected value of the SMC estimators defined in Section 2.3.

**Definition 3.1.** If \( \pi^N \) is an SMC estimator of \( \pi \), then the expected estimator is \( \bar{\pi}^N \triangleq \mathbb{E}^N[\pi^N] \), where for a measurable set \( A \), \( \mathbb{E}^N[\pi^N](A) \triangleq \mathbb{E}^N[\pi^N(A)] \).

To connect the expected estimator to the goal of sampling we note that the marginal distribution of a sample from \( \pi^N \) is \( \bar{\pi}^N \). To the best of our knowledge, there has been minimal investigation of expected SMC estimators, with [9, Chapter 8] a notable exception. For example, the bound

\[
\text{KL}(\bar{\pi}^R_t||\pi_t) \leq \frac{C}{N},
\]

(34)
can be extracted as a special case of a more general propagation-of-chaos result [9, Theorem 8.3.2]. Our interest will be in the other direction of the KL divergence, \( \text{KL}(\pi||\bar{\pi}^N) \). In a certain sense, the direction of the KL divergence we investigate is the more “natural,” since \( \text{KL}(\mu||\nu) \) is the expected number of additional bits required to encode samples from \( \mu \) when using a code for \( \nu \) instead [7]. In other words, it is the amount of information lost by using samples from \( \nu \) instead of samples from \( \mu \). Beyond this heuristic justification, however, we shall see that the quantities that arise when studying \( \text{KL}(\pi||\bar{\pi}^N) \) are intimately related to those appearing in the study of particle Gibbs and the related iterated conditional SMC algorithm.

3.1.1. Special Case: Importance Sampling. Before detailing our results in full generality, to provide intuition as to how well \( \bar{\pi}^N \) approximates \( \pi \) as \( N \) increases, we briefly consider the special case of importance sampling, which is equivalent to either SIS or SIR when \( t = 1 \). We write \( M = M_1, g = g_1, \pi = \pi_1, Z = Z_1, X = X_1, \ldots \)
\( \pi^{1,N} = \pi^{S,N}_1 \), and \( \tilde{\pi}^{1,N} = \tilde{\pi}^{S,N}_1 \). Fig. 1a gives an example of \( \pi \), \( M \), and \( g \) in the case that \( E = \mathbb{R} \) and \( \pi \) and \( M \) are densities with respect to Lebesgue measure. Fig. 1b shows an example of an importance sampling estimate with \( N = 20 \) particles. Fig. 2 shows the density of \( \bar{\pi}^{1,N}_I \), along with \( \pi \) and \( M \), for \( N = 4, 8, 16, 32 \) particles. Informally, for a “small” number of particles, \( \bar{\pi}^{1,N}_I \) is strongly distorted toward the proposal distribution \( M \): for \( N \) not too large, with non-trivial probability all the samples from \( M \) will be in a region of low \( \pi \)-probability. Hence, all the potentials (normalized by a factor of \( 1/Z \)) will be small (\( \ll 1 \)). But in order to form the probability measure \( \pi^{1,N}_I \) the sum of the weights is normalized to 1, creating a overweighting in regions of high \( M \)-probability. However, as \( N \) increases, the probability of producing a sample from \( M \) in a region of high \( \pi \)-probability (and thus with a large potential) increases, which induces a better approximation to \( \pi \).

Using KL divergence to quantify the distortion in \( \bar{\pi}^{1,N}_I \), we have the following:

**Theorem 3.2.**

\[
\text{KL}(\pi || \bar{\pi}^{1,N}_I) \leq \log \left( 1 + \mathbb{V}\left[ Z^{-1}g(X) \right] \right) \leq \frac{\mathbb{V}[Z^{-1}g(X)]}{N}. \tag{35}
\]

The variance \( \mathbb{V}[Z^{-1}g(X)] \) is, in fact, the \( \chi^2 \)-divergence from \( \pi \) to \( M \) since

\[
\mathbb{V}[Z^{-1}g(X)] = \int \left( Z^{-1}g(x) - 1 \right)^2 M(dx) \tag{36}
\]

\[
= \int \left( \frac{d\pi}{dM}(x) - 1 \right)^2 M(dx) \tag{37}
\]

\[
= d_{\chi^2}(\pi, M). \tag{38}
\]

A classical result for bounding the KL divergence in terms of the \( \chi^2 \) divergence can be recovered by taking \( N = 1 \). We then have \( \text{KL}(\pi || \bar{\pi}^{1,N}_I) = \text{KL}(\pi || M) \), and so Theorem 3.2 implies that

\[
\text{KL}(\pi || M) \leq \log \left( 1 + d_{\chi^2}(\pi, M) \right). \tag{39}
\]

### 3.1.2. Basic SMC Results.

Many of our results concerning \( \bar{\pi}^N \) can be seen as analogous to existing operator-perspective results, but from the measure perspective. To make the analogies transparent, we first give an operator-perspective result, followed by the comparable measure-perspective result.

A typical \( L^p \) bound for SIR states that, for any \( p \geq 1 \) and any \( \phi \in \mathcal{B}_b(E) \)

\[
\mathbb{E}^{R,N} \left[ \left| \pi^{R,N}_1(\phi) - \pi_{1,t}(\phi) \right|^{1/p} \right]^{1/p} \leq \frac{a(t)b(\phi)c_t}{\sqrt{N}}, \tag{40}
\]

where \( c_t \) is a constant that depends only on \( \{M_s, g_s\}_{s \in [t]} \). We will show that for SIS

\[
\text{KL}(\pi_{1,t} || \pi^{S,N}_1) \leq \frac{S_t}{N} \tag{41}
\]

and for SIR

\[
\text{KL}(\pi_{1,t} || \pi^{R,N}_1) \leq \frac{R_t}{N} + \Theta(N^{-2}), \tag{42}
\]

where \( S_t \) and \( R_t \) are constants depending only on \( \{M_s, g_s\}_{s \in [t]} \). All these bounds hold under very mild assumptions.
3.1.3. Adaptive SMC Results. Recall (cf. Section 2.3.3) that the effective sample size (ESS) of the $\alpha$SMC particle weights at time $t$ is

$$E_t^2 \triangleq \frac{\sum_{n=1}^{N}(\tilde{W}_n^t)^2}{\sum_{n=1}^{N}(\tilde{W}_n^t)^2}$$

and the so-called adaptive particle filter resamples particles when $E_t^2 < \zeta N$ for some fixed $\zeta \in (0, 1]$. There are myriad heuristic arguments for using the ESS criterion [20, 22] as well as some theoretical analyses of the behavior of adaptive resampling algorithms under a variety of technical assumptions [8, 10, 23]. Whiteley, Lee, and Heine [23] provided a rigorous justification for the use of ESS from the operator viewpoint. Roughly speaking, they showed that if the ESS is not allowed to fall below $\zeta N$, for a fixed parameter $\zeta \in (0, 1]$, then the SMC algorithm does in fact behave as if there are $\zeta N$ particles. More formally, for $\phi \in B_b(E)$ and $p \geq 1$, under appropriate regularity conditions,

$$\sup_{t \geq 1} E_t^2 \geq \zeta N \implies \sup_{t \geq 1} \mathbb{E}^{\alpha,N} \left[ |\tilde{\eta}^{\alpha,N}_t(\phi) - \eta_t(\phi)|^p \right]^{1/p} \leq \frac{a(p)b(\phi)c_\infty}{\sqrt{\zeta N}}.$$  

Comparing Eq. (44) to [9, Theorem 7.4.4], which states that

$$\sup_{t \geq 1} \mathbb{E}^{R,N} \left[ |\tilde{\eta}^{R,N}_t(\phi) - \eta_t(\phi)|^p \right]^{1/p} \leq \frac{a(p)b(\phi)c_\infty}{\sqrt{N}},$$

we see that the condition $\sup_{t \geq 1} E_t^2 \geq \zeta N$ ensures that the effective number of particles in the time-uniform $L^p$ error bound is $\zeta N$ compared to $N$ particles if SIR is used. So in this technical sense ESS is a measure of the effective sample size.

We show that a different notion of ESS leads to similar results for $\alpha$SMC from the measure perspective, though we focus on fixed-time bounds, whereas the bounds
of [23] are time-uniform. We define the class of $p$-ESS measures, where $p \in [1, \infty]$ is the parameter of the ESS measure. For $p \in (1, \infty]$, the $p$-ESS is defined to be

$$E^p_t \triangleq \left( \frac{\|\dot{W}_t\|_1}{\|W_t\|_p} \right)^{p/(p-1)},$$

(46)

where if $p = \infty$, then $p/(p-1) \triangleq 1$. The standard ESS given by Eq. (43) corresponds to 2-ESS. We prove that, under standard regularity conditions,

$$\sup_{s \in [t]} E_s^\infty \geq \zeta N \implies KL(\pi_{1,t} || \bar{\pi}_{1,N}^{\alpha,N}) \leq \frac{\mathcal{R}_s'}{\zeta N} + \Theta(N^{-2}),$$

(47)

where $\mathcal{R}_s'$ is a constant depending only on $\{M_s, g_s\}_{s \in [t]}$.

3.2. Sampling with PG and i-cSMC. In [2], conditions are given under which the i-cSMC process is uniformly ergodic and the PG sampler is geometrically ergodic. Specifically, under regularity conditions, there exists $\rho_{t,N} = O(1/N)$ such that for all $y \in Y$ and for $P = P_\theta^{R,N}$,

$$d_{TV}(\delta_y P^k, \pi_\theta) \leq \rho_{t,N}^k,$$

(48)

and the PG sampler is geometrically ergodic as soon as the Gibbs sampler is geometrically ergodic. Furthermore, under a mixing-type condition, for any $C > 0$, if the number of particles is chosen to be $N = Ct$, then there exists $0 < \rho_C < 1$ such that $\sup_{t \geq 1} \rho_{t,N} < \rho_C$.

We give similar results for the $\alpha$-SMC process and the $\alpha$PG sampler. We show that under appropriate regularity conditions, there exists $\rho_{t,N} = O(1/N)$ such that for all $y \in Y$ and for $P = P_\theta^{\alpha,N}$,

$$\sup_{s \in [t]} E_s^\infty \geq \zeta N \implies \begin{cases} d_{TV}(\delta_y P^k, \pi_\theta) \leq \rho_{t,N}^k \\ \text{and} \\ \text{the $\alpha$PG sampler is geometrically ergodic} \\ \text{as soon as the Gibbs sampler is}. \end{cases}$$

(49)

Furthermore, under a mixing-type condition and a regularity condition on the matrices $\alpha \in \mathcal{A}_N$, for any $C > 0$, if $N = Ct$, then there exists $0 < \rho_C < 1$ such that $\sup_{t \geq 1} \rho_{t,N} < \rho_C$. As in sampling with $\alpha$SMC, maintaining a lower bound on the $\infty$-ESS is an important ingredient to the proof guaranteeing the convergence of the i-$\alpha$SMC process and the $\alpha$PG sampler.

4. Basic Results

In this section we give convergence rates of KL divergences for $\bar{\pi}_{1,t}^{S,N}$ and $\bar{\pi}_{1,t}^{R,N}$.

4.1. Convergence Rates. The key quantities in this section are

$$S_t \triangleq \mathbb{V} [Z_t^{-1} g_{1,t}(X_{1,t})], \quad \mathcal{R}_t \triangleq Z_t^{-1} \sum_{s=1}^t G_{0,s} \pi_{1,t}(G_{s,t+1}) - t,$$

$$\mathcal{G}_t \triangleq Z_t^{-1} \int \mathbb{E}_{y_{1,t},1}[\bar{Z}_t] \pi_{1,t}(dy_{1,t}).$$

(50)
Recall that for \( x_s \in E \), \( G_{s,t}(x_s) = \mathbb{E}[g_{s,t-1}(X_{s,t-1}) \mid X_s = x_s] \). For \( 1 \leq \ell \leq s \leq t \), let
\[
\mathcal{T}_{\ell,s} \triangleq \{ (\tau_1, \ldots, \tau_\ell) : t - s + 1 < \tau_1 < \cdots < \tau_\ell = t + 1 \} \quad (51)
\]
and, for \( \tau \in \mathcal{T}_{\ell,s} \), define
\[
C_\ell(\tau, y_1, t) \triangleq \ell^{-1} \prod_{i=1}^{\ell} G_{\tau_i, \tau_i+1}(y_{\tau_i}). \quad (52)
\]
We will sometimes write \( C_\ell(y) \) or \( C_\ell(\tau, y_1, t) \) instead of \( C_\ell(\tau, y_1, t) \). We will show (Proposition 4.8) that
\[
G_t = Z^{-1} \sum_{\ell=1}^{t} \sum_{\tau \in \mathcal{T}_{\ell,t+1}} G_{0,\tau_1,\pi_1,t}(C_\ell(\tau)). \quad (53)
\]
Hence, \( G_t \) is related to \( R_t \):
\[
G_t = \frac{(N-1)^t}{N^t} \sum_{\ell=1}^{t-1} \ell^{+1-\ell} \sum_{\tau \in \mathcal{T}_{\ell+1,t}} G_{0,\tau_1,\pi_1,t}(C_\ell(\tau)) + \Theta(N^{-2}) \quad (54)
\]
\[
= 1 + \frac{R_t}{N} + \Theta(N^{-2}). \quad (55)
\]

**Theorem 4.1.** For SIS,
\[
\text{KL}(\pi_{1,t} \mid \pi_{1,t}^{S,N}) \leq \log \left(1 + \frac{S_t}{N} \right) \leq \frac{S_t}{N} \quad (56)
\]
and for SIR,
\[
\text{KL}(\pi_{1,t} \mid \pi_{1,t}^{R,N}) \leq \log G_t = \log \left(1 + \frac{R_t}{N} + \Theta(N^{-2}) \right) \leq \frac{R_t}{N} + \Theta(N^{-2}). \quad (57)
\]

Pinsker’s inequality can be used to bound the total variation distance, though the SIR convergence rate is not optimal since, as [9] shows, \( d_{TV}(\pi, \pi_{R,N}) = O(1/N) \):

**Corollary 4.2.**
\[
d_{TV}(\pi, \pi_{S,N}) \leq \sqrt{\frac{1}{2} \log \left(1 + \frac{S_t}{N} \right)} \leq \frac{S_t}{2N}.
\]
and
\[
d_{TV}(\pi, \pi_{R,N}) \leq \sqrt{\frac{1}{2} \log \left(1 + \frac{R_t}{N} + \Theta(N^{-2}) \right)} \leq \sqrt{\frac{R_t}{2N}} + \Theta(N^{-1}).
\]

The following technical lemma will repeatedly prove useful:

**Lemma 4.3.** Let \( X \) and \( Y \) be random elements in Borel spaces \( (S,S) \) and \( (T,T) \), respectively, let \( \psi : S \times T \to \mathbb{R}_+ \) be measurable, and let \( \mu \) be the distribution of \( X \). If
\[
\nu = \mathbb{E}[\psi(X,Y) \delta_X],
\]
then \( \nu \ll \mu \) and
\[
\frac{d\nu}{d\mu}(X) = \mathbb{E}[\psi(X,Y) \mid X] \text{ a.s.} \quad (59)
\]
Proof. Because $S$ is Borel, there exists an $f$ satisfying $f(X) = \mathbb{E}[\psi(X, Y) \mid X]$ a.s. It follows from the chain rule of conditional expectation and then some elementary manipulations that, for all $A \in \mathcal{S}$,

$$
\nu(A) = \mathbb{E}[f(X)\delta_X(A)] = \mathbb{E}[f(X)1_A(X)] = \int_A f(x)\mu(dx),
$$

and so $f$ is a version of the Radon-Nikodym derivative $d\nu/d\mu$.

\[ \Box \]

**Proposition 4.4.** For the SIS algorithm, $\bar{\pi}^{S,N}_{1,t} \ll \pi_{1,t}$ and

\[
\frac{d\bar{\pi}^{S,N}_{1,t}}{d\pi_{1,t}}(y_{1,t}) = \mathbb{E}^{S,N}\left[\frac{Z_t}{Z_t} \mid \tilde{X}_{1,t}^1 \equiv y_{1,t}\right].
\]

Although Proposition 4.6 (and Proposition 4.4 below) follow as special cases of Proposition 5.1, the proofs of these results contain key ideas in simplified form, so we have included them.

**Proof.** We have

\[
\bar{\pi}^{S,N}_{1,t} = \mathbb{E}^{S,N}\left[\sum_{n=1}^{N} \frac{g_{1,t}(X^n_{1,t})}{\sum_{k=1}^{N} g_{1,t}(X^k_{1,t})} \delta_{X^n_{1,t}}\right]
\]

and so, by Lemma 4.3 and the definition of $\hat{Z}_t$,

\[
\frac{d\bar{\pi}^{S,N}_{1,t}}{dM_{1,t}}(y_{1,t}) = g_{1,t}(y_{1,t}) \cdot \mathbb{E}^{S,N}\left[\frac{1}{Z_t} \mid \tilde{X}_{1,t}^1 \equiv y_{1,t}\right],
\]

where $M_{1,t} \triangleq M_1M_2 \cdots M_t$. Upon noting that $\frac{d\pi_{1,t}}{dM_{1,t}}(y_{1,t}) = g_{1,t}(y_{1,t})/Z_t$, the result follows.

\[ \Box \]

**Lemma 4.5.** $\pi_{1,t}$ and $\bar{\pi}^{S,N}_{1,t}$ are absolutely continuous with respect to each other.

**Proof.** By Proposition 4.4, it suffices to show that $\pi_{1,t} \ll \bar{\pi}^{S,N}_{1,t}$. Note that for all $A \in \mathcal{B}(E')$, $\bar{\pi}^{S,N}_{1,t}(A) = 0 \implies$ there exists $B \subset A$ such that $M_{1,t}(B) = 0$ and $\forall y_{1,t} \in A \setminus B, g_{1,t}(y_{1,t}) = 0$. But since $\pi_{1,t} \ll M_{1,t}$, $M_{1,t}(B) = 0 \implies \pi_{1,t}(B) = 0$ and since for $y_{1,t} \in A \setminus B$, $g_{1,t}(y_{1,t}) = 0$, $\pi_{1,t}(A \setminus B) = 0$ as well. So $\pi_{1,t}(A) = 0$.

**Proof of Theorem 4.1 (SIS portion).** By Proposition 4.4 and Jensen’s inequality

\[
\frac{d\bar{\pi}^{S,N}_{1,t}}{d\pi_{1,t}}(y_{1,t}) = \mathbb{E}^{S,N}\left[\frac{Z_t}{Z_t} \mid \tilde{X}_{1,t}^1 \equiv y_{1,t}\right]
\]

\[
\geq \mathbb{E}^{S,N}\left[\sum_{k=1}^{N} Z_t^{-1} g_{1,t}(X^k_{1,t}) \mid \tilde{X}_{1,t}^1 \equiv y_{1,t}\right]
\]

\[
= \frac{N}{N - 1 + Z_t^{-1} g_{1,t}(y_{1,t})}.
\]
Lemma 4.5 implies that $\frac{d\pi_{1,t}}{d\pi_{1,t}} = (\frac{d\pi_{1,t}^{S,N}}{d\pi_{1,t}^{S,N}})^{-1}$, which together with Jensen’s inequality yields

$$\text{KL}(\pi_{1,t} || \pi_{1,t}^{S,N}) = \pi_{1,t} \left( \log \frac{d\pi_{1,t}}{d\pi_{1,t}^{S,N}} \right) \leq \log \pi_{1,t} \left( \frac{d\pi_{1,t}^{S,N}}{d\pi_{1,t}^{S,N}} \right)$$

(67)

$$= \log M_{1,t} \frac{d\pi_{1,t}}{d\pi_{1,t}^{S,N}} \frac{d\pi_{1,t}}{dM_{1,t}}$$

(68)

$$\leq \log M_{1,t} \left( \frac{N - 1 + Z_{1,t}^{-1} g_{1,t} Z_{1,t}^{-1} g_{1,t}}{N} \right)^{-1}$$

(69)

$$= \log \left( 1 + \frac{M_{1,t}(Z_{1,t}^{-1} g_{1,t})^2 - 1}{N} \right)$$

(70)

$$= \log \left( 1 + \frac{\mathbb{V}[Z_{1,t}^{-1} g_{1,t} X_{1,t}]}{N} \right).$$

(71)

To prove the SIR portion of Theorem 4.1 a few additional definitions are needed. We can write the joint density of SIR as

$$\psi(x_{1,t}, a_{1,t-1}) \triangleq \left( \prod_{n=1}^{N} M_{1}(x_{1}^{n,t}) \right) \prod_{s=2}^{t} \left( r(a_{s-1} | w_{s-1}) \prod_{n=1}^{N} M_{s}(x_{s-1}^{n,s-1}, x_{s}^{n}) \right),$$

(72)

where

$$w_{s}^{n} \triangleq w_{s}^{n}(x_{s}) \triangleq \frac{g_{s}(x_{1}^{n,t})}{\sum_{k=1}^{M} g_{s}(x_{1}^{k,t})}$$

(73)

and the carrier measure is implicit. For trajectory $y_{1,t} \in E^{t}$ and lineage $k_{1,t} \in [N]^{t}$, we can write the density of the cSMC process with law $\pi_{y_{1,t}, k_{1,t}, (X_{1,t}, A_{1,t-1})}$ as

$$\tilde{\psi}_{y_{1,t}, k_{1,t}}(x_{1,t}, a_{1,t-1}) = \prod_{s=1}^{t} \mathbb{I}(a_{s} = k_{s}) \mathbb{I}(y_{s}^{k_{s}} = x_{s}^{k_{s}}) \left( \prod_{n \neq k_{1}^{t}}^{N} M_{1}(x_{1}^{n,t}) \right)$$

$$\times \prod_{s=2}^{t} \left( r(a_{s-1} | w_{s-1}) \prod_{n \neq k_{1}^{t}}^{N} M_{s}(x_{s-1}^{n,s-1}, x_{s}^{n}) \right).$$

(74)

Let $b_{k}^{t} \triangleq k$ and for $s = t - 1, \ldots, 1$, $b_{s}^{k} \triangleq a_{s+1}^{k}$, so $x_{1,t}^{k} = (x_{1}^{k_{1}}, x_{2}^{k_{2}}, \ldots, x_{t}^{k_{t}})$. Furthermore, if $k = k_{1}^{t}$, then $k_{s}^{1} = b_{s}^{k}$ and, by implicitly identifying $x_{1,t}^{k}$ with $y_{1,t}$ and $b_{t}^{k}$ with $k_{1}^{t}$, we can rewrite the cSMC density in the “collapsed” form

$$\tilde{\psi}(k, x_{1,t}, a_{1,t-1}) \triangleq \left( \prod_{n \neq b_{1}^{t}}^{N} M_{1}(x_{1}^{n,t}) \right) \prod_{s=2}^{t} \left( r(a_{s-1} | w_{s-1}) \prod_{n \neq b_{1}^{t}}^{N} M_{s}(x_{s-1}^{n,s-1}, x_{s}^{n}) \right)$$

$$= \frac{\psi(x_{1,t}, a_{1,t-1})}{M_{1}(x_{1}^{b_{1}^{t}})} \prod_{s=2}^{t} r(b_{s-1}^{k_{s-1}} | w_{s-1}) M_{s}(x_{s-1}^{b_{s-1}^{k_{s-1}}}, x_{s}^{b_{s}^{k}}).$$

(75)
Proof. Consider the density
\[ \pi_{1,t}(k, x_{1:t}, a_{1:t-1}) = \pi_{1,t}(x_{1:t}) \psi(k, x_{1:t}, a_{1:t-1}). \]

Then (similarly to [1, Theorem 2])
\begin{align*}
\psi(x_{1:t}, a_{1:t-1}) w_t^k & = \frac{w_t^k M_1(x_{1:t}) \prod_{s=2}^{t} \text{mult}(b_{s-1}^k | w_{s-1}) M_s(x_{s-1:t}, b_s^k)}{N^{-1} \pi_{1,t}(x_{1:t})} \\
& = \frac{M_1(x_{1:t}) \prod_{s=2}^{t} M_s(x_{s-1:t}, x_s) \prod_{s=1}^{t} w_s b_s}{N^{-1} \pi_{1,t}(x_{1:t})} \\
& = \frac{M_1(x_{1:t}) \prod_{s=2}^{t} M_s(x_{s-1:t}, x_s) \prod_{s=1}^{t} g_s(x_{1:s})}{\pi_{1,t}(x_{1:t}) N^{-t} \prod_{s=1}^{t} \sum_{n=1}^{N} g_s(x_{1:s})} \\
& = \frac{Z_t}{Z_t(x_{1:t}, a_{1:t-1})},
\end{align*}

where
\[ Z_t(x_{1:t}, a_{1:t-1}) = E_{R,N} \left[ \tilde{Z}_t | \tilde{X}_{1:t} = x_{1:t}, A_{1:t-1} = a_{1:t-1} \right]. \]

We thus have
\[ \tilde{\pi}_{1,t}^{R,N}(dy_{1:t}) \]
\begin{align*}
& = \sum_{a_{1:t-1}, k} \int \psi(x_{1:t}, a_{1:t-1}) w_t^k \delta_{x_{1:t}}(dy_{1:t}) dx_{1:t} \\
& = \sum_{a_{1:t-1}, k} \int \psi(x_{1:t}, a_{1:t-1}) w_t^k \tilde{\pi}_{1,t}(k, x_{1:t}, a_{1:t-1}) \delta_{x_{1:t}}(dy_{1:t}) dx_{1:t} \\
& = \sum_{a_{1:t-1}, k} \left\{ \frac{Z_t}{Z_t(x_{1:t}, a_{1:t-1})} \tilde{\pi}_{1,t}(k, x_{1:t}, a_{1:t-1}) \delta_{x_{1:t}}(dy_{1:t}) \right\} \\
& = \left\{ \frac{N^{-t+1}}{Z_t(x_{1:t}, a_{1:t-1})} \tilde{\psi}(1, x_{1:t}, a_{1:t-1}) \delta_{y_{1:t}}(dy_{1:t}) \right\} \pi_{1,t}(dy_{1:t}).
\end{align*}

The result follows from Lemma 4.3. □

Definition 4.7. A collection of lineages \( k^1, \ldots, k^t \in [N]^t \) are said to be distinct when \( k^s \neq k^{s'} \) for all \( s \in [t] \) and \( j, j' \in [t] \) where \( j \neq j' \).

The next result is a straightforward generalization of [2, Proposition 9].
Proposition 4.8. For all \( t \geq 1, i \geq 1, N \geq i, y_{1,i}, \ldots, y_{1,i} \in E^1 \), and distinct lineages \( k_{1,i} \),

\[
\mathbb{E}_{y_{1,i}, k_{1,i}}^{R,N}[\tilde{Z}_t] = \frac{1}{N^i} \sum_{\ell=1}^{t+1} (N-i)^{t+1-\ell} \sum_{\tau \in \mathcal{T}_{t+1}} G_{0,\tau_1} \prod_{m=1}^{\ell-1} \sum_{j=1}^{i} G_{\tau_m,\tau_{m+1}}(y_{\tau_m}^j). \tag{87}
\]

In particular, for \( y_{1,i} \in E^i \),

\[
\mathbb{E}_{y_{1,i}}^{R,N}[\tilde{Z}_t] = \frac{1}{N^i} \sum_{\ell=1}^{t+1} (N-i)^{t+1-\ell} \sum_{\tau \in \mathcal{T}_{t+1}} G_{0,\tau_1} \ell C_\ell(\tau, y_{1,i}). \tag{88}
\]

Lemma 4.9. \( \pi_{1,i} \) and \( \pi_{1,i}^{R,N} \) are absolutely continuous with respect to each other.

Proof. The reasoning is analogous to the proof of Lemma 4.5. \( \square \)

Proof of Theorem 4.1 (SIR portion). The SIR bound follows from Propositions 4.6 and 4.8, Lemma 4.9, and Jensen’s inequality. \( \square \)

Remark 4.1. As one should expect in the case of \( g_s = 1 \) for \( s \in [t] \),

\[
\text{KL}(\pi_{1,i}||\pi_{1,i}^{S,N}) = \text{KL}(\pi_{1,i}||\pi_{1,i}^{R,N}) = 0,
\]

and so SIS and SIR are equivalent from the measure perspective. Indeed, from the measure perspective, all that is required is a single sample from the chain \( (X_t)_{t \geq 1} \). However, when \( \pi_{1,i}^{S,N} \) and \( \pi_{1,i}^{R,N} \) are used as estimators, \( \pi_{1,i}^{S,N} \) is superior to \( \pi_{1,i}^{R,N} \). Specifically, for \( \phi \in B_b(E^1) \), SIS produces \( N \) independent samples, so

\[
\mathbb{V}[\pi_{1,i}^{S,N}(\phi)] = \frac{\mathbb{V}[\phi(X_{1,i})]}{N}.
\]

For simplicity, consider a version of \( \pi_{1,i}^{R,N} \) obtained by first generating \( N \) samples from \( \pi_{1,i} \), then applying multinomial resampling to obtain the final \( N \) samples. In this case it is easy to show that

\[
\mathbb{V}[\pi_{1,i}^{R,N}(\phi)] = \frac{(2N-1)\mathbb{V}[\phi(X_{1,i})]}{N^2} \approx \frac{2\mathbb{V}[\pi_{1,i}^{S,N}(\phi)]}{N},
\]

so SIS is superior to SIR from the operator perspective.

4.2. Quantitative Bounds. We can obtain explicit bounds in the SIR case using results from Andrieu, Lee, and Vihola [2] to derive bounds on the expected value of the normalization constant. To obtain bounds on the KL divergence from a bound on \( \mathbb{E}_{y_{1,i}}^{R,N}[\tilde{Z}_t] \) we use the following simple lemma, which follows immediately from the definition of \( \mathcal{G}_t \) and Theorem 4.1.

Lemma 4.10. If \( Z_t^{-1}\mathbb{E}_{y_{1,i}}^{R,N}[\tilde{Z}_t] \leq B_N \) for some \( B_N \) that does not depend on \( y_{1,i} \), then

\[
\text{KL}(\pi_{1,i}||\pi_{1,i}^{R,N}) \leq \log B_N.
\]

In the case of bounded potential functions, a bound is straightforward to obtain. However, it requires the number of particles \( N \) to grow exponentially in \( t \) in order to remain constant. Let \( \bar{g}_t \triangleq \sup_{x \in E} g_s(x) \).
Proposition 4.11. Assume that \( \overline{g}_s < \infty \) for all \( s \in [t] \). Then for all \( t \geq 1, i \geq 1, N \geq i, y_{1,t}^1, \ldots, y_{1,t}^i \in E^t \), and distinct lineages \( k_{1,t} \),
\[
\mathbb{E}_{y_{1,t}^i, k_{1,t}}[\tilde{Z}_t] \leq Z_t \left\{ 1 + \left[ 1 - (1 - i/N)^t \right] \left[ Z_t^{-1} \prod_{s=1}^t \overline{g}_s - 1 \right] \right\}.
\]
(89)

Proof. The proof is a straightforward generalization of that for [2, Proposition 12] with some additional bookkeeping for \( i \) (instead of 2) fixed trajectory lineages. □

Corollary 4.12. Assume that \( \overline{g}_s < \infty \) for all \( s \in [t] \). Then
\[
\text{KL}(\pi_{1,t} || \overline{\pi}_{R,t}^N) \leq \log \left\{ 1 + \left[ 1 - (1 - N^{-1})^t \right] Z_t^{-1} \prod_{s=1}^t \overline{g}_s - 1 \right\}
\]
\[
\leq \frac{t (Z_t^{-1} \prod_{s=1}^t \overline{g}_s - 1)}{N}.
\]

Proof. Combine Lemma [4.10] and Proposition [4.11] □

To obtain better dependence on \( t \), stronger assumptions on \( \{M_t, g_t\} \) are required.

Assumption 4.A. There exists a constant \( \beta > 0 \) such that for any \( t, s \in \mathbb{N} \),
\[
\sup_{x \in E} \frac{G_{0,t} G_{t,t+s}(x)}{G_{0,t+s}} \leq \beta.
\]
(90)

Proposition 4.13. If Assumption [4.A] holds, then for all \( t \geq 1, i \geq 1, N \geq i, y_{1,t}^1, \ldots, y_{1,t}^i \in E^t \), and distinct lineages \( k_{1,t} \),
\[
\mathbb{E}_{y_{1,t}^i, k_{1,t}}[\tilde{Z}_t] \leq Z_t \left( 1 + \frac{i(\beta - 1)}{N} \right)^t.
\]
(90)

Proof. The proof is a simple generalization of that for [2, Proposition 14]. □

Corollary 4.14. If Assumption [4.A] holds, then
\[
\text{KL}(\pi_{1,t} || \overline{\pi}_{R,t}^N) \leq t \log \left( 1 + \frac{\beta - 1}{N} \right) \leq \frac{(\beta - 1)t}{N}.
\]

Proof. Combine Lemma [4.10] and Proposition [4.13] □

Assumption [4.A] is implied by a standard “strong mixing” condition which is often employed in SMC analyses (e.g., [9, 23]).

Assumption 4.B. There exists some \( m \geq 1 \) such that:
(a) for some \( 1 \leq \gamma < \infty \), for any \( t \geq 1 \) and \( x, x' \in E \) and \( A \in \mathcal{E} \),
\[
M_{t,t+m}(x, A) \leq \gamma M_{t,t+m}(x', A),
\]
where \( M_{t,t+m} \triangleq M_t M_{t+1} \cdots M_{t+m} \);
(b) for some \( 1 \leq \delta < \infty \), the potentials satisfy
\[
\sup_{x, x' \in E} \frac{g_t(x)}{g_t(x')} \leq \delta^{1/m}.
\]
Lemma 4.15 ([2, Lemma 17]). If Assumption 4.B holds, then for all $t, s \geq 1$,
\[
\sup_{x, x' \in E} \frac{G_{t, t+s}(x)}{G_{t, t+s}(x')} \leq \gamma \delta,
\]
so Assumption 4.A holds.

5. SMC with Adaptive Resampling

We now turn to the $\alpha$SMC algorithmic framework to investigate the convergence properties of SMC with adaptive resampling. As summarized earlier, Whiteley, Lee, and Heine [23] show that maintaining a lower bound of $E_t^2 \geq \zeta N$ guarantees time-uniform convergence of $\alpha$SMC in the $L^p$ sense at an $O(\frac{1}{N})$ rate. We will prove an analogous result from the measure perspective using a quantity $E_t^\infty$ we call the infinite-ESS. We show that maintaining a lower bound of $E_t^\infty \geq \zeta N$ suffices to guarantee fixed-time convergence of $\alpha$SMC in KL divergence at an $O(\frac{1}{N})$ rate. Note that our result is weaker than that of Whiteley, Lee, and Heine [23] in the sense that theirs is time-uniform while ours is only for a fixed time. We will show that $E_t^\infty$ is a stricter notion of effective sample size than $E_t^2$.

We can write the joint density of the $\alpha$SMC sampler as
\[
\psi^\alpha(x_{1:t}, a_{1:t-1}) \triangleq \left( \prod_{n=1}^{N} M_1(x_1^n) \right) \left( \prod_{s=2}^{T} \prod_{n=1}^{N} r_n(a_{s-1}^n | x_{s-1}, x_1, \ldots, x_{s-2}) M_s(x_{s-1}^{a_{s-1}^n}, x_s^n) \right), \tag{91}
\]
where
\[
r_n(k | w_{t-1}, x_{1:t-1}) \triangleq \frac{\alpha_{t-1}^n w_{t-1}^k g_{t-1}(x_{t-1}^k)}{w_t^n}, \tag{92}
\]
We will work under the following assumption:

**Assumption 5.C.** For all $N$, all $\alpha \in \mathbb{A}_N$ are doubly stochastic.

**Remark 5.1.** Assumption 5.C is the same as Assumption (B++) in [23], though there it is stated as assuming each $\alpha$ admits the uniform distribution as an invariant measure.

We begin by noting that, under Assumption 5.C, the density of the $\alpha$SMC process with law $P_{t, t}^{\alpha,N}(X_{1:t}, A_{1:t-1}, F_{1:t})$ can be written in the following “collapsed” form, by implicitly identifying $x_t^{a_t}$ with $y_{1:t}$ (cf. Eq. (75)):
\[
\tilde{\psi}^\alpha(x_{1:t}, a_{1:t-1}, f_{1:t})
= \psi^\alpha(x_{1:t}, a_{1:t-1}) \prod_{s=2}^{t} I_s a_{s-1}^{f_s f_{s-1}}
\frac{NM_1(x_1^n) \prod_{s=2}^{s} f_{s-1} | w_{s-1}, x_{1:s-1}) M_s(x_{s-1}^{a_{s-1}^n}, x_s^n)}{NM_1(x_1^n) \prod_{s=2}^{s} f_{s-1} | w_{s-1}, x_{1:s-1}) M_s(x_{s-1}^{a_{s-1}^n}, x_s^n)}
\frac{1}{N} \prod_{n \neq f_1} M_1(x_1^n) \prod_{s=2}^{t} I_s a_{s-1}^{f_s f_{s-1}} \prod_{n \neq f_s} r_n(a_{s-1} | w_{s-1}, x_{1:s-1}) M_s(x_{s-1}^{a_{s-1}^n}, x_s^n), \tag{93}
\]
where $I_s \triangleq 1(a_{s-1}^{f_s} = f_{s-1})$. Let $\mathcal{F}_s$ be the $\sigma$-algebra generated by $X_{1:s}, A_{1:s-1},$ and $F_{1:s}$, where by convention we let $\mathcal{F}_0$ be the trivial $\sigma$-algebra.
Proposition 5.1. If Assumption \textbf{5.C} holds, then the $\alpha$SMC estimator satisfies
\[
\frac{d\hat{\pi}^\alpha_{1,t}}{d\pi_{1,t}}(y_{1,t}) = \mathbb{E}_{y_{1,t}} \left[ \frac{Z_t}{Z_{t}} \right] \geq \frac{Z_t}{\mathbb{E}_{y_{1,t}}[Z_t]} \tag{94}
\]

Proof. Consider the density
\[
\hat{\pi}^\alpha_{1,t}(x_{1,t}, a_{1,t-1}, f_{1,t}) \triangleq \pi_{1,t}(x_{1,t})\tilde{\psi}^\alpha(x_{1,t}, a_{1,t-1}, f_{1,t}).
\]

Then
\[
\begin{align*}
\hat{\pi}^\alpha_{1,t}(x_{1,t}, a_{1,t-1}, f_{1,t}) g_t(x^f_t) w_t^f &= M_1(x^f_1) \prod_{s=2}^{t} f_{s-1}(x_{s-1}, x_{s-1}) \mathcal{M}_s(x_{s-1}^f, x_{s}^f) g_t(x^f_t) w_t^f \\
&= \frac{\pi_{1,t}(x^f_1) \prod_{s=2}^{t} I_{s}\alpha_{s-1}^{f_{s-1}} N^{-1} \sum_{n=1}^{N} g_t(x^n_t) w^n_t}{\pi_{1,t}(x^f_1) \prod_{s=2}^{t} I_{s}\alpha_{s-1}^{f_{s-1}} N^{-1} \sum_{n=1}^{N} g_t(x^n_t) w^n_t} \\
&= \frac{1}{\tilde{Z}_t(x_{1,t}, a_{1,t-1}) \prod_{s=2}^{t} I_{s}},
\end{align*}
\]

where
\[
\tilde{Z}_t(x_{1,t}, a_{1,t-1}) \triangleq \mathbb{E}_{y_{1,t}}[\tilde{Z}_t | \tilde{X}_{1,t} = x_{1,t}, A_{1,t-1} = a_{1,t-1}].
\]

Using the convention that $0/0 = 0$, it follows that
\[
\begin{align*}
\hat{\pi}^\alpha_{1,t} (dy_{1,t}) &= \sum_{a_{1,t-1}, f_{1,t}} \left\{ \psi^\alpha(x_{1,t}, a_{1,t-1}) g_t(x^a_t) w^a_t \frac{\delta_{x^a_t}}{\pi^a_{1,t}} (dy_{1,t}) \right\} \\
&= \sum_{a_{1,t-1}, f_{1,t}} \left\{ \frac{\psi^\alpha(x_{1,t}, a_{1,t-1}) g_t(x^f_t) w^f_t}{\pi^f_{1,t}(x_{1,t}, a_{1,t-1}, f_{1,t})} \sum_{n=1}^{N} g_t(x^n_t) w^n_t \right\} \\
&\quad \times \hat{\pi}^\alpha_{1,t}(x_{1,t}, a_{1,t-1}, f_{1,t}) \delta_{x^f_t}(dy_{1,t}) \\
&= \sum_{a_{1,t-1}, f_{1,t}} \left\{ \frac{Z_t}{\tilde{Z}_t(x_{1,t}, a_{1,t-1})} \hat{\pi}^\alpha_{1,t}(x_{1,t}, a_{1,t-1}, f_{1,t}) \delta_{x^f_t}(dy_{1,t}) \right\} \\
&= \left\{ \sum_{a_{1,t-1}, f_{1,t}} \frac{Z_t}{\tilde{Z}_t(x_{1,t}, a_{1,t-1})} \tilde{\psi}^\alpha(k, x_{1,t}, a_{1,t-1}) \delta_{x^f_t}(dx^f_t) \right\} \pi_{1,t}(dy_{1,t}).
\end{align*}
\]

The result follows from Lemma \textbf{4.3}. \hfill \qed

With Proposition \textbf{5.1} in hand, our next task is to understand the quantity $\mathbb{E}_{y_{1,t}}[\tilde{Z}_t]$. To do so, we will require a generalized notion of effective sample size, which includes two commonly used definitions as special cases.
Definition 5.2. For parameter \( p \in [1, \infty] \), let \( p^* = \frac{p}{p-1} \) be the conjugate exponent of \( p \) (so \( 1/p + 1/p^* = 1 \)). The \( p \)-effective sample size (\( p \)-ESS) at time \( t \) is

\[
\mathcal{E}_t^p \triangleq \begin{cases} 
\left( \frac{\|W_t\|_1}{\|W_t\|_p} \right)^{p^*} & p > 1 \\
\prod_{n=1}^{\infty} (W_n^t)^{p^*/p} / \|W_t\|_1 & p = 1.
\end{cases}
\]

(105)

The following proposition catalogs some elementary properties of \( p \)-ESS (see Appendix A.1 for a proof).

Proposition 5.3. The \( p \)-ESS has the following properties:

1. The 1-ESS satisfies
   \[
   \mathcal{E}_t^1 = \lim_{p \downarrow 1} \mathcal{E}_t^p = \mathcal{E}_t^{ent} \triangleq e^H(W_t),
   \]
   where \( H(W) \triangleq -\sum_n W_n \log W_n / \|W\|_1 \).

2. For all \( p \in [1, \infty] \), \( 1 \leq \mathcal{E}_t^p \leq N \). The lower bound is achieved if and only if all but one of the weights is zero. The upper bound is achieved if and only if all the weights are equal.

3. For \( 1 \leq p < q \leq \infty \), \( \mathcal{E}_t^p \geq \mathcal{E}_t^q \geq N - (1 - q^*/p^*) \mathcal{E}_t^p \), with equality if and only if \( \mathcal{E}_t^p \) and \( \mathcal{E}_t^q \) equal 1 or \( N \).

Part (1) of Proposition 5.3 shows that the 1-ESS corresponds to the entropic ESS, which is a common choice of ESS in applications [6]. Note that 2-ESS corresponds to the standard definition of ESS. Part (2) formalizes the sense in which \( p \)-ESS measures effective sample size: it always takes on values between 1 and \( N \), and it takes on the value \( k \in [N] \) when \( k \) particles have weight \( 1/k \). Part (3) shows that the larger \( p \), the more stringent the notion of effective sample size \( p \)-ESS is.

In order to prove our convergence result for \( \alpha \text{-SMC} \), we will require a lower bound on the \( \infty \)-ESS.

Assumption 5.D. There exists some \( 0 < \zeta \leq 1 \) such that for all \( s \in [t] \), \( \mathcal{E}_s^\infty \geq \zeta N \).

At a technical level, Whiteley, Lee, and Heine [23] uses the 2-ESS lower bound guarantee to bound the \( L_2 \) norm of the weights in terms of their \( L_1 \) norm. Similarly, we will use the \( \infty \)-ESS lower bound guarantee to bound the sup-norm of the weights in terms of their \( L_1 \) norm. Specifically, under Assumption 5.D,

\[
\zeta N \leq \mathcal{E}_s^\infty = \frac{\|W_s\|_1}{\|W_s\|_\infty} = \frac{\|W_s\|_1}{\sup_n W_n^s},
\]

so for all \( n \in [N] \) and \( s \in [t] \), \( W_n^s \leq \frac{\|W_s\|_1}{\zeta N} \).

The expectation \( \mathbb{E}_{\gamma_1,t}[\tilde{Z}_t] \) can be deconstructed in a similar manner to \( \mathbb{E}_{\gamma_1,t}[\tilde{Z}_t] \). Recall that for \( \tau \in \mathcal{T}_\ell \),

\[
C_\ell^y(\tau) = \prod_{i=1}^{\ell-1} C_{\tau_i,\tau_{i+1}}^y,
\]

(108)
Proposition 5.4. If Assumptions 5.C and 5.D hold, then for all $t \geq 1$, $i \geq 1$, $N \geq i$, $y_{1,t}, \ldots, y_{i,t} \in E^t$,

$$E_{y_{1,t}}[\hat{Z}_t] \leq \frac{1}{N(\zeta N)^{t+1}} \sum_{\ell=1}^{t+1} \sum_{\tau \in T_{\ell+1}} (\zeta N)^{t+1-\ell} \left( \frac{N-i}{\zeta N} \right)^{1/(\tau_1+1)} G_{0,\tau_1} \prod_{m=1}^{\ell-1} \sum_{j=1}^i G_{\tau_m,\tau_{m+1}}(y_{j,m}^t)$$

(109)

In particular, in the case of $i = 1$, we have

$$E_{y_{1,t}}[\hat{Z}_t] \leq \frac{1}{N(\zeta N)^{t+1}} \sum_{\ell=1}^{t+1} (\zeta N)^{t+1-\ell} \left( \frac{N-1}{\zeta N} \right)^{1/(\tau_1+1)} G_{0,\tau_1} C^y_\tau(\tau)$$

(110)

See Appendix A.2 for a proof.

5.1. Quantitative Bounds. Proposition 5.4 leads to quantitative bounds very similar to those for SIR. Indeed, the SIR results are essentially a special case of the $\alpha$SMC results, though with slightly worse constants in the $\alpha$SMC case.

Theorem 5.5. If Assumptions 5.C and 5.D hold, then for $t \geq 1$,

$$KL(\pi_{1,t}||\hat{\pi}_{1,t}^{\alpha,N}) \leq Z_t^{-1} \sum_{s=1}^t G_{0,s} \pi_{1,t}(G_{s,t+1})-1 + \Theta(N^{-2}).$$

(111)

Proof. We have

$$E_{y_{1,t}}[\hat{Z}_t] \leq \frac{1}{N(\zeta N)^{t+1}} \sum_{\ell=1}^{t+1} \sum_{\tau \in T_{\ell+1}} (\zeta N)^{t+1-\ell} \left( \frac{N-1}{\zeta N} \right)^{1/(\tau_1+1)} G_{0,\tau_1} C^y_\tau(\tau)$$

(112)

$$= \frac{Z_t(N-1)}{N} + \frac{G_{1,t+1}^y}{N} + \frac{1}{\zeta N} \sum_{s=1}^t G_{0,s} G_{s,t+1}^y + \Theta(N^{-2})$$

(113)

$$\leq Z_t + \frac{\sum_{s=1}^t G_{0,s} G_{s,t+1}^y - Z_t}{\zeta N} + \Theta(N^{-2}).$$

(114)

The result then follows from Proposition 5.1.

It is now possible to give generalized versions of Propositions 4.11 and 4.13 and their corollaries:

Proposition 5.6. Fix any $t \geq 1$ and assume that for $s \in [t]$, $\overline{g}_s \triangleq \sup_{x \in E} g_s(x) < \infty$. Then if Assumptions 5.C and 5.D hold, for all $i \geq 1$, $N \geq i$, $y_{1,t}, \ldots, y_{i,t} \in E^t$,

$$E_{y_{1,t}}[\hat{Z}_t] \leq Z_t \left\{ 1 + Z_t^{-1} \prod_{s=1}^t \overline{g}_s \left[ \left( 1 + \frac{1}{\zeta N} \right)^t - 1 \right] \right\}$$

(115)

and therefore, for $N \geq \zeta^{-1}$,

$$KL(\pi_{1,t}||\hat{\pi}_{1,t}^{\alpha,N}) \leq \log \left\{ 1 + Z_t^{-1} \prod_{s=1}^t \overline{g}_s \left[ \left( 1 + \frac{1}{\zeta N} \right)^t - 1 \right] \right\}$$

(116)

$$\leq Z_t^{-1} \prod_{s=1}^t \overline{g}_s \left[ \left( 1 + \frac{1}{\zeta N} \right)^t - 1 \right]$$

(117)

$$\leq Z_t^{-1} \prod_{s=1}^t \overline{g}_s.$$
Proof. We have
\[
E_{y_{1,t}}^{\alpha,N}[\tilde{Z}_t] \leq \sum_{\ell=1}^{t+1} \sum_{\tau \in T_{\ell,t+1}} (\zeta N)^{-\ell+1} G_{0,\tau_1} C^{\ell}_\tau(\tau)
\]
(119)
\[
\leq Z_t + \prod_{s=1}^t g_s \sum_{\ell=1}^{t+1} \binom{t}{\ell-1} (\zeta N)^{t-1}
\]
(120)
\[
= Z_t + \prod_{s=1}^t g_s \sum_{\ell=1}^t \binom{t}{\ell} (\zeta N)^{t-1}
\]
(121)
\[
= Z_t + \prod_{s=1}^t g_s \left[ \left(1 + \frac{1}{\zeta N}\right)^t - 1 \right].
\]
(122)
\[
\square
\]

Proposition 5.7. If Assumptions 4.A, 5.C and 5.D hold, then for all \( t \geq 1, i \geq 1, N \geq i, y_{1,t}, \ldots, y_{i,t} \in E^t \),
\[
E_{y_{1,t}}^{\alpha,N}[\tilde{Z}_t] \leq Z_t \left(1 + \frac{\beta}{\zeta N}\right)^t
\]
(123)
and therefore
\[
\text{KL}(\pi_{1,t}||\bar{\pi}_{1,t}^{\alpha,N}) \leq t \log \left(1 + \frac{\beta}{\zeta N}\right) \leq \frac{\beta t}{\zeta N}.
\]
(124)

Proof. The proof mirrors that for [2, Proposition 14]. Observe that for \( s \in [t+1] \),
\[
G_{0,t+1} = G_{0,t+1} \bar{G}_{0,s} = G_{0,s} \eta_s(G_{s,t+1}),
\]
so we can write for \( \ell \in [t], \tau \in T_{\ell,t+1} \),
\[
Z_t = G_{0,t+1} \prod_{i=1}^{\ell-1} \eta_{\tau_i}(G_{\tau_{i,t+1}}).
\]
(125)
Combined with Assumption 4.A and writing \( \bar{G}_{s,t} \equiv \sup_{x \in E} G_{s,t}(x) \),
\[
\sum_{\ell=1}^{t+1} (\zeta N)^{-\ell+1} \sum_{\tau \in T_{\ell,t+1}} G_{0,\tau_1} \prod_{i=1}^{\ell-1} G^{y}_{\tau_{i,t+1}}
\]
(126)
\[
\leq Z_t + Z_t \sum_{\ell=2}^{t+1} (\zeta N)^{-\ell+1} \sum_{\tau \in T_{\ell,t+1}} G_{0,\tau_1} \prod_{i=1}^{\ell-1} \bar{G}_{\tau_{i,t+1}}
\]
(127)
\[
= Z_t \sum_{\ell=1}^{t+1} \binom{t}{\ell-1} (\zeta N)^{t-1} \beta^{t-1}
\]
(128)
\[
= Z_t \left(1 + \frac{\beta}{\zeta N}\right)^t.
\]
(129)
\[
\square
\]

6. CONVERGENCE IN THE MARGINAL AND PREDICTIVE CASES

So far we have focused on the convergence of \( \bar{\pi}_{1,t}^{S,N}, \bar{\pi}_{1,t}^{R,N} \) and \( \bar{\pi}_{1,t}^{\alpha,N} \), which are all expectations of estimators of \( \pi_{1,t} \). Very similar results can be derived for
the expected estimators of $\pi_t, \eta_{1,t}$, and $\eta$. The following result allows us to easily generalize all the quantitative bounds given so far to these marginal and/or predictive measures. Define the reverse probability kernels $\tilde{\pi}_{1,t-1}(x_t, dx_{1,t-1})$ and $\tilde{\eta}_{1,t-1}(x_t, dx_{1,t-1})$ such that

$$\pi_{1,t}(dx_{1,t}) = \pi_t(dx_t)\tilde{\pi}_{1,t-1}(x_t, dx_{1,t-1})$$

(130) and

$$\eta_{1,t}(dx_{1,t}) = \eta_t(dx_t)\tilde{\eta}_{1,t-1}(x_t, dx_{1,t-1}).$$

(131)

**Proposition 6.1.** If Assumption 5.C holds, then the expected $\alpha$SMC estimators satisfy

$$\frac{d\tilde{\pi}_{1,t,N}^\alpha}{d\pi_t}(y_t) = \int E_{y_{1,t}}^\alpha \left[ \frac{Z_t}{Z_t^\alpha} \right] \tilde{\pi}_{1,t-1}(y_t, dy_{1,t-1})$$

(132)

$$\frac{d\tilde{\pi}_{1,t,N}^\alpha}{d\eta_{1,t}}(y_t) = \int E_{y_{1,t}}^\alpha \left[ \frac{Z_t^\alpha}{Z_t} \right] \tilde{\eta}_{1,t-1}(y_t, dy_{1,t-1})$$

(133)

$$\frac{d\tilde{\eta}_{1,t,N}^\alpha}{d\eta_t}(y_t) = \int E_{y_{1,t}}^\alpha \left[ \frac{Z_t^\alpha}{Z_t} \right] \tilde{\eta}_{1,t-1}(y_t, dy_{1,t-1}).$$

(134)

**Proof.** For Eq. (132), by Proposition 5.1 we have

$$\tilde{\pi}_{1,t,(x_{1,t}, a_{1,t-1}, f_{1,t})} = \eta_{1,t}(x_{1,t})\tilde{\pi}_{1,t-1}(y_{1,t}, dy_{1,t-1})$$

(135)

$$\tilde{\pi}_{1,t} = \int E_{y_{1,t}}^\alpha \left[ \frac{Z_t}{Z_t^\alpha} \right] \tilde{\pi}_{1,t-1}(y_t, dy_{1,t-1})\pi_t(dy_t).$$

(136)

For Eq. (133), consider the density

$$\tilde{\eta}_{1,t}^\alpha(x_{1,t}, a_{1,t-1}, f_{1,t}) \triangleq \eta_{1,t}(x_{1,t})\tilde{\pi}_{1,t}(y_{1,t}, dy_{1,t}),$$

(137)

Then

$$\tilde{\eta}_{1,t}^\alpha(x_{1,t}, a_{1,t-1}, f_{1,t}) \sum_{n=1}^N w_t^n$$

$$= M_1(x_{1,t}) \prod_{s=2}^t \psi_{f_{s-1}, w_{s-1}, x_{1,s-1}, m_s(x_{s-1}^{f_{s-1}}, x_s^{f_s})} = M_1(x_{1,t}) \prod_{s=2}^t I_s \frac{\alpha^{f_{s-1}, 1}N-1}{\sum_{n=1}^N w_t^n}$$

(138)

$$= M_1(x_{1,t}) \prod_{s=2}^t \frac{\alpha^{f_{s-1}, 1}N-1}{\sum_{n=1}^N w_t^n} \prod_{s=2}^t I_s \frac{\alpha^{f_{s-1}, 1}N-1}{\sum_{n=1}^N w_t^n} \sum_{n=1}^N w_t^n$$

(139)

$$= M_1(x_{1,t}) \prod_{s=2}^t \frac{\alpha^{f_{s-1}, 1}N-1}{\sum_{n=1}^N w_t^n} \prod_{s=2}^t I_s \frac{\alpha^{f_{s-1}, 1}N-1}{\sum_{n=1}^N w_t^n} \sum_{n=1}^N w_t^n$$

(140)

$$= \frac{Z_t^\alpha}{Z_t^\alpha} \prod_{s=2}^t I_s \frac{\alpha^{f_{s-1}, 1}N-1}{\sum_{n=1}^N w_t^n} \prod_{s=2}^t I_s$$

(141)

where

$$\tilde{\eta}_{1,t}^\alpha(x_{1,t}, a_{1,t-1}) \triangleq E_{y_{1,t}}^\alpha \left[ \frac{Z_t^\alpha}{Z_t} \right] \tilde{\pi}_{1,t} = x_{1,t}, A_{1,t-1} = a_{1,t-1}. $$

(142)
We thus have

\[ \bar{\eta}_{1, t}^\alpha, \bar{N}(dy_{1, t}) \]

\[ = \sum_{a_1, t-1, a_t} \int \left\{ \frac{\psi^\alpha(x_1, t, a_1, t-1)w_i^{a_t}}{\sum_{n=1}^N w_i^n} \delta_{x_{1, t}}(dy_{1, t}) \right\} \]

\[ = \sum_{a_1, t-1, f_1, t} \int \left\{ \frac{\psi^\alpha(x_1, t, a_1, t-1)w_i^{f_t}}{\bar{\eta}_{1, t}^\alpha(x_1, t, a_1, t-1, f_1, t)} \sum_{n=1}^N w_i^n \right\} \]

\[ \times \bar{\eta}_{1, t}^\alpha(x_1, t, a_1, t-1, f_1, t) \delta_{x_{1, t}}(dy_{1, t}) \}

\[ = \sum_{a_1, t-1, f_1, t} \int \left\{ \frac{Z_i'}{Z_i'(x_1, t, a_1, t-1)} \bar{\eta}_{1, t}^\alpha(x_1, t, a_1, t-1, f_1, t) \delta_{x_{1, t}}(dy_{1, t}) \right\} \]

\[ = \left\{ \sum_{a_1, t-1, f_1, t} \int \frac{Z_i'}{Z_i'(x_1, t, a_1, t-1)} \bar{\eta}_{1, t}^\alpha(x_1, t, a_1, t-1, f_1, t) \delta_{x_{1, t}}(dy_{1, t}) \right\} \bar{\pi}_{1, t}(dy_{1, t}) \]

proving the first equality. The second equality follows since the estimator of the normalization constant can be written as

\[ \hat{Z}_t' = N^{-1} \sum_n W_t^n = N^{-1} \sum_n \alpha_t^{n_k} W_t^{k}g_{t-1}^k \]

\[ = N^{-1} \sum_k W_t^{k}g_{t-1}^k = \hat{Z}_t, \]

where the penultimate equality follows from Assumption 5.C.

Eq. (134) follows by the same argument used to prove Eq. (132), but using Eq. (133).

For \( \pi_t \), we have bounds identical to those for \( \pi_{1, t} \):

**Theorem 6.2.** For the SIS and SIR algorithms,

\[ \text{KL}(\pi_t || \bar{\pi}_t^{S,N}) \leq \log \left( 1 + \frac{S_t}{N} \right) \leq \frac{S_t}{N} \]  

and

\[ \text{KL}(\pi_t || \bar{\pi}_t^{R,N}) \leq \log G_t. \]

**Proof.** For SIS,

\[ \frac{d\pi_t^{S,N}}{d\pi_t}(y_{1, t}) = \int_{\psi^{S,N}} \left[ \frac{N}{\sum_{n=1}^N Z_t^{-1}g_{t}(X_{1, t}^n)} \right] X_{1, t}^1 = y_{1, t} \bar{\pi}_{1, t-1}(y_t, dy_{1, t-1}) \]

\[ \geq \frac{N}{N - 1 + \int Z_t^{-1}g_{t}(y_{1, t})\bar{\pi}_{1, t-1}(y_t, dy_{1, t-1})}. \]
Hence,
\[
\text{KL}(\pi_t || \bar{\pi}_t^{S,N}) \leq \int \log \left( \frac{N - 1 + \int Z_{t-1}^{-1} g_{t,1}(y_{1,t}) \bar{\pi}_{1,t-1}(y_{t}, dy_{t-1})}{N} \right) \pi_t(dy_t) \\
\leq \log \left( \frac{N - 1 + \int Z_{t-1}^{-1} g_{t,1}(y_{1,t}) \bar{\pi}_{1,t-1}(y_{t}, dy_{t-1})}{N} \pi_t(dy_t) \right) \\
= \log \left( 1 + \frac{\pi_{1,t}(Z_{t-1}^{-1} g_{1,t})}{N} - 1 \right) \\
= \log \left( 1 + \frac{S_t}{N} \right). 
\]

For SIR,
\[
\frac{d\bar{\pi}_t^{R,N}}{d\pi_t}(y_t) \geq \frac{Z_t}{\int \mathbb{E}_{y_{1,t}^{R,N}[Z_{t}]^{1}}(y_{t}, dy_{t-1})}. \tag{153}
\]
and the bound follows analogously to the SIS case. \qed

Using Proposition 5.1 and Theorem 6.2 together with the results from Sections 4.2 and 5.1, the following quantitative bounds follow immediately:

**Theorem 6.3.** For all \( t \geq 1 \),

1. if for \( s \in [t], \bar{g}_s < \infty \), then
   \[
   \text{KL}(\pi_t || \bar{\pi}_t^{R,N}) \leq \frac{t(Z_{t-1}^{-1} \prod_{s=1}^{t} \bar{g}_s - 1)}{N}, \tag{154}
   \]
   while if Assumptions 5.C and 5.D also hold, then for \( N \geq \zeta^{-1} \)
   \[
   \text{KL}(\pi_t || \bar{\pi}_t^{\alpha,N}) \leq \frac{Z_{t-1}^{-1} \prod_{s=1}^{t} \bar{g}_s}{\zeta N} \tag{155}
   \]
   and
   2. if Assumption 4.A holds, then
      \[
      \text{KL}(\pi_t || \bar{\pi}_t^{R,N}) \leq \frac{(\beta - 1)t}{N} \tag{156}
      \]
      while if Assumptions 5.C and 5.D also hold, then
      \[
      \text{KL}(\pi_t || \bar{\pi}_t^{\alpha,N}) \leq \frac{\beta t}{\zeta N} \tag{157}
      \]

**Remark 6.1.** From the operator perspective, \( \pi_t^{S,N}, \pi_t^{R,N} \), and \( \pi_t^{\alpha,N} \) generally approximate \( \pi_t \) far better than \( \pi_{1,t}, \pi_{1,t}^{R,N}, \) and \( \pi_{1,t}^{\alpha,N} \) approximate \( \pi_{1,t} \). It is quite natural for SMC to produce better estimates of the marginal expectation since, while both the marginal and joint estimators involve the same number of particles, the joint expectation involves an integral over a much higher-dimensional space. So it is somewhat surprising that the KL divergence bounds we obtain in the marginal case are identical to those in the full joint distribution case already considered. But in fact, intuition suggests that the KL divergence case will behave very differently from that of functional approximation. Since only a single sample is being drawn from \( \pi_{1,t}^{S,N}, \pi_{1,t}^{R,N}, \) or \( \pi_{1,t}^{\alpha,N} \), the quality of the full sample compared to the marginal sample does not suffer from the same “curse of dimensionality.”
Using Proposition 5.1 together with the results of Sections 4.2 and 5.1 we obtain similar results for estimators of the predictive measure $\eta_{1,t}$ to those for the estimators of $\pi_{1,t}$:

**Theorem 6.4.** For all $t \geq 1$,

1. if for $s \in [t-1], \bar{g}_s < \infty$, then

$$\text{KL}(\eta_{1,t}||\bar{\eta}_{R,N}^{1,t}) \leq \frac{(t-1)(Z_{t-1}^{t-1} \prod_{s=1}^{t-1} g_s - 1)}{N}, \quad (158)$$

while if Assumptions 5.C and 5.D also hold, then for $N \geq \zeta^{-1}$

$$\text{KL}(\eta_{1,t}||\bar{\eta}_{R,N}^{1,t}) \leq \frac{Z_{t-1}^{t-1} 2^t \prod_{s=1}^{t-1} \bar{g}_s}{\zeta N}; \quad (159)$$

and

2. if Assumption 4.A holds, then

$$\text{KL}(\eta_{1,t}||\bar{\eta}_{R,N}^{1,t}) \leq \frac{(\beta - 1)(t-1)}{N}, \quad (160)$$

while if Assumptions 5.C and 5.D also hold, then

$$\text{KL}(\eta_{1,t}||\bar{\eta}_{R,N}^{1,t}) \leq \frac{\beta(t-1)}{\zeta N}. \quad (161)$$

7. **Particle MCMC**

We now turn to particle MCMC methods, specifically i-cSMC and particle Gibbs. We will leverage results from Section 5 to prove mixing and convergence results for adaptive versions of i-cSMC and particle Gibbs under the condition that a lower bound on the $\infty$-ESS is maintained by the algorithm. Briefly, let us recall the setting and some key definitions from Section 2.5. We parameterize the Markov chain $(X_t)_{t \geq 1}$ and the potentials by a global parameter $\theta \in \Theta$ for which there is a prior distribution $\varpi(d\theta)$. Hence, $M_s$ is replaced by $M_s^\theta$ and $g_s$ is replaced by $g_s^\theta$. Throughout this section we fix $t$ and let $(Y, Y')$ be the measurable space with $Y \triangleq E^t$ and $Y \triangleq B(Y)$. Since $t$ is fixed we will suppress most of the time notation. The target distribution is

$$\pi(d\theta \times dy) \triangleq \gamma(d\theta \times dy)/Z, \quad (162)$$

where

$$\gamma(d\theta \times dy) \triangleq \prod_{s=1}^{t} g_s^\theta(y_s) M_s^\theta(y_{s-1}, dy_s) \varpi(d\theta) \quad \text{and} \quad Z \triangleq \gamma(1). \quad (163)$$

We denote the conditional distributions given $\theta$ or $y$ by, respectively, $\pi_\theta(dy)$ or $\pi_y(d\theta)$.

The particle Gibbs (PG) sampler uses the MCMC kernel

$$\Pi_G(\theta, y, d\theta \times dz) \triangleq \pi_y(d\theta)\Pi_\theta(y, dz). \quad (164)$$

First, in Section 7.1, we discuss the standard PG algorithm, where $\Pi_\theta = P^\theta_{R,N}$, the i-cSMC kernel. In particular, we summarize results from [2] and describe the close technical connections between their techniques and those from Section 4. In Section 7.2 we use results from Section 5 to prove convergence and mixing results for iterated conditional $\alpha$SMC and $\alpha$PG, where $\Pi_\theta = P^\alpha_{R,N}$, the i-c$\alpha$SMC kernel.
7.1. Convergence of PG and i-cSMC. Recall from Section 2.5 that the i-cSMC kernel is given by
\[ P_{\theta}^{R,N}(y,dz) = E_{y,z,\theta}^{R,N}[\delta_{\lambda}^{A_1}(dz)]. \] (165)
We will first state some properties of i-cSMC processes, then the particle Gibbs sampler. For notational convenience, we will write \( E_{y,z,k,\theta}^{R,N} [.] \equiv E_{y_{1:t},z_{1:t},k_{1:t},\theta}^{R,N} [.] \) for expectations with respect to the law of the cSMC process with trajectories \( y^1 = y \) and \( y^2 = z \), and lineages \( k_{1:t}^1 = 1 \) and \( k_{1:t}^2 = k \).

**Proposition 7.1** ([2] Proposition 6). For \( y \in Y \) and \( N \geq 1 \),
\[ P_{\theta}^{R,N}(y,dz) \geq \frac{Z(1 - 1/N)^t}{E_{y,z,k,\theta}^{R,N}} \pi_{\theta}(dz). \] (166)

From Proposition 7.1 is follows that if \( \sup_{y,z} E_{y,z,k,\theta}^{R,N} [Z] \leq B_N \), the cSMC kernel satisfies the minimization condition
\[ P_{\theta}^{R,N}(y,dz) \geq \varepsilon_{t,N} \pi_{\theta}(dz), \] (167)
where \( \varepsilon_{t,N} \equiv \frac{Z(1 - 1/N)^t}{E_{y,z,k,\theta}^{R,N}}. \) The minimization condition, in turn, implies uniform ergodicity and a number of other types of convergence guarantees for the i-cSMC process. For a stationary Markov chain \((\xi_k)_{k \geq 0}\) with \( \mu \)-reversible Markov kernel \( K \), the asymptotic variance for the function \( \phi \in L^2(S,\mu) \) is defined to be
\[ \mathbb{V}[\phi, K] \equiv \lim_{k \to \infty} \mathbb{V} \left[ k^{-1/2} \sum_{i=1}^k (\phi(\xi_i) - \pi(\phi)) \right]. \] (168)

**Proposition 7.2** ([2] Proposition 7). Let \( \mu \) be a probability measure on \((S,S)\) and let \( \Pi : S \times S \to [0,1] \) be a probability kernel that is reversible with respect to \( \mu \). If the stationary Markov chain defined by \( \Pi \) is \( \psi \)-irreducible and aperiodic and there exists \( \varepsilon > 0 \) such that for all \( y \in Y \), \( \Pi(y,dz) \geq \varepsilon \mu(dz) \), then
(1) for any probability measure \( \nu \ll \mu \) and \( k \geq 1 \),
\[ d_{\chi^2}(\nu \Pi^k,\mu) \leq d_{\chi^2}(\nu,\mu)(1 - \varepsilon)^k, \] (169)
(2) for any \( y \in Y \) and \( k \geq 1 \),
\[ d_{TV}(\delta_y \Pi^k,\mu) \leq (1 - \varepsilon)^k, \] (170)
and
(3) for any \( \phi \in L^2(S,\mu) \)
\[ \mathbb{V}[(\phi, \Pi) \leq (2\varepsilon^{-1} - 1)\mathbb{V}[\phi, \Pi]] + \mathbb{V}[\phi, \Pi]. \] (171)

**Corollary 7.3.** If \( \sup g_s^\theta < \infty \) for all \( s \in \{0,1,2\} \), then all the results of Proposition 7.2 apply with \( \Pi = \Pi_{s}^{R,N}[.] \equiv \Pi_{s}^{R,N}[.] \) and \( \varepsilon = \varepsilon_{t,N} \), where \( 1 - \varepsilon_{t,N} = O(1/N) \). If in addition Assumption 4.A holds, then for every \( C > 0 \) there exists an \( \varepsilon_C > 0 \) such that for all \( t > 1 \), if \( N = Ct \), then \( \varepsilon_{t,N} \geq \varepsilon_C > 0 \).

**Proof.** The first part follows from Propositions 7.1, 7.2 and 4.11. The second then follows from Proposition 4.13. \qed

**Remark 7.1.** The second part of the corollary states that, if \( \sup g_s^\theta < \infty \) and Assumption 4.A hold, then scaling \( N \) linearly with \( t \) ensures a uniform convergence rate, as measured by \( \chi^2 \)-divergence, total variation distance, or asymptotic variance.
Recall that \( \gamma_\theta(dy) = \prod_{s=1}^t g^s_\theta(y_s) M^s_\theta(y_{s-1}, dy_s) \). The results of [2, Section 7] show how to use the convergence guarantees of the i-cSMC kernel to give conditions for the geometric ergodicity of the PG sampler:

**Theorem 7.4.** If

\[
\pi \text{-ess sup}_\theta \prod_{t=1}^t \gamma_\theta(1) < \infty
\]

or there exists \( 1 \leq \beta < \infty \) such that for any \( t, s \in \mathbb{N} \),

\[
\pi \text{-ess sup}_\theta, x G_{0,t,s}(x) \leq \beta.
\]

then as soon as the Gibbs sampler is geometrically ergodic, the PG chain is geometrically ergodic.

### 7.2. Convergence of \( \alpha_{PG} \) and \( i_{-cSMC} \)

We now generalize the results of the previous section to allow for adaptive resampling. Of particular note is that our results show that \( \infty \)-ESS is a measure of effective sample size for adaptive resampling in this setting. Recall from Section 2.5 that the \( i_{-cSMC} \) kernel can be used to define the \( \alpha_{PG} \) algorithm. For \( y \in Y \), \( i_{-cSMC} \) kernel is given by

\[
P_{\theta}^{\alpha,N}(y, dz) \triangleq \mathbb{E}_{y,1,\theta}^{\alpha,N} \left[ \delta_{X_{t-1}'}(dz) \right].
\]

We call the family of Markov chains with transition kernels of the form \( P_{\theta}^{\alpha,N}(y, dz) \) \( i_{-cSMC} \) processes.

Our primary technical goal will be to prove a minorization condition for \( P_{\theta}^{\alpha,N}(y, dz) \) by generalizing Proposition 7.1. We begin by noting that, under Assumption 5.C, the normalization constants for the \( c_2_{-cSMC} \) process are given by

\[
C_1 \triangleq \frac{N}{N-1}
\]

and, for \( s = 2, \ldots, t \),

\[
C_s \triangleq \left( 1 - \sum_{k=1}^N \alpha_{s-1}^{kn} \alpha_{s-1}^{kn'} \right)^{-1}.
\]

Let

\[
\kappa_N \triangleq \max_{n \neq n', \alpha \in \mathcal{A}_N} \sum_{k=1}^N \alpha^{kn} \alpha^{kn'} \quad \text{and} \quad \kappa'_N \triangleq \kappa_N \vee 1/N,
\]

so \( C_s \leq \frac{1}{1-\kappa_N} \) for \( s = 2, \ldots, t \). Thus, for all \( s \in [t] \), \( C_s \leq \frac{1}{1-\kappa_N} \). Here \( a \vee b \) denotes the maximum of \( a, b \in \mathbb{R} \).

**Proposition 7.5.** For \( y \in Y \) and \( N \geq 2 \),

\[
P_{\theta}^{\alpha,N}(y, dz) \geq \frac{Z_t}{\mathbb{E}_{y,z,\theta}^{\alpha,N} \left[ Z_t \prod_{s=1}^t C_s \right]} \pi(dz) \geq \frac{Z_t}{\mathbb{E}_{y,z,\theta}^{\alpha,N} \left[ Z_t \right]} \pi(dz).
\]

The proof of Proposition 7.5 can be found in Appendix A.3. We recover Proposition 7.1 as a special case of Proposition 7.5 since, for SIR, \( \kappa_N = 1/N \).

Using Proposition 5.4, we obtain:
Proposition 7.6. If Assumptions $5.C$ and $5.D$ hold, then for all $N \geq 2$, 
\[
E_{\alpha,N}^{y,z,\theta}[\hat{Z}_t] \leq \frac{1}{N(\zeta N)^{t-1}} \sum_{\ell=1}^{t+1} \sum_{\tau \in \mathcal{T}_{\ell,t+1}} (\zeta N)^{t-\ell} \left( \frac{N-2}{\zeta N} \right)^{\mathbb{1}_{(\ell,>1)}} \text{G}_{0,\tau_1} C_{\ell}^{y,z}(\tau),
\]
where 
\[
C_{\ell}^{y,z}(\tau) \triangleq \prod_{i=1}^{\ell-1} \left( G_{\tau_i,\tau_{i+1}}(y_{\tau_i}) + G_{\tau_i,\tau_{i+1}}(z_{\tau_i}) \right).
\]
Hence, if Assumptions $5.C$ and $5.D$ hold and the potentials are bounded, then 
\[
E_{\alpha,N}^{y,z,\theta}[\hat{Z}_t] = 1 + O\left( \frac{N-1}{N} \right),
\]
while if Assumptions $4.A$, $5.C$ and $5.D$ hold, then 
\[
E_{\alpha,N}^{y,z,\theta}[\hat{Z}_t] \leq Z_t \left( 1 + \frac{2\beta}{\zeta N} \right)^t.
\]

Corollary 7.7. If Assumptions $4.A$, $5.C$ and $5.D$ hold, then for all $y \in Y$, 
\[
P_{\theta}^{y,N}(y, dz) \geq \epsilon_{t,N} \pi(dz),
\]
where 
\[
\epsilon_{t,N} \triangleq \frac{(1 - \frac{1}{N})(1 - \kappa_N)^{t-1}}{(1 + \frac{2\beta}{\zeta N})^t}.
\]
Furthermore, if $N \geq \frac{2\beta}{Ct(1-\kappa'_N) - \kappa'_N}$ for some constant $C > 0$, where $\kappa'_N \triangleq \kappa_N \lor N^{-1}$, then 
\[
\epsilon_{t,N} \geq \exp\left( -\frac{1}{\zeta C} \right).
\]
In particular, assuming $\kappa'_N \leq B/N$ for some constant $B > 0$, if $N \geq Ct + B$, then 
\[
\epsilon_{t,N} \geq \exp\left( -\frac{2\beta}{\zeta C} - B \right).
\]

Proof. The first part follows from Propositions 7.5 and 7.6. For the second part, we then have 
\[
\epsilon_{t,N} \geq \left( \frac{1 + \frac{2\beta}{\zeta N}}{1 - \kappa'_N} \right)^{-t} = \left( 1 + \frac{1}{1 - \kappa'_N} \left( \frac{2\beta}{\zeta N} + \kappa'_N \right) \right)^{-t} \geq \left( 1 + \frac{1}{\zeta C t} \right)^{-t} \geq \exp\left( -\frac{1}{\zeta C} \right).
\]
The final part follows after noting that if $\kappa'_N \leq B/N$, then 
\[
\frac{1}{1 - \kappa'_N} \left( \frac{2\beta}{\zeta N} + \kappa'_N \right) \geq \frac{1}{1 - B/N} \left( \frac{2\beta}{\zeta N} + B/N \right) = \frac{1}{N - B} \left( \frac{2\beta}{\zeta} + B \right).
\]
\[\square\]

Remark 7.2. In the case of SIR, Corollary 7.7 is almost as good as [2, Corollary 15]: the former result replaces $\beta - 1$ with $\beta$.

The following, which generalizes Corollary 7.3 and Theorem 7.4, is a straightforward consequence of Propositions 7.2 and 5.6 and Corollary 7.7:

Theorem 7.8. If Assumptions $5.C$ and $5.D$ hold, then for $N \geq 2$, the $i$-coSMC process with kernel $P = P_{\theta}^{\alpha,N}$
(1) is reversible with respect to $\pi$ and defines a positive operator, 
(2) if the potentials are bounded there exists $\epsilon_{t,N} = 1 + O(1/N)$ such that 
(a) for all $y \in Y$, $P(y, dz) \geq \epsilon_{t,N} \pi_\theta(dz)$, 
(b) for any measure $\nu \ll \pi_\theta$ and $k \geq 1$, 
\[
d_{\chi^2}(\nu P^k, \pi_\theta) \leq d_{\chi^2}(\nu, \pi_\theta)(1 - \epsilon_{t,N})^k, \tag{189}\]
(c) for any $y \in Y$ and $k \geq 1$, 
\[
d_{TV}(\delta_y P^k, \pi_\theta) \leq (1 - \epsilon_{t,N})^k, \tag{190}\]
(d) for any $\phi \in L^2(Y, \pi_\theta)$ 
\[
\mathbb{V}[\phi, P] \leq (2 \epsilon_{t,N}^{-1} - 1) \mathbb{V}_\pi[\phi], \tag{191}\]
(3) if in addition Assumption 4.A holds and there is a constant $B > 0$ such that 
$\kappa_N' \leq B/N$, then for every $C > 0$ there exists an $\epsilon_{B,C,\zeta} > 0$ such that with 
$N \geq Ct - B$, for any $t > 1$, 
\[
\epsilon_{t,N} \geq \epsilon_{B,C,\zeta} > 0. \tag{192}\]
Furthermore, if 
\[
\pi\text{-ess sup}_\theta \Pi_{s=1}^t \bar{g}_{\theta,s} \frac{1}{\gamma_\theta(1)} < \infty \tag{193}\]
or there exists $1 \leq \beta < \infty$ such that for any $t, s \in \mathbb{N}$, 
\[
\pi\text{-ess sup}_\theta \frac{G_{0,t}G_{t,t+s}(x)}{G_{0,t+s}} \leq \beta. \tag{194}\]
then, when Assumptions 5.C and 5.D hold and the Gibbs sampler is geometrically 
ergodic, the $\alpha$PG chain is geometrically ergodic.

Notably, our results show that $\infty$-ESS is a notion of effective sample size in the 
setting of the i-cSMC and $\alpha$PG algorithms.

Remark 7.3. At a high level, we have seen that the expected value of $\tilde{Z}_t$ arises in the 
study of the mixing properties of iterated conditional SMC (i-cSMC) Markov chains, 
which are related to the convergence of particle Gibbs (PG) samplers. In order to 
show geometric ergodicity for particle Gibbs samplers, bounds on the expected 
value of $\tilde{Z}_t$ can be used, with growth of the expectation as $t$ increases determining 
how well the particle Gibbs algorithm scales. Bounds on the expected value of $\tilde{Z}_t$ 
also allow one to obtain bounds on $\text{KL}(\pi_{1,t}||\bar{\pi}^{R,N}_{1,t})^2$. We have also shown that an 
analogous connection exists for the expected value of $\tilde{Z}_t$ between adaptive particle 
Gibbs and adaptive SMC for sampling. Hence, as a slogan, good performance of 
(adaptive) particle Gibbs is equivalent to good performance of (adaptive) SMC for 
sampling.

\footnote{In the i-cSMC setting, the expectation is with respect to the “doubly conditional SMC kernel,” whereas we require expectations with respect to the “conditional SMC kernel.” However, this is only a small technical difference and, as we have seen, the same techniques apply in both cases.}
A. Proof of Proposition 5.3

We conclude that

\[
\frac{1}{\sum_{n=1}^{N} v_n} \sum_{n=1}^{N} v_n (p - 1)^k \log^k (v_n) / k! \left(1/p\right)
\]

with the lower (upper) bound achieved if and only if

\[
\frac{1}{\sum_{n=1}^{N} v_n} \sum_{n=1}^{N} v_n (p - 1)^k \log^k (v_n) / k! = 1
\]

ward algebraic manipulation. To prove the limit equality, write

\[
\frac{\exp (\log(1 + \sum_{k=1}^{\infty} x^{-k} \log^k (v_1) / k!))}{\exp (\log(1 + \sum_{k=1}^{\infty} \sum_{n=1}^{N} v_n \cdot x^{-k} \log^k (v_n) / k!))}
\]

For (2), apply the Fact with

\[
\frac{1}{\sum_{n=1}^{N} v_n} \sum_{n=1}^{N} v_n (p - 1)^k \log^k (v_n) / k! = 1
\]

To prove the remaining parts, we make repeated use of the following:

\[
\lim_{N \to \infty} \frac{\exp (\log(1 + \sum_{k=1}^{\infty} x^{-k} \log^k (v_1) / k!))}{\exp (\log(1 + \sum_{k=1}^{\infty} \sum_{n=1}^{N} v_n \cdot x^{-k} \log^k (v_n) / k!))}
\]

For (2), apply the Fact with \( r = 1, s = p > 1, \) and note that in this case

\[
1/r - 1/s = 1 - 1/p = 1/p.
\]

We then have \( 1 \leq \|v\|_1 / \|v\|_p \leq N^{1/r - 1/s} \|v\|_s, \) with the lower (upper) bound achieved if and only if \( v \) has one non-zero entry (\( v \) has all equal entries).

For (3), in the case that \( p > 1, \) note that

\[
\|v\|_1^{q - p} \geq N^{(q - p)/q} \|v\|_q^{q - p} = N^{1/p - 1/q} \|v\|_q^{q - p},
\]

where the final equality follows since

\[
1 - p_* / q_* = 1 - p_* (1 - 1/q) = 1 - p_* + p_* / q = -p_* / p + p_* / q.
\]

We conclude that

\[
\left(\frac{\|v\|_1}{\|v\|_p}\right)^{p_*} \geq \frac{\|v\|_1^{p_*}}{N^{p_* (1/p - 1/q)]} \|v\|_q^{q - p_*}}
\]

\[
\geq \frac{\|v\|_1^{p_*}}{N^{p_* (1/p - 1/q)]} \|v\|_q^{q - p_*}}
\]

\[
= \left(\frac{\|v\|_1}{\|v\|_q}\right)^{q_*}
\]
The proof relies on the following lemma.

**Lemma A.1.** If \( y_{1,t} \in \mathbb{R} \), then

1. for \( s = 2, \ldots, t \) and any functions \( \phi_s^n : E \to \mathbb{R}, n \in [N] \),

\[
\mathbb{E}_{y_{1,t}}^\alpha \left[ \sum_n \phi_s^n(X^n_s) \mid F_{s-1} \right] \\
= \sum_{s' \neq s} \alpha_{s, s+1} \phi_s^{s'}(y_{s'}) + \sum_{n \neq n'} \sum_{s' \neq s} \alpha_{s, s+1} \phi_s^{s'}(y_{s'}) \frac{w_{n,n'}^s Q_{s,s+1}^k}{w_{n,n'}^s};
\]

(210)

2. for \( \tau \in [t-s] \),

\[
\mathbb{E}_{y_{1,t}}^\alpha \left[ \sum_n W_n^s G_{s, s+\tau}^n \mid F_{s-1} \right] \\
\leq \frac{1}{\zeta N} \sum_n w_{s-1}^n g_{n}^{s-1} G_{s, s+\tau}^n + \sum_n w_{s-1}^n G_{s-1, s+\tau}^n;
\]

(211)

and

3. for \( s = 1, \ldots, t-1 \),

\[
N \mathbb{E}_{y_{1,t}}^\alpha \left[ \tilde{Z}_t \mid F_{t-s} \right] \leq A_{t-s} + B_{t-s},
\]

(212)

where,

\[
A_{t-s} \triangleq (\zeta N)^{-s+1} \sum_n w_{t-s}^n g_{t-s}^n \left( \sum_{\ell=1}^s (\zeta N)^{s-1-\ell} G_{t-s+1, \tau}^\ell \right)
\]

(213)

\[
B_{t-s} \triangleq (\zeta N)^{-s+1} \sum_n w_{t-s}^n \left( \sum_{\ell=1}^s (\zeta N)^{s-\ell} G_{t-s, \tau}^\ell \right).
\]

(214)

The case of \( p = 1 \) follows from the \( p > 1 \) case and part (1).
Proof. For (1),

\[
\mathbb{E}_{Y_{1,t}}^{a,N} \left[ \sum_{n} \phi^{n}_{s}(X^{n}_{s}) \mid \mathcal{F}_{s-1} \right] \\
= \sum_{f_{s}} \sum_{g_{s-1}} \alpha_{s-1}^{f_{s}-1} \prod_{k \neq f_{s}} r_{k}(a_{s-1}^{k} | w_{s-1}, x_{1,s-1}) \mathbb{E}_{Y_{1,s}}^{a,N} \left[ \sum_{n} \phi^{n}_{s}(X^{n}_{s}) \mid \mathcal{F}_{s-1}, A_{s-1} = a_{s-1}, F_{s} = f_{s} \right] \\
= \sum_{f_{s}} \sum_{g_{s-1}} \alpha_{s-1}^{f_{s}-1} \prod_{k \neq f_{s}} r_{k}^{k_{a_{s-1}^{k}}} w_{s-1}^{k} g_{s-1}(x_{s-1}^{a_{s-1}^{k}}) \left( \phi_{f_{s}}^{f_{s}}(y_{s}) + \sum_{n \neq f_{s}} \mathbb{E} \left[ \phi^{n}_{s}(X_{s}) \mid X_{s-1} = x^{n}_{s-1} \right] \right) \\
= \sum_{f_{s}} \alpha_{s-1}^{f_{s}-1} \phi_{f_{s}}^{f_{s}}(y_{s}) + \sum_{f_{s} \neq n} \sum_{k} \alpha_{s-1}^{f_{s}-1} \frac{\alpha_{n}^{k} w_{n}^{k}}{w_{s}^{k}} g_{s-1}(x_{s-1}^{n}) \mathbb{E} \left[ \phi^{n}_{s}(X_{s}) \mid X_{s-1} = x^{n}_{s-1} \right] \\
= \sum_{f_{s}} \alpha_{s-1}^{f_{s}-1} \phi_{f_{s}}^{f_{s}}(y_{s}) + \sum_{f_{s} \neq n} \sum_{k} \alpha_{s-1}^{f_{s}-1} \frac{\alpha_{n}^{k} w_{n}^{k}}{w_{s}^{k}} Q_{s-1,s}^{k}(\phi^{n}_{s})
\]

For (2), choosing \(\phi^{n}_{s}(x) = w_{n}^{s} G_{s,s+\tau}(x)\), we have

\[
\mathbb{E}_{y_{1,t}}^{a,N} \left[ \sum_{n} W_{n}^{s} G_{s,s+\tau}^{n} \mid \mathcal{F}_{s-1} \right] \\
= \sum_{f_{s}} \alpha_{s-1}^{f_{s}-1} w_{s}^{f_{s}} G_{s,s+\tau}^{y} + \sum_{f_{s} \neq n} \sum_{k} \alpha_{s-1}^{f_{s}-1} \frac{\alpha_{n}^{k} w_{n}^{k}}{w_{s}^{k}} G_{s-1,s}^{k} \mathbb{E} \left[ \phi^{n}_{s}(X_{s}) \mid X_{s-1} = x^{n}_{s-1} \right] \\
= G_{s,s+\tau}^{y} \sum_{f_{s}} \alpha_{s-1}^{f_{s}-1} w_{s}^{f_{s}} + \sum_{f_{s} \neq n} \sum_{k} \alpha_{s-1}^{f_{s}-1} \frac{\alpha_{n}^{k} w_{n}^{k}}{w_{s}^{k}} G_{s-1,s+\tau}^{k} \\
\leq G_{s,s+\tau}^{y} \sum_{f_{s}} \alpha_{s-1}^{f_{s}-1} \left\| w_{s}^{f_{s}} \right\|_{1} + \sum_{f_{s} \neq n} \sum_{k} \alpha_{s-1}^{f_{s}-1} \frac{\alpha_{n}^{k} w_{n}^{k}}{w_{s}^{k}} G_{s-1,s+\tau}^{k} \\
= G_{s,s+\tau}^{y} \sum_{f_{s}} \alpha_{s-1}^{f_{s}-1} \frac{\left\| w_{s}^{f_{s}} \right\|_{1}}{\zeta_{N}} + \sum_{f_{s} \neq n} \sum_{k} \alpha_{s-1}^{f_{s}-1} \frac{\alpha_{n}^{k} w_{n}^{k}}{w_{s}^{k}} G_{s-1,s+\tau}^{k} \\
= \frac{1}{\zeta_{N}} \sum_{k} w_{s-1}^{k} g_{s-1}^{k} G_{s,s+\tau}^{y} + \sum_{k} w_{s-1}^{k} G_{s-1,s+\tau}^{k} \\
\text{where the inequality follows from Assumption 5.D and we have repeatedly used Assumption 5.C.}
\]

To show (3), we start by using (2) with \(s = t\) and \(\tau = 1\):

\[
\mathbb{E}_{y_{1,t}}^{a,N} \left[ \sum_{n} W_{n}^{s} g_{t-1}^{n} \mid \mathcal{F}_{t-1} \right] = \frac{1}{\zeta_{N}} \sum_{k} w_{t-1}^{k} g_{t-1}^{k} + \zeta_{N} \sum_{m} w_{t-1}^{m} G_{t-1,t+1}^{m} \\
= A_{t-1} + B_{t-1},
\]

Hence, (3) holds for \(s = 1\). We now assume that the bound holds for some \(s \in \{1, \ldots, t-2\}\) and establish that it also holds for \(s + 1\). Using the inductive
hypothesis,
\[ N E_{y_t, t}^{α, N} \left[ \tilde{Z}_t | F_{t-s-1} \right] = \mathbb{E}_{y_t, t}^{α, N} \left[ N E_{y_t, t}^{α, N} \left[ \tilde{Z}_t | F_{t-s} \right] | F_{t-s-1} \right] \]
\[ \leq \mathbb{E}_{y_t, t}^{α, N} \left[ A_{t-s} + B_{t-s} | F_{t-s-1} \right]. \] (224)

Using (2), we have
\[ A := \mathbb{E}_{y_t, t}^{α, N} \left[ A_{t-s} | F_{t-s-1} \right] \]
\[ = (ζN)^{-s+1} \left( \sum_{t=1}^{s} \sum_{τ \in T_{t,s}} (ζN)^{s-1-ℓ} G_{t-s+1, τ_1}^{y} C_{ℓ}^{y}(τ) \right) \mathbb{E}_{y_t, t}^{α, N} \left[ \sum_n W_{t-s}^{n} G_{t-s}^{n} \left| F_{t-s-1} \right. \right] \]
\[ \leq (ζN)^{-s} \left( \sum_{t=1}^{s} \sum_{τ \in T_{t,s}} (ζN)^{s-1-ℓ} G_{t-s+1, τ_1}^{y} C_{ℓ}^{y}(τ) \right) \]
\[ \times \left( \sum_n w_{t-s-1}^{n} G_{t-s-1}^{n} + ζN \sum_n w_{t-s-1}^{n} G_{t-s-1, t-s+1}^{n} \right) \] (227)

and
\[ B := \mathbb{E}_{y_t, t}^{α, N} \left[ B_{t-s} | F_{t-s-1} \right] \]
\[ = (ζN)^{-s+1} \left( \sum_{t=1}^{s} \sum_{τ \in T_{t,s}} (ζN)^{s-1-ℓ} E_{y_t, t}^{α, N} \left[ \sum_n W_{t-s}^{n} G_{t-s, τ_1}^{n} | F_{t-s-1} \right] C_{ℓ}^{y}(τ) \right) \]
\[ \leq (ζN)^{-s} \left( \sum_{t=1}^{s} \sum_{τ \in T_{t,s}} (ζN)^{s-1-ℓ} \left( \sum_n w_{t-s-1}^{n} G_{t-s-1}^{n} + ζN \sum_n w_{t-s-1}^{n} G_{t-s-1, t-s+1}^{n} \right) C_{ℓ}^{y}(τ) \right). \] (228)

Hence,
\[ A + B \leq (ζN)^{-s} \sum_n w_{t-s-1}^{n} G_{t-s-1}^{n} \left( \sum_{t=1}^{s} \sum_{τ \in T_{t,s}} (ζN)^{s-1-ℓ} G_{t-s, t-s+1}^{y} C_{ℓ}^{y}(τ) \right) \]
\[ + (ζN)^{-s} \sum_n w_{t-s-1}^{n} G_{t-s-1}^{n} \left( \sum_{t=1}^{s} \sum_{τ \in T_{t,s}} (ζN)^{s-1-ℓ} G_{t-s}^{n} C_{ℓ}^{y}(τ) \right) \]
\[ + (ζN)^{-s} \sum_n w_{t-s-1}^{n} \left( \sum_{t=1}^{s} \sum_{τ \in T_{t,s}} (ζN)^{s-ℓ} G_{t-s-1, t-s+1}^{n} C_{ℓ}^{y}(τ) \right) \]
\[ + (ζN)^{-s} \sum_n \sum_{τ \in T_{t,s}} (ζN)^{s-ℓ+1} G_{t-s-1, τ_1}^{n} C_{ℓ}^{y}(τ). \] (232)
Summing the parenthesized double sums of the first two terms yields

\[ \sum_{\ell=1}^{s} \sum_{\tau \in T_{t,s}} (\zeta N)^{s-\ell} G_{t-s,t-s+1}^{y} G_{t-s+1,\tau_1}^{y} C_{\ell}^{y}(\tau) + \sum_{\ell=1}^{s} \sum_{\tau \in T_{t,s}} (\zeta N)^{s-\ell} G_{t-s,\tau_1}^{y} C_{\ell}^{y}(\tau) \]

\[ = \sum_{\ell=1}^{s+1} \sum_{\tau \in T_{t,s+1}} (\zeta N)^{s-\ell} G_{t-s,\tau_1}^{y} C_{\ell}^{y}(\tau) + \sum_{\ell=1}^{s+1} \sum_{\tau \in T_{t,s+1}} (\zeta N)^{s-\ell} G_{t-s-1,\tau_1}^{y} C_{\ell}^{y}(\tau) \]  \hspace{1cm} (233)

\[ = \sum_{\ell=1}^{s+1} \sum_{\tau \in T_{t,s+1}} (\zeta N)^{s-\ell} G_{t-s,\tau_1}^{y} C_{\ell}^{y}(\tau), \]  \hspace{1cm} (234)

so the first two terms are equal to \( A_{t-(s+1)} \). Summing the parenthesized double sums of the last two terms yields

\[ \sum_{\ell=1}^{s} \sum_{\tau \in T_{t,s}} (\zeta N)^{s-\ell} G_{t-s-1,t-s+1}^{y} G_{t-s+1,\tau_1}^{y} C_{\ell}^{y}(\tau) + \sum_{\ell=1}^{s} \sum_{\tau \in T_{t,s}} (\zeta N)^{s-\ell} G_{t-s-1,\tau_1}^{y} C_{\ell}^{y}(\tau) \]

\[ = \sum_{\ell=1}^{s+1} \sum_{\tau \in T_{t,s+1}} (\zeta N)^{s-\ell+1} G_{t-s-1,\tau_1}^{y} C_{\ell}^{y}(\tau) + \sum_{\ell=1}^{s+1} \sum_{\tau \in T_{t,s+1}} (\zeta N)^{s-\ell+1} G_{t-s-1,\tau_1}^{y} C_{\ell}^{y}(\tau) \]  \hspace{1cm} (235)

\[ = \sum_{\ell=1}^{s+1} \sum_{\tau \in T_{t,s+1}} (\zeta N)^{s-\ell+1} G_{t-s-1,\tau_1}^{y} C_{\ell}^{y}(\tau), \]  \hspace{1cm} (236)

so the last two terms are equal to \( B_{t-(s+1)} \).

Using part (3) of Lemma [A.1] with \( s = t - 1 \), we have

\[ N \mathbb{E}_{y_1,t}^{\alpha,N} \left[ \tilde{Z}_t \right] \leq \mathbb{E}_{y_1,t}^{\alpha,N} \left[ A_1 + B_1 \right]. \]  \hspace{1cm} (237)

Therefore,

\[ \mathbb{E}_{y_1,t}^{\alpha,N} [A_1] = (\zeta N)^{-t+2} \mathbb{E}_{y_1,t}^{\alpha,N} \left[ \sum_{n} g_{n}^{y} \right] \left( \sum_{\ell=1}^{t-1} \sum_{\tau \in T_{t-\ell,s-1}} (\zeta N)^{t-2-\ell} G_{2,\tau_1}^{y} C_{\ell}^{y}(\tau) \right) \]  \hspace{1cm} (238)

\[ = (\zeta N)^{-t+2} (G_{1,2}^{y} + (N-1)G_{0,2}) \left( \sum_{\ell=1}^{t-1} \sum_{\tau \in T_{t-\ell-1,s-1}} (\zeta N)^{t-2-\ell} G_{2,\tau_1}^{y} C_{\ell}^{y}(\tau) \right) \]  \hspace{1cm} (239)

and

\[ \mathbb{E}_{y_1,t}^{\alpha,N} [B_1] = (\zeta N)^{-t+2} \left( \sum_{\ell=1}^{t-1} \sum_{\tau \in T_{t-\ell,s-1}} (\zeta N)^{t-1-\ell} \mathbb{E}_{y_1,t}^{\alpha,N} \left[ \sum_{n} G_{n}^{y} \right] C_{\ell}^{y}(\tau) \right) \]  \hspace{1cm} (240)

\[ = (\zeta N)^{-t+2} \left( \sum_{\ell=1}^{t-1} \sum_{\tau \in T_{t-\ell,s-1}} (\zeta N)^{t-1-\ell}(G_{1,\tau_1}^{y} + (N-1)G_{0,\tau_1}) C_{\ell}^{y}(\tau) \right). \]  \hspace{1cm} (241)
Hence, using arguments analogous to those from the proof of Lemma A.1 and the fact that $G_{0,1} = 1$ yields

$$E_{y_1,t}^\alpha [A_1 + B_1] = (\zeta N)^{-t+2} \sum_{\ell=1}^{t} \sum_{\tau \in T_{t,\ell}} (\zeta N)^{t-1-\ell} G_{1,\tau_1}^{y} C_{\ell}^{y}(\tau) \quad (242)$$

$$+ \frac{N-1}{\zeta N} (\zeta N)^{-t+2} \sum_{\ell=1}^{t} \sum_{\tau \in T_{t,\ell}} (\zeta N)^{t-\ell} G_{0,\tau_1}^{y} C_{\ell}^{y}(\tau) \quad (243)$$

$$= (\zeta N)^{-t+2} \sum_{\ell=1}^{t+1} \sum_{\tau \in T_{t+1,\ell}} (\zeta N)^{t-\ell} G_{0,\tau_1}^{y} C_{\ell}^{y}(\tau) \quad (244)$$

$$+ \frac{N-1}{\zeta N} (\zeta N)^{-t+2} \sum_{\ell=1}^{t+1} \sum_{\tau \in T_{t+1,\ell}} (\zeta N)^{t-\ell} G_{0,\tau_1}^{y} C_{\ell}^{y}(\tau) \quad (245)$$

A.3. Proof of Proposition 7.5. First, observe that we can write the i-coSMC kernel as

$$P^\alpha_{\theta,N}(y, dz) = E_{y,d}^\alpha \left[ \sum_{k \in [N]} T^\alpha_k(X_{1,t}, A_{1,t}, dz) \right], \quad (246)$$

where

$$T^\alpha_k(x_{1,t}, a_{1,t}, dz) \triangleq \delta_{\gamma_N}(dz) \mathbb{I}(k_t = a_t) \prod_{s=1}^{t-1} \mathbb{I}(k_s = a_{s+1}^{k_s+1}). \quad (247)$$

Next note that

$$\sum_{k_1=1}^{N} \mathbb{I}(x_{1,t}^k \in S) E_{y_1,d}^\alpha [X_1 \in dx_1, F_{1}^{y} = F_{1}^{y}]$$

$$= \sum_{k_1=1}^{N} \mathbb{I}(x_{1,t}^k \in S) \frac{1}{N} \delta_{y_1}(dx_{1}^{y}) \prod_{n \neq f_1} M_1(dx_{1}^{n})$$

$$\geq \frac{N}{C_1} \sum_{k_1=1}^{N} \int_E \mathbb{I}(x_{1,t}^k \in S) \frac{C_1}{N^2} \mathbb{I}(f_1^{y} \neq k_1) \delta_{y_1}(dx_{1}^{y}) \delta_{z_1}(dx_{1}^{k_1}) \prod_{n \neq f_1, k_1} M_1(dx_{1}^{n}) M_1(dz_1)$$

$$= \frac{N}{C_1} \sum_{k_1=1}^{N} \int_E \mathbb{I}(x_{1,t}^k \in S) E_{y_1,d}^\alpha [X_1 \in dx_1, F_{1}^{y} = f_1^{y}, F_{1}^{z} = k_1] M_1(dz_1).$$

For the remainder of the proof, to keep notation compact when writing laws, instead of writing, e.g., $X_s \in x_s$ or $F_{s}^{y} = f_{s}^{y}$, whenever a random variable is instantiated to be (the differential) of the lowercase version of itself, we will write only the random
variance: for example, $X_s$ or $F_s^y$. Now, for $s = 2, \ldots, t$,

\[
\sum_{k_s=1}^{N} \mathbb{I}(x_{1,t}^k \in S)\mathbb{I}(k_{s-1} = a_{s-1}^{k_s})\mathbb{P}_{y,\theta}^{\alpha,N}[X_s, A_{s-1}, F_s^y \mid X_{1,s-1}, A_{1,s-2}, F_{s-1}^y]
\]

\[
= \sum_{k_s=1}^{N} \mathbb{I}(x_{1,t}^k \in S)\mathbb{I}(k_{s-1} = a_{s-1}^{k_s})\alpha_{s-1}^{f_s^y f_s^y-1}\mathbb{I}(a_{s-1}^{f_s^y} = f_s^y)\mathbb{I}(a_{s-1}^{k_s} = k_{s-1})
\]

\[
\geq \sum_{k_s=1}^{N} \frac{1}{C_s\alpha_{s-1}^{k_s}} \int_E \mathbb{I}(x_{1,t}^k \in S)C_s\mathbb{I}(f_s^y \neq k_s)\mathbb{I}(a_{s-1}^{f_s^y} = f_s^y)\mathbb{I}(a_{s-1}^{k_s} = k_{s-1})
\]

\[
\times \alpha_{s-1}^{f_s^y f_s^y-1}\alpha_{s-1}^{k_s} \delta_{y_s}(dx_s^{f_s^y}) \delta_{z_s}(dx_s^{k_s})
\]

\[
\times \prod_{n \neq f_s^y, k_s} r_n(a_{s-1}^{n} \mid w_{s-1}, x_{1,s-1})M_s(x_{s-1}^{a_{s-1}^{n}}, x_{s}^{n})
\]

\[
\times M_s(x_{s-1}^{k_{s-1}}, dz_s)r_k_s(k_{s-1} \mid w_{s-1}, x_{1,s-1})
\]

Using Eqs. (248) and (250), we have (note that the terms such as those involving $a_0$ should be ignored)

\[
\sum_{k \in [N]} \mathbb{I}(x_{1,t}^k \in S)\mathbb{I}(k_t = \alpha_t)\prod_{s=1}^{t} \mathbb{I}(k_{s-1} = a_{s-1}^{k_s})\mathbb{P}_{y,\theta}^{\alpha,N}[X_{1,t}, A_{1,t}, F_{1,t}^y]
\]

\[
= \sum_{k \in [N]} \mathbb{I}(x_{1,t}^k \in S)\mathbb{I}(k_t = \alpha_t)\mathbb{P}_{y,\theta}^{\alpha,N}[A_t \mid X_{1,t}, A_{1,t-1}]
\]

\[
\geq \sum_{k \in [N]} \int_E \mathbb{I}(x_{1,t}^k \in S)\mathbb{I}(k_t = \alpha_t)\frac{\mathbb{P}_{y,\theta}^{\alpha,N}[X_s, A_{s-1}, F_s^y \mid X_{1,s-1}, A_{1,s-2}, F_{s-1}^y]}{(\prod_{s=1}^{t-1} C_s)(\prod_{s=1}^{t-1} \alpha_{s-1}^{k_{s-1}})} \sum_{n=1}^{N} w_t^{a_t} g_t(x_t^{a_t})
\]

\[
\times \prod_{s=1}^{t} M_s(x_{s-1}^{k_{s-1}}, dz_s)r_k_s(k_{s-1} \mid w_{s-1}, x_{1,s-1})
\]

\[
\times \prod_{s=1}^{t} \mathbb{P}_{y,\theta}^{\alpha,N}[X_s, A_{s-1}, F_s^y, F_s^z = k_s \mid X_{1,s-1}, A_{1,s-2}, F_{s-1}^y, F_{s-1}^z = k_{s-1}]
\]
\[
\begin{align*}
&= \sum_{k \in [N[t]} \int \frac{N \mathbf{1}(z_{1,t} \in S)}{(\prod_{s=1}^{t} C_s) \sum_{n=1}^{N} w^n_t g_t(x^n_t)} \pi_{y,z,\theta}^{\alpha,N} [X_{1,t}, A_{1,t-1}, F_{1,t}, F_{1,t} = k] Q_{0,t+1}(dz_{1,t}) \\
&= \int_S \sum_{k \in [N[t]} \frac{Z_t}{Z_t \prod_{s=1}^{t} C_s} \pi_{y,z,\theta}^{\alpha,N} [X_{1,t}, A_{1,t-1}, F_{1,t}, F_{1,t} = k] \pi(dz_{1,t})
\end{align*}
\]
from which the result follows.

Acknowledgments

Thanks to Arnaud Doucet for critical feedback and numerous helpful suggestions, to Cameron Freer for feedback on early versions of this work, to Josh Tenenbaum for discussions that helped to inspire this work, and to Vikash Mansinghka for suggesting we investigate the expected value of SMC estimators. JHH was supported by the U.S. Government under FA9550-11-C-0028 and awarded by the DoD, Air Force Office of Scientific Research, National Defense Science and Engineering Graduate (NDSEG) Fellowship, 32 CFR 168a. This research was carried out in part while DMR held a Research Fellowship at Emmanuel College, Cambridge, with funding also from a Newton International Fellowship through the Royal Society.

References

[1] C. Andrieu, A. Doucet, and R. Holenstein. “Particle Markov chain Monte Carlo methods”. In: Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72.3 (2010), pp. 269–342.
[2] C. Andrieu, A. Lee, and M. Vihola. “Uniform Ergodicity of the Iterated Conditional SMC and Geometric Ergodicity of Particle Gibbs samplers”. In: arXiv.org (Dec. 2013). arXiv:1312.6432v1 [math.PR].
[3] C. Andrieu and G. O. Roberts. “The pseudo-marginal approach for efficient Monte Carlo computations”. In: The Annals of Statistics 37.2 (Apr. 2009), pp. 697–725.
[4] C. Andrieu and M. Vihola. “Convergence properties of pseudo-marginal Markov chain Monte Carlo algorithms”. In: The Annals of Applied Probability (2014).
[5] N. Chopin and S. S. Singh. “On the particle Gibbs sampler”. In: arXiv.org (Apr. 2013). arXiv:1304.1887v1 [stat.CO].
[6] J. Cornebise, E Moulines, and J. Olsson. “Adaptive methods for sequential importance sampling with application to state space models”. In: Statistics and Computing 18.4 (2008), pp. 461–480.
[7] T. M. Cover and J. A. Thomas. Elements of Information Theory. John Wiley & Sons, 1991.
[8] D. Crisan and O Obamubi. “Particle filters with random resampling times”. In: Stochastic processes and their applications 122.4 (2012), pp. 1332–1368.
[9] P. Del Moral. Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications. New York: Springer, 2004.
[10] P. Del Moral, A. Doucet, and A. Jasra. “On adaptive resampling strategies for sequential Monte Carlo methods”. In: Bernoulli 18.1 (Feb. 2012), pp. 252–278.
[11] P. Del Moral, A. Doucet, and A. Jasra. “Sequential Monte Carlo samplers”. In: Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68.3 (2006), pp. 411–436.
[12] A. Doucet, N. de Freitas, and N. Gordon. *Sequential Monte Carlo in Practice*. New York: Springer, 2001.
[13] A. Doucet, S. J. Godsill, and C. Andrieu. “On sequential Monte Carlo sampling methods for Bayesian filtering”. In: *Statistics and Computing* 10.3 (2000), pp. 197–208.
[14] A. Doucet and A. M. Johansen. “A tutorial on particle filtering and smoothing: fifteen years later”. In: *Handbook of Nonlinear Filtering*. Ed. by D Crisan and B Rozovsky. Cambridge: Cambridge University Press, 2010.
[15] R. Holenstein. “Particle Markov Chain Monte Carlo”. PhD thesis. Vancouver: University Of British Columbia, 2009.
[16] N. Kantas, A. Doucet, S. S. Singh, and J. M. Maciejowski. “An overview of sequential Monte Carlo methods for parameter estimation in general state-space models”. In: *15th IFAC Symposium on System Identification*. 2009, pp. 774–785.
[17] H. R. Künsch. “Particle filters”. In: *Bernoulli* 19.4 (Sept. 2013), pp. 1391–1403.
[18] A. Lee and K. Latuszynski. “Variance bounding and geometric ergodicity of Markov chain Monte Carlo kernels for approximate Bayesian computation”. In: *Biometrika* 101.3 (2014), pp. 655–671.
[19] F. Lindsten, R. Douc, and E Moulines. “Uniform ergodicity of the Particle Gibbs sampler”. In: *arXiv.org* (Jan. 2014). arXiv:1401.0683v1 [math.ST]
[20] J. S. Liu. “Metropolized independent sampling with comparisons to rejection sampling and importance sampling”. In: *Statistics and Computing* 6.2 (1996), pp. 113–119.
[21] J. S. Liu and R. Chen. “Blind deconvolution via sequential imputations”. In: *Journal of the American Statistical Association* 90.430 (1995), pp. 567–576.
[22] R. M. Neal. “Annealed importance sampling”. In: *Statistics and Computing* 11.2 (2001), pp. 125–139.
[23] N. Whiteley, A. Lee, and K. Heine. “On the role of interaction in sequential Monte Carlo algorithms”. In: *arXiv.org* (Sept. 2013). arXiv:1309.2918v1 [stat.CO].

Massachusetts Institute of Technology
URL: http://jhhuggins.org/
E-mail address: jhuggins@mit.edu

University of Toronto
URL: http://danroy.org/
E-mail address: droy@utstat.toronto.edu