arXiv:2404.13302v1 [stat.CO] 20 Apr 2024

Monte Carlo sampling with integrator snippets

Christophe Andrieu, Mauro Camara Escudero and Chang Zhang

April 23, 2024

School of Mathematics, University of Bristol

Abstract

Assume interest is in sampling from a probability distribution μ defined on $(\mathsf{Z}, \mathscr{Z})$. We develop a framework to construct sampling algorithms taking full advantage of numerical integrators of ODEs, say $\psi\colon \mathsf{Z} \to \mathsf{Z}$ for one integration step, to explore μ efficiently and robustly. The popular Hybrid/Hamiltonian Monte Carlo (HMC) algorithm [17, 29] and its derivatives are example of such a use of numerical integrators. However we show how the potential of integrators can be exploited beyond current ideas and HMC sampling in order to take into account aspects of the geometry of the target distribution. A key idea is the notion of integrator snippet, a fragment of the orbit of an ODE numerical integrator ψ, and its associate probability distribution $\bar{\mu}$, which takes the form of a mixture of distributions derived from μ and ψ. Exploiting properties of mixtures we show how samples from $\bar{\mu}$ can be used to estimate expectations with respect to μ. We focus here primarily on Sequential Monte Carlo (SMC) algorithms, but the approach can be used in the context of Markov chain Monte Carlo algorithms as discussed at the end of the manuscript. We illustrate performance of these new algorithms through numerical experimentation and provide preliminary theoretical results supporting observed performance.

Contents

1 Overview and motivation: SMC sampler with HMC

Assume interest is in sampling from a probability distribution μ on a probability space $(\mathsf{Z}, \mathscr{Z})$. The main ideas of sequential Monte Carlo (SMC) samplers to simulate from μ are (a) to define a sequence of probability distributions $\{\mu_n, n \in [\![0, P]\!]\}$ on $(\mathsf{Z}, \mathscr{Z})$ where $\mu_P = \mu$, μ_0 is chosen by the user, simple to sample from and the sequence $\{\mu_n, n \in [\![P-1]\!]\}$ "interpolates" μ_0 and μ_P, (b) and to propagate a cloud of samples $\{z_n^{(i)} \in \mathsf{Z}, i \in [\![N]\!]\}$ for $n \in [\![0, P]\!]$ to represent $\{\mu_n, n \in [\![0, P]\!]\}$ using an importance sampling/resampling mechanism [16]. Notation and definitions used throughout this paper can be found in Appendix A.

After initialisation, for $n \in [\![P]\!]$, samples $\{z_{n-1}^{(i)}, i \in [\![N]\!]\}$, representing μ_{n-1}, are propagated thanks to a user-defined mutation Markov kernel $M_n \colon \mathsf{Z} \times \mathscr{Z} \to [0, 1]$, as follows. For $i \in [\![N]\!]$ sample $\tilde{z}_n^{(i)} \sim M_n(z_{n-1}^{(i)}, \cdot)$ and compute the importance weights, assumed to exist for the moment,

$$\omega_n^{(i)} = \frac{\mathrm{d}\mu_n \widehat{\otimes} L_{n-1}}{\mathrm{d}\mu_{n-1} \otimes M_n}\left(z_{n-1}^{(i)}, \tilde{z}_n^{(i)}\right), \tag{1}$$

where $L_{n-1} \colon \mathsf{Z} \times \mathscr{Z} \to [0, 1]$ is a user-defined "backward" Markov kernel required to define importance sampling on $\mathsf{Z} \times \mathsf{Z}$ and swap the rôles of $z_{n-1}^{(i)}$ and $\tilde{z}_n^{(i)}$, in the sense that for $f \colon \mathsf{Z} \to \mathbb{R}$ μ_n-integrable,

$$\int f(z') \frac{\mathrm{d}\mu_n \widehat{\otimes} L_{n-1}}{\mathrm{d}\mu_{n-1} \otimes M_n}(z, z') \mu_{n-1}(\mathrm{d}z) M_n(z, \mathrm{d}z') = \mu_n(f).$$

We adopt this measure theoretic formulation of importance weights out of necessity in order to take into account situations such as when μ_{n-1} and μ_n both have densities with respect to the Lebesgue measure but M_n is a Metropolis-Hastings (MH) update, which does not possess such a density – more details are provided in Appendices A-B but can be omitted to understand how the algorithms considered in the manuscript proceed.

The mutation step is followed by a selection step where for $i \in [\![N]\!]$, $z_n^{(i)} = \tilde{z}_n^{(a_i)}$ for a_i the random variable taking values in $[\![N]\!]$ with $\mathbb{P}(a_i = k) \propto \omega_n^{(k)}$. The procedure is summarized in Alg. 1.

Given $\{M_n, n \in [\![P]\!]\}$, theoretically optimal choice of $\{L_{n-1}, n \in [\![P]\!]\}$ is well understood but tractability is typically obtained by assuming that M_n is μ_{n-1}-invariant, or considering approximations of $\{L_{n-1}, n \in [\![P]\!]\}$ and that M_n is μ_n-invariant. This makes Markov chain Monte Carlo (MCMC) kernels very attractive choices for M_n, and the use of measure theoretic tools inevitable.

A possible choice of MCMC kernel is that of the hybrid Monte Carlo method, a MH update using a discretization of Hamilton's equations [17, 29], which can be thought of as a particular instance of a more general strategy relying on numerical integrators of ODEs possessing properties of interest. More specifically, assume that interest is in sampling π defined on $(\mathsf{X}, \mathscr{X})$. First the problem is embedded into that of sampling from the joint distribution $\mu(\mathrm{d}z) := \pi \otimes \varpi(\mathrm{d}z) = \pi(\mathrm{d}x)\varpi(\mathrm{d}v)$ defined on $(\mathsf{Z}, \mathscr{Z}) = (\mathsf{X} \times \mathsf{V}, \mathscr{X} \otimes \mathscr{V})$, where v is an auxiliary variable facilitating sampling. Following the SMC framework we set $\mu_n(\mathrm{d}z) := \pi_n \otimes \varpi_n(\mathrm{d}z)$ for $n \in [\![0, P]\!]$, a sequence of distributions on $(\mathsf{Z}, \mathscr{Z})$ with $\pi_P = \pi$, $\{\pi_n, n \in [\![0, P-1]\!]\}$ probabilities on $(\mathsf{X}, \mathscr{X})$ and $\{\varpi_n, n \in [\![0, P]\!]\}$ on $(\mathsf{V}, \mathscr{V})$. With $\psi \colon \mathsf{Z} \to \mathsf{Z}$ an integrator of an ODE of interest, one can use $\psi^k(z)$ for some $k \in \mathbb{N}$ as a proposal in a MH update mechanism; v is a source of randomness allowing "exploration", resampled every now and then. Again, hereafter we let $z =: (x, v) \in \mathsf{X} \times \mathsf{V}$ be the corresponding components of z.

Example 1 (Leapfrog integrator of Hamilton's equations)**.** Assume that $\mathsf{X} = \mathsf{V} = \mathbb{R}^d$, that $\{\pi_n, \varpi_n, n \in [\![0, P]\!]\}$ have densities, denoted $\pi_n(x)$ and $\varpi_n(v)$, with respect to the Lebesgue measure and let $x \mapsto U_n(x) := -\log \pi_n(x)$. For $n \in [\![0, P]\!]$ and U_n differentiable, Hamilton's equations for the potential $H_n(x, v) := U_n(x) + \frac{1}{2}|v|^2$ are

$$\dot{x}_t = v_t, \dot{v}_t = -\nabla U_n(x_t), \tag{2}$$

for $i \in [\![N]\!]$ **do**
  Sample $z_0^{(i)} \sim \mu_0(\cdot)$;
  Set $\omega_0^{(i)} = 1$
end
for $n \in [\![P]\!]$ **do**
  for $i \in [\![N]\!]$ **do**
    Sample $\tilde{z}_n^{(i)} \sim M_n\big(z_{n-1}^{(i)}, \cdot\big)$;
    Compute $w_n^{(i)}$ as in (1).
  end
  for $i \in [\![N]\!]$ **do**
    Sample $a_i \sim \mathrm{Cat}\big(\omega_n^{(1)}, \ldots, \omega_n^{(N)}\big)$
    Set $z_n^{(i)} = \tilde{z}_n^{(a_i)}$
  end
end

Algorithm 1: Generic SMC sampler

and possess the important property that $H_n(x_t, v_t) = H_n(x_0, v_0)$ for $t \geq 0$. The corresponding leapfrog integrator is given, for some $\varepsilon > 0$, by

$$\psi_n(x, v) = {}_{\mathrm{B}}\psi \circ {}_{\mathrm{A}}\psi_n \circ {}_{\mathrm{B}}\psi(x, v) \tag{3}$$
$${}_{\mathrm{B}}\psi(x, v) := \big(x, v - \tfrac{1}{2}\varepsilon \nabla U_n(x)\big), \quad {}_{\mathrm{A}}\psi_n(x, v) = (x + \varepsilon\, v, v)\,.$$

We point out that, with the exception of the first step, only one evaluation of $\nabla U(x)$ is required per integration step since the rightmost ${}_{\mathrm{B}}\psi$ in (3) recycles the last computation from the last iteration. Let $\sigma \colon \mathsf{Z} \to \mathsf{Z}$ such that for any $f \colon \mathsf{Z} \to \mathsf{Z}$, $f \circ \sigma(x, v) = f(x, -v)$, it is standard to check that $\psi_n^{-1} = \sigma \circ \psi_n \circ \sigma$ and note that $\mu_n \circ \sigma = \mu_n$ in this particular case. In its most basic form the hybrid Monte Carlo MH update leaving μ_n invariant proceeds as follows, for $(z, A) \in \mathsf{Z} \times \mathscr{Z}$

$$M_{n+1}(z, A) = \int \varpi_n(\mathrm{d}v') \big[\alpha_n(x, v'; T)\mathbf{1}\{\psi_n^T(x, v') \in A\} + \bar{\alpha}_n(x, v'; T)\mathbf{1}\{\sigma(x, v) \in A\}\big]\,, \tag{4}$$

with, for some user defined $T \in \mathbb{N}$,

$$\alpha_n(z; T) := 1 \wedge \frac{\mu_n \circ \psi_n^T(z)}{\mu_n(z)}\,, \tag{5}$$

and $\bar{\alpha}_n(z; T) = 1 - \alpha_n(z; T)$.

Other ODEs, capturing other properties of the target density, or other types of integrators are possible. However a common feature of integrator based updates is the need to compute recursively an integrator snippet $\mathsf{z} := \big(z, \psi(z), \psi^2(z), \ldots, \psi^T(z)\big)$, for a given mapping $\psi \colon \mathsf{Z} \to \mathsf{Z}$, of which only the endpoint $\psi^T(z)$ is used. This raises the question of recycling intermediate states, all the more so that computation of the snippet often involves quantities shared with the evaluation of U_n. In Example 1, for instance, expressions for $\nabla U_n(x)$ and $U_n(x)$ often involve the same computationally costly quantities and evaluation of the density $\mu_n(x)$ where $\nabla U_n(x)$ has already been evaluated is therefore often virtually free; consider for example $U(x) = x^\top \Sigma^{-1} x$ for a covariance matrix Σ, then $\nabla U(x) = 2\Sigma^{-1} x$.

In turn these quantities offer the promise of being able to exploit points used to generate the snippet while preserving accuracy of the estimators of interest, through importance sampling or reweighting. For example one may consider virtual HMC updates i.e. given $(z, A) \in \mathsf{Z} \times \mathscr{Z}$ and $k \in [\![P-1]\!]$ define

$$P_{n,k}(z, A) := \alpha_n(z; k)\mathbf{1}\{\psi_n^k(z) \in A\} + \bar{\alpha}_n(z; k)\mathbf{1}\{\sigma(z) \in A\}\,. \tag{6}$$

Noting that $\mu_n P_{n,k} = \mu_n$ one deduces that for $z \sim \mu_n$

$$\frac{1}{T+1}\sum_{k=0}^{T} \alpha_n(z; k) f \circ \psi_n^k(z) + \bar{\alpha}_n(z; k) f(z)\,,$$

is an unbiased estimator of $\mu(f)$. This post-processing procedure is akin to waste-recycling [14] and its generalizations but, crucially, does not affect the dynamic of the Markov chain generating the samples. An alternative algorithm exploiting the snippet fully, with an active effect on the dynamics, could use the following mixture of Markov chain transition kernels [23, 29, 22],

$$\mathfrak{M}_{n+1}(z, A) = \frac{1}{T+1}\sum_{k=0}^{T} \int \varpi_n(\mathrm{d}v') P_{n,k}(x, v'; A)\,,$$

which is also related to the strategy adopted in [11] with the extra chance algorithm. As we shall see our work shares the same objective but we adopt a strategy more closely related to [28]; see Section 5 for a more detailed discussion.

The present paper is concerned with exploring such recycling procedures in the context of SMC samplers, although the ideas we develop can be straightforwardly applied to particle filters in the context of state-space models and to MCMC as discussed later. The manuscript is organised as follows.

In Section 2 we first provide a high level description of particular instances of the class of algorithms considered and provide a justification through reinterpretation as standard SMC algorithms in Subsection 2.2. In Subsection 2.3 we discuss direct extensions of our algorithms, some of which we explore in the present manuscript. This work has links with related recent attempts in the literature [33, 14, 38]; these links are discussed and contrasted with our work in Subsection 2.4 where some of the motivations behind these algorithms are also discussed. Initial exploratory simulations demonstrating the interest of our approach are provided in Subsections 2.5 and 2.6.

In Section 3 we introduce the more general framework of Markov snippets Monte Carlo and associated formal justifications. Subsection 3.3 details the link with the scenario considered in Section 2. In Subsection 3.4 we provide general results facilitating the practical calculation of some of the Radon-Nikodym involved, highlighting why some of the usual constraints on mutation and backward kernels in SMC can be lifted here.

In Section 4 we provide elements of a theoretical analysis explaining expected properties of the algorithms proposed in this manuscript, although a fully rigorous theoretical analysis is beyond the present methodological contribution.

In Section 5 we explore the use of some of the ideas developed here in the context of MCMC algorithms and establish links with earlier suggestions, such as "windows of states" techniques proposed in the context of HMC [28, 29]. Notation, definitions and basic mathematical background can be found in Appendices A-B.

A Python implementation of the algorithms developed in this paper is available at https://github.com/MauroCE/IntegratorSnippets.

2 An introductory example

Assume interest is in sampling from a probability distribution μ on $(\mathsf{Z}, \mathscr{Z})$ as described above Example 1 using an SMC sampler relying on the leapfrog integrator of Hamilton's equations. As in the previous section we introduce

an interpolating sequence of distributions $\{\mu_n, n \in [\![0, P]\!]\}$ on $(\mathsf{Z}, \mathscr{Z})\}$ and assume for now the existence of densities for $\{\mu_n, n \in [\![0, P]\!]\}$ with respect to a common measure υ, say the Lebesgue measure on $\mathbb{R}^{2d}$, denoted $\mu_n(z) := \mathrm{d}\mu_n/\mathrm{d}\upsilon(z)$ for $z \in \mathsf{Z}$ and $n \in [\![0, P]\!]$ and denote ψ_n the corresponding integrator, which again is measure preserving in this setup.

2.1 An SMC-like algorithm

Primary interest in this paper is in algorithms of the type given in Alg. 2 and variations thereof; throughout $T \in \mathbb{N} \setminus \{0\}$.

Sample $z_0^{(i)} \overset{\text{iid}}{\sim} \mu_0$ for $i \in [\![N]\!]$.
for $n \in [\![P]\!]$ **do**
  for $i \in [\![N]\!]$ **do**
    for $k \in [\![0, T]\!]$ **do**
      Compute $z_{n-1,k}^{(i)} := \psi_n^k(z_{n-1}^{(i)})$ and

$$\bar{w}_{n,k}\big(z_{n-1}^{(i)}\big) := \frac{\mu_n\big(z_{n-1,k}^{(i)}\big)}{\mu_{n-1}\big(z_{n-1}^{(i)}\big)} = \frac{\mu_n \circ \psi_n^k\big(z_{n-1}^{(i)}\big)}{\mu_{n-1}\big(z_{n-1}^{(i)}\big)},$$

    end
  end
  for $j \in [\![N]\!]$ **do**
    Sample $[\![N]\!] \times [\![0, T]\!] \ni (b_j, a_j) \sim \mathrm{Cat}\big(\{\bar{w}_{n,k}(z_{n-1}^{(i)}), (i,k) \in [\![N]\!] \times [\![0, T]\!]\}\big)$
    Set $\bar{z}_n^{(j)} := (\bar{x}_{n-1}^{(j)}, \bar{v}_{n-1}^{(j)}) = z_{n-1,a_j}^{(b_j)}$
    Rejuvenate the velocities $z_n^{(j)} = (\bar{x}_{n-1}^{(j)}, v_n^{(j)})$ with $v_n^{(j)} \sim \varpi_n$.
  end
end

Algorithm 2: Unfolded Hamiltonian Snippet SMC algorithm

The SMC sampler-like algorithm in Alg. 2 therefore involves propagating N "seed" particles $\{z_{n-1}^{(i)}, i \in [\![N]\!]\}$, with a mutation mechanism consisting of the generation of N integrator snippets $\mathsf{z} := \big(z, \psi_n(z), \psi_n^2(z), \ldots, \psi_n^T(z)\big)$ started at every seed particle $z \in \{z_{n-1}^{(i)}, i \in [\![N]\!]\}$, resulting in $N \times (T+1)$ particles which are then whittled down to a set of N seed particles using a standard resampling scheme; after rejuvenation of velocities this yields the next generation of seed particles–this is illustrated in Fig. 1. This algorithm should be contrasted with standard implementations of SMC samplers where, after resampling, a seed particle normally gives rise to a single particle in the mutation step, in Fig. 1 the last state on the snippet. Intuitively validity of the algorithm follows from the fact that if $\big\{(z_{n-1}^{(i)}, 1), i \in [\![N]\!]\big\}$ represent μ_{n-1}, then $\big\{\big(z_{n-1,k}^{(i)}, \bar{w}_{n,k}(z_{n-1}^{(i)})\big)\big), (i,k) \in [\![N]\!] \times [\![0, T]\!]\big\}$ represents μ_n in the sense that for $f \colon \mathsf{Z} \to \mathbb{R}$ summable, one can use the approximation

$$\mu_n(f) \approx \sum_{i=1}^{N} \sum_{k=0}^{T} f \circ \psi_n^k\big(z_{n-1}^{(i)}\big) \frac{\mu_n \circ \psi_n^k\big(z_{n-1}^{(i)}\big)/\mu_{n-1}\big(z_{n-1}^{(i)}\big)}{\sum_{j=1}^{N} \sum_{l=0}^{T} \mu_n \circ \psi_n^l\big(z_{n-1}^{(j)}\big)/\mu_{n-1}\big(z_{n-1}^{(j)}\big)}, \tag{7}$$

where the self-normalization of the weights is only required in situations where the ratio $\mu_n(z)/\mu_{n-1}(z)$ is only known up to a constant. We provide justification for the correctness of Alg. 2 and the estimator (7) by recasting

the procedure as a standard SMC sampler targetting a particular sequence of distributions in Section 2.2 and using properties of mixtures. Direct generalizations are provided in Section 2.3.

2.2 Justification outline

We now outline the main ideas underpinning the theoretical justification of Alg. 2. Key to this is establishing that Alg. 2 is a standard SMC sampler targetting a particular sequence of probability distributions $\{\bar{\mu}_n, n \in [\![0, P]\!]\}$ from which samples can be processed to approximate expectations with respect to $\{\mu_n, n \in [\![0, P]\!]\}$. This has the advantage that no fundamentally new theory is required and that standard methodological ideas can be re-used in the present scenario, while the particular structure of $\{\bar{\mu}_n, n \in [\![0, P]\!]\}$ can be exploited for new developments. This section focuses on identifying $\{\bar{\mu}_n, n \in [\![0, P]\!]\}$ and establishing some of their important properties. Similar ideas are briefly touched upon in [14], but we will provide full details and show how these ideas can be pushed further, in interesting directions.

First for $(n, k) \in [\![0, P]\!] \times [\![0, T]\!]$ let $\psi_{n,k} \colon \mathsf{Z} \to \mathsf{Z}$ be measurable and invertible mappings, define $\mu_{n,k}(\mathrm{d}z) := \mu_n^{\psi_{n,k}^{-1}}(\mathrm{d}z)$, i.e. the distribution of $\psi_{n,k}^{-1}(z)$ when $z \sim \mu_n$. It is worth pointing out that invertibility of these mappings is not necessary, but facilitates interpretation throughout, as illustrated by the last statement about the distribution of $\psi_{n,k}^{-1}(z)$. Earlier we have focused on the scenario where $\psi_{n,k} = \psi_n^k$ for an integrator $\psi_n \colon \mathsf{Z} \to \mathsf{Z}$, but this turns out not to be a requirement, although it is our main motivation. Useful applications of this general perspective can be found in Subsection 2.3. Introduce the probability distributions on $\left([\![0, T]\!] \times \mathsf{Z}, \mathscr{P}([\![0, T]\!]) \otimes \mathscr{Z}\right)$

$$\bar{\mu}_n(k, \mathrm{d}z) = \frac{1}{T+1} \mu_{n,k}(\mathrm{d}z) \,,$$

for $n \in [\![0, P]\!]$. We will show that Alg. 2 can be interpreted as an SMC sampler targeting the sequence of marginal distributions on $\left(\mathsf{Z}, \mathscr{Z}\right)$

$$\bar{\mu}_n(\mathrm{d}z) = \frac{1}{T+1} \sum_{k=0}^{T} \mu_{n,k}(\mathrm{d}z) \,, \tag{8}$$

which we may refer to as a mixture, for $n \in [\![0, P]\!]$. Note that the conditional distribution is

$$\bar{\mu}_n(k \mid z) = \frac{1}{T+1} \frac{\mathrm{d}\mu_{n,k}}{\mathrm{d}\bar{\mu}_n}(z) \,, \tag{9}$$

which can be computed in most scenarios of interest.

Example 2. For $n \in [\![0, P]\!]$ assume the existence of a σ-finite dominating measure $\upsilon \gg \mu_n$, therefore implying the existence of a density $\mu_n(z) := \mathrm{d}\mu_n / \mathrm{d}\upsilon(z)$; υ could be the Lebesgue measure. Assuming that υ is $\psi_{n,k}$-invariant, or "volume preserving", then Lemma 32 implies, for $k \in [\![0, T]\!]$,

$$\mu_{n,k}(z) := \frac{\mathrm{d}\mu_{n,k}}{\mathrm{d}\upsilon}(z) = \mu_n \circ \psi_{n,k}(z) \text{ and } \bar{\mu}_n(z) := \frac{\mathrm{d}\bar{\mu}_n}{\mathrm{d}\upsilon}(z) = \frac{1}{T+1} \sum_{k=0}^{T} \mu_n \circ \psi_{n,k}(z) \,,$$

and therefore

$$w_{n,k}(z) := \frac{\mathrm{d}\mu_{n,k}}{\mathrm{d}\left(\sum_{l=0}^{T} \mu_{n,l}\right)}(z) = \frac{\mu_n \circ \psi_{n,k}(z)}{\sum_{l=0}^{T} \mu_n \circ \psi_{n,l}(z)} \,.$$

When υ is the Lebesgue measure and $\{\psi_{n,k}, k \in [\![P]\!]\}$ are not volume preserving, additional multiplicative Jacobian determinant-like terms may be required (see Lemma 32 and the additional requirement that $\{\psi_{n,k}, k \in [\![P]\!]\}$ be

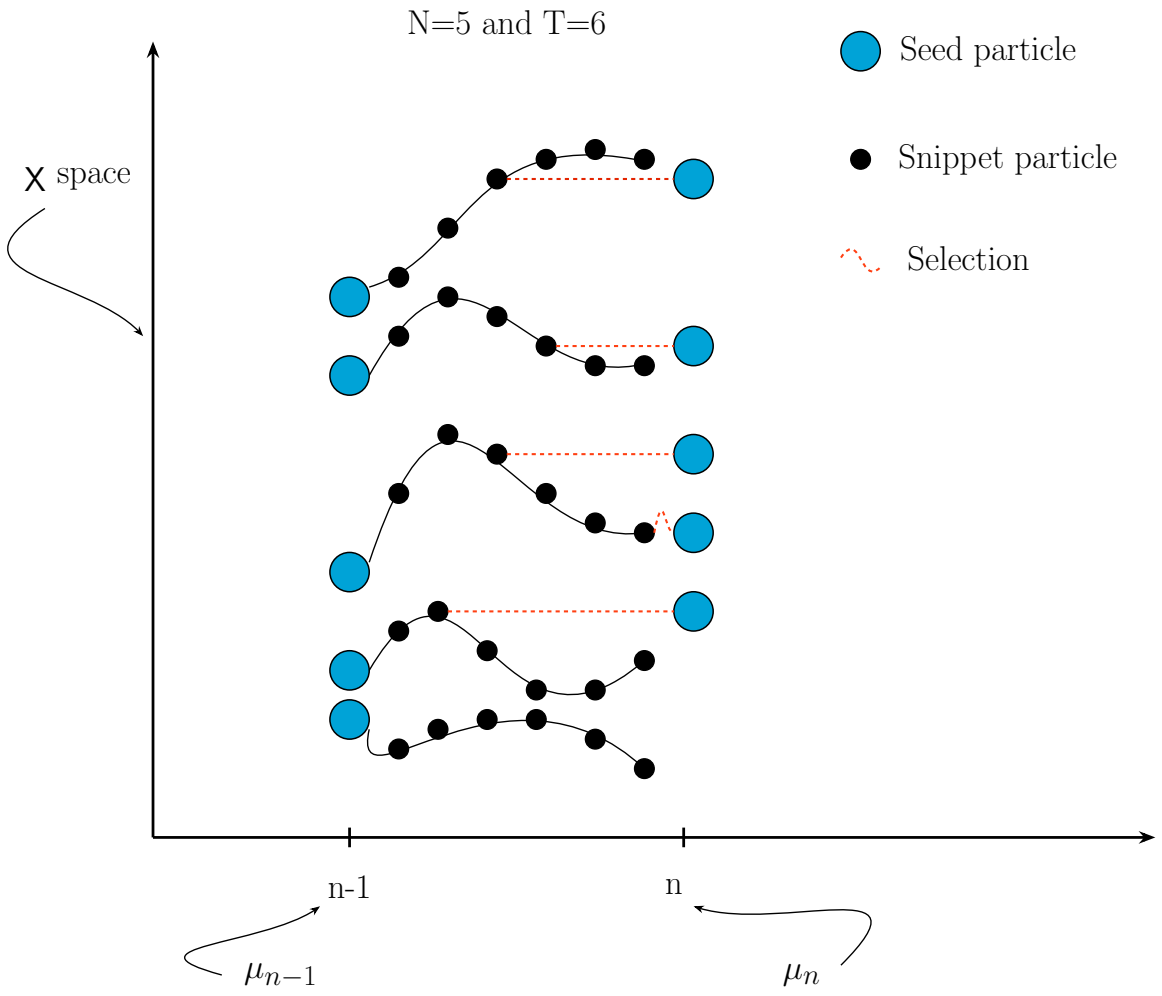

Figure 1: Illustration of the transition from μ_{n-1} to μ_n with integrator snippet SMC. A snippet grows (black dots) out of each seed particle (blue). The middle snippet gives rises through selection (dashed red) to two seed particles while the bottom snippet does not produce any seed particle.

differentiable). However this extra term may be more complex and require application specific treatment; adoption of the measure theoretic notation circumvents such peripheral considerations at this stage, which can be ignored until actual implementation of the algorithm.

A central point throughout this paper is how samples from μ_n can be used to obtain samples from the marginal $\bar{\mu}_n$ and vice-versa, thanks to the mixture structure relating the two distributions. Naturally, given $z \sim \mu_n$, sampling $k \sim \mathcal{U}(\llbracket 0, T \rrbracket)$ and returning $\psi_{n,k}^{-1}(z)$ yields a sample from the marginal $\bar{\mu}_n$. Now assuming $z \sim \bar{\mu}_n$ and then sampling $k \sim \bar{\mu}_n(\cdot \mid z)$ naturally yields $(k, z) \sim \bar{\mu}_n$ and hence intuitively $\psi_{n,k}(z) \sim \mu_n$. We now formally establish a more general result concerned with the estimation of expectations with respect to μ_n from samples from $\bar{\mu}_n$. For $f \colon \mathsf{Z} \to \mathbb{R}$ μ_n-integrable and $k \in \llbracket 0, T \rrbracket$, a change of variable (see Theorem 31) yields

$$\int f \circ \psi_{n,k}(z) \mu_{n,k}(\mathrm{d}z) = \int f \circ \psi_{n,k}(z) \mu_n^{\psi_{n,k}^{-1}}(\mathrm{d}z) = \int f(z) \mu_n(\mathrm{d}z) \,,$$

which formalizes the fact, at first sight of limited interest, that $\mu_n^{\psi_{n,k}^{-1}}$ is the distribution of $\psi_{n,k}^{-1}(z)$ for $z \sim \mu_n$ and therefore $\psi_{n,k} \circ \psi_{n,k}^{-1}(z) \sim \mu_n$. The relevance of this remark comes from the identity

$$\begin{aligned}
\int f(z) \mu_n(\mathrm{d}z) &= \frac{1}{T+1} \sum_{k=0}^{T} \int f \circ \psi_{n,k}(z) \mu_{n,k}(\mathrm{d}z) \\
&= \sum_{k=0}^{T} \int f \circ \psi_{n,k}(z) \bar{\mu}_n(k, \mathrm{d}z) \qquad (10) \\
&= \int \Big\{ \sum_{k=0}^{T} f \circ \psi_{n,k}(z) \bar{\mu}_n(k \mid z) \Big\} \bar{\mu}_n(\mathrm{d}z), \qquad (11)
\end{aligned}$$

which implies that samples from the mixture $\bar{\mu}_n$ can be used to unbiasedly estimate $\mu_n(f)$ thanks to an appropriate weighted average along the snippet $k \mapsto f \circ \psi_{n,k}(z)$. Using $f = \mathbf{1}_A$ for $A \in \mathscr{Z}$ formally establishes the earlier claim that if $(k, z) \sim \bar{\mu}_n$ then $\psi_{n,k}(z) \sim \mu_n$. Note that, as suggested by Example 2, construction of the estimator will only require evaluations of the density μ_n and function f at $z, \psi_{n,1}(z), \psi_{n,2}(z), \ldots, \psi_{n,T}(z)$.

We now turn to the description of an SMC algorithm targeting $\{\bar{\mu}_n, n \in \llbracket 0, P \rrbracket\}$, Alg. 3, and then establish that it is probabilistically equivalent to Alg. 2 in a sense made precise below. With $z = (x, v)$ for $n \in \llbracket P \rrbracket$ we introduce the following mutation kernel

$$\bar{M}_n(z, \mathrm{d}z') := \sum_{k=0}^{T} \bar{\mu}_{n-1}(k \mid z) \, R_n(\psi_{n-1,k}(z), \mathrm{d}z'), \quad R_n(z, \mathrm{d}z') := (\delta_x \otimes \varpi_{n-1})(\mathrm{d}z') \,, \qquad (12)$$

where we note that the refreshment kernel has the property that $\mu_{n-1} R_n = \mu_{n-1}$. One can show that the near optimal kernel $\bar{L}_{n-1} \colon \mathsf{Z} \times \mathscr{Z} \to [0, 1]$ is given for any $(z, A) \in \mathsf{Z} \times \mathscr{Z}$ by

$$\bar{L}_{n-1}(z, A) := \frac{\mathrm{d}\bar{\mu}_{n-1} \otimes \bar{M}_n(A \times \cdot)}{\mathrm{d}\bar{\mu}_{n-1} \bar{M}_n}(z) \,, \qquad (13)$$

and is well defined $\bar{\mu}_{n-1}\bar{M}_n$-a.s. (see Lemma 4, given at the end of this subsection) and satisfies the property that for any $A, B \in \mathscr{Z}$

$$\bar{\mu}_{n-1} \otimes \bar{M}_n(A \times B) = \int_B \bar{\mu}_{n-1} \bar{M}_n(\mathrm{d}z) \bar{L}_{n-1}(z, A) \,,$$

that is $\bar{L}_{n-1}(z, A)$ is a conditional probability of A given z for the joint probability distribution $\bar{\mu}_{n-1} \otimes \bar{M}_n$. With the assumption $\mu_{n-1} \gg \bar{\mu}_n$, from Lemma 4, the SMC sampler importance weights at step $n \in [\![P]\!]$ are

$$\bar{w}_n(z, z') := \frac{\mathrm{d}\bar{\mu}_n \widehat{\otimes} \bar{L}_{n-1}}{\mathrm{d}\bar{\mu}_{n-1} \otimes \bar{M}_n}(z, z') = \frac{\mathrm{d}\bar{\mu}_n}{\mathrm{d}\mu_{n-1}}(z') \,. \tag{14}$$

that is for $f \colon \mathsf{Z} \to \mathbb{R}$ such that $\bar{\mu}(|f|) < \infty$,

$$\int f(z') \frac{\mathrm{d}\bar{\mu}_n}{\mathrm{d}\mu_{n-1}}(z') \bar{\mu}_{n-1} \otimes \bar{M}_n\big(\mathrm{d}(z, z')\big) = \bar{\mu}(f) \,.$$

The corresponding standard SMC sampler is given in Alg. 3 where,

- the weighted particles $\{(\check{z}_n^{(i)}, 1), i \in [\![N]\!]\}$ represent $\bar{\mu}_n$ and $\{(\check{z}_{n,k}^{(i)}, w_{n,k}(\check{z}_n^{(i)})), (i, k) \in [\![N]\!] \times [\![0, T]\!]\}$ represent μ_n from (11),
- steps 3-8 correspond to sampling from the mutation kernel $\bar{M}_{n+1}$ in (12),
- $\{(z_n^{(i)}, 1), i \in [\![N]\!]\}$ represent μ_n,
- $\{\big(z_n^{(i)}, \bar{w}_{n+1}(z_n^{(i)})\big), i \in [\![N]\!]\}$ represent $\bar{\mu}_{n+1}$, and so do $\{(\check{z}_{n+1}^{(i)}, 1), i \in [\![N]\!]\}$.

1 sample $\check{z}_0^{(i)} \overset{\text{iid}}{\sim} \bar{\mu}_0 = \mu_0$ for $i \in [\![N]\!]$.
2 **for** $n = 0, \ldots, P-1$ **do**
3 **for** $i \in [\![N]\!]$ **do**
4 **for** $k \in [\![0, T]\!]$ **do**
5 compute $\check{z}_{n,k}^{(i)} = (\check{x}_{n,k}^{(i)}, \check{v}_{n,k}^{(i)}) := \psi_{n,k}(\check{z}_n^{(i)}), w_{n,k}\big(\check{z}_n^{(i)}\big)$
6 **end**
7 sample $a_i \sim \mathrm{Cat}\left(w_{n,0}\big(\check{z}_n^{(i)}\big), w_{n,1}\big(\check{z}_n^{(i)}\big), w_{n,2}\big(\check{z}_n^{(i)}\big), \ldots, w_{n,T}\big(\check{z}_n^{(i)}\big)\right)$
8 set $z_n^{(i)} = (\check{x}_{n,a_i}^{(i)}, v_n^{(i)})$ with $v_n^{(i)} \sim \varpi_n$
9 $$\bar{w}_{n+1}\big(z_n^{(i)}\big) = \frac{\mathrm{d}\bar{\mu}_{n+1}}{\mathrm{d}\mu_n}\big(z_n^{(i)}\big) \,,$$
10 **end**
11 **for** $j \in [\![N]\!]$ **do**
12 sample $b_j \sim \mathrm{Cat}\left(\bar{w}_{n+1}\big(z_n^{(1)}\big), \ldots, \bar{w}_{n+1}\big(z_n^{(N)}\big)\right)$
13 set $\check{z}_{n+1}^{(j)} = z_n^{(b_j)}$
14 **end**
15 **end**

Algorithm 3: Folded Hamiltonian Snippet SMC algorithm

Notice that we assume here $\psi_0 = \mathrm{Id}$, and hence that $\bar{\mu}_0 = \mu_0$, and that the weights $w_{n,k}$ appear as being computed twice in step 4 and step 9 when evaluating the resampling weights at the previous iteration, for the only

reason that it facilitates exposition. The identities (11) and (9) suggest, for $n \in [\![P]\!]$ the estimator of $\mu_n(f)$, for $f : \mathsf{Z} \to \mathbb{R}$ μ_n-integrable,

$$\begin{aligned}\check{\mu}_n(f) &= \frac{1}{N}\sum_{i=1}^{N}\sum_{k=0}^{T}\frac{1}{T+1}\frac{\mathrm{d}\mu_{n,k}}{\mathrm{d}\bar{\mu}_n}(\check{z}_n^{(i)}) f\circ\psi_{n,k}(\check{z}_n^{(i)}).\\ &= \frac{1}{N}\sum_{i=1}^{N}\sum_{k=0}^{T}\frac{\mu_n\circ\psi_{n,k}}{\sum_{l=0}^{T}\mu_n\circ\psi_{n,l}}(\check{z}_n^{(i)}) f\circ\psi_{n,k}(\check{z}_n^{(i)}),\end{aligned} \tag{15}$$

where the second line is correct under the assumptions of Example 2. Further, when $\mu_{n-1} \gg \mu_{n,k}$ for $k \in [\![0,T]\!]$ one can also write (11) for $f : \mathsf{Z} \to \mathbb{R}$ summable as

$$\mu_n(f) = \int \left\{\frac{1}{T+1}\sum_{k=0}^{T} f\circ\psi_{n,k}(z)\frac{\mathrm{d}\mu_{n,k}}{\mathrm{d}\mu_{n-1}}(z)\right\}\mu_{n-1}(\mathrm{d}z),$$

which can be estimated, using self-renormalization when required, with

$$\hat{\mu}_n(f) = \sum_{i=1}^{N}\sum_{k=0}^{T}\frac{\frac{\mathrm{d}\mu_{n,k}}{\mathrm{d}\mu_{n-1}}(z_{n-1}^{(i)})}{\sum_{j=1}^{N}\sum_{l=0}^{T}\frac{\mathrm{d}\mu_{n,l}}{\mathrm{d}\mu_{n-1}}(z_{n-1}^{(j)})} f\circ\psi_{n,k}(z_{n-1}^{(i)}), \tag{16}$$

therefore justifying the estimator suggested in (7) in the particular case where the conditions of Example 2 are satisfied, once we establish the equivalence of Alg. 3 and Alg. 2 in Proposition 3. In fact it can be shown (Proposition 3) that, with $\mathbb{E}_i$ referring to the expectation of the probability distribution underpinning Alg. i,

$$\mathbb{E}_3\left(\check{\mu}_n(f) \mid z_{n-1}^{(j)}, j \in [\![N]\!]\right) = \hat{\mu}_n(f),$$

a form of Rao-Blackwellization implying lower variance for $\hat{\mu}_n(f)$ while the two estimators share the same bias. The result is in fact stronger since it states that the estimators are convex ordered, a property which we however do not exploit further here. Interestingly a result we establish later in the paper, Proposition 15, suggests that the variance $\check{\mu}_n(f)$ is smaller than that of the standard Monte Carlo estimator assuming N samples $z_n^{(i)} \overset{\text{iid}}{\sim} \mu_n$, due to the control variate nature of integrator snippets estimators.

An interesting point is that computation of the weight $\mu_n \circ \psi_{n,k}/\bar{\mu}_n$ only requires knowledge of μ_n up to a normalizing constant, that is the estimator is unbiased if $\check{z}_n^{(i)} \sim \bar{\mu}_n$ for $i \in [\![N]\!]$ even if μ_n is not completely known, while the estimator (7) will most often require self-normalisation, hence inducing a bias.

We now provide the probabilistic argument justifying Alg. 2 and the shared notation $\{z_n^{(i)}, i \in [\![N]\!]\}$ in Alg. 2 and Alg. 3.

Proposition 3. *Alg. 2 and Alg. 3 are probabilistically equivalent. More precisely, letting $\mathbb{P}_i$ refer to the probability of Algorithm i for $i \in \{2,3\}$,*

1. *for $n \in [\![P-1]\!]$ the distributions of $\{z_n^{(i)}, i \in [\![N]\!]\}$ conditional upon $\{z_{n-1}^{(i)}, i \in [\![N]\!]\}$ in Alg. 3 and Alg. 2 are the same,*
2. *the joint distributions of $\{z_n^{(i)}, i \in [\![N]\!], n \in [\![0, P-1]\!]\}$ are the same under $\mathbb{P}_2$ and $\mathbb{P}_3$,*
3. *for $n \in [\![P]\!]$, any $f : \mathsf{Z} \to \mathbb{R}$ μ_n-integrable with $\check{\mu}_n(f)$ and $\hat{\mu}_n(f)$ as in (15) and (16),*

$$\mathbb{E}_3\left(\check{\mu}_n(f) \mid z_{n-1}^{(j)}, j \in [\![N]\!]\right) = \hat{\mu}_n(f).$$

Proof of Proposition 3. In Alg. 3 for any $A \in \mathscr{Z}^{\otimes N}$ we have with $(b_1, \dots, b_N)$ the random vector taking values in $[\![N]\!]^N$ involved in the resampling step,

$$\begin{aligned}\mathbb{P}_3\big(z_n^{(1:N)} \in A \mid z_{n-1}^{(1:N)}\big) =& \mathbb{E}_3\left\{\mathbf{1}_A\{(\check{x}_{n,a_i}^{(i)}, v_n^{(i)}), i \in [\![N]\!]\}\mathbf{1}\{\check{z}_n^{(i)} = z_{n-1}^{(b_i)}, i \in [\![N]\!]\} \mid z_{n-1}^{(j)}, j \in [\![N]\!]\right\} \\ =& \mathbb{E}_3\left\{\mathbf{1}_A\{(x_{n-1,a_i}^{(b_i)}, v_n^{(i)}), i \in [\![N]\!]\} \mid z_{n-1}^{(j)}, j \in [\![N]\!]\right\}.\end{aligned}$$

Letting for $(\alpha, \beta) \in [\![0, T]\!]^N \times [\![N]\!]^N$

$$E_A(\alpha, \beta) = E_A(\alpha, \beta, z_{n-1}^{(1:N)}) := \int \mathbf{1}_A\{(x_{n-1,\alpha_i}^{(\beta_i)}, v_n^{(i)}), i \in [\![N]\!]\}\varpi_n^{\otimes N}(\mathrm{d}v_n^{(1:N)})$$

from the tower property we have

$$\mathbb{E}_3\left\{\mathbf{1}_A\{(x_{n-1,a_i}^{(b_i)}, v_n^{(i)}), i \in [\![N]\!]\} \mid z_{n-1}^{(j)}, b_j = \beta_j, j \in [\![N]\!]\right\} = \sum_{\alpha \in [\![0,T]\!]^N} E_A(\alpha, \beta) \prod_{i=1}^N \frac{1}{T+1} \frac{\mathrm{d}\mu_{n,\alpha_i}}{\mathrm{d}\bar{\mu}_n}\big(z_{n-1}^{(\beta_i)}\big).$$

Since

$$\mathbb{P}_3\left(b_1 = \beta_1, \dots, b_N = \beta_N \mid z_{n-1}^{(i)}, i \in [\![N]\!]\right) = \left(\sum_{j=1}^N \frac{\mathrm{d}\bar{\mu}_n}{\mathrm{d}\mu_{n-1}}\big(z_{n-1}^{(j)}\big)\right)^{-N} \prod_{i=1}^N \frac{\mathrm{d}\bar{\mu}_n}{\mathrm{d}\mu_{n-1}}\big(z_{n-1}^{(\beta_i)}\big)$$

we deduce

$$\begin{aligned}\mathbb{P}_3\big(z_n^{(1:N)} \in A \mid z_{n-1}^{(1:N)}\big) \propto& \sum_{\alpha,\beta} E_A(\alpha, \beta, z_{n-1}^{(1:N)}) \prod_{i=1}^N \frac{\mathrm{d}\bar{\mu}_n}{\mathrm{d}\mu_{n-1}}\big(z_{n-1}^{(\beta_i)}\big) \prod_{i=1}^N \frac{\mathrm{d}\mu_{n,\alpha_i}}{\mathrm{d}\bar{\mu}_n}\big(z_{n-1}^{(\beta_i)}\big) \\ =& \sum_{\alpha,\beta} E_A(\alpha, \beta, z_{n-1}^{(1:N)}) \prod_{i=1}^N \frac{\mathrm{d}\bar{\mu}_n}{\mathrm{d}\mu_{n-1}}\big(z_{n-1}^{(\beta_i)}\big) \frac{\mathrm{d}\mu_{n,\alpha_i}}{\mathrm{d}\bar{\mu}_n}\big(z_{n-1}^{(\beta_i)}\big) \\ =& \sum_{\alpha,\beta} E_A(\alpha, \beta, z_{n-1}^{(1:N)}) \prod_{i=1}^N \frac{\mathrm{d}\mu_{n,\alpha_i}}{\mathrm{d}\mu_{n-1}}\big(z_{n-1}^{(\beta_i)}\big).\end{aligned}$$

Now notice that

$$\begin{aligned}\mathbb{P}_2\left(\bar{z}_n^{(j)} = z_{n-1,\alpha_j}^{(\beta_j)} \mid z_{n-1}^{(1:N)}\right) &= \frac{\frac{\mathrm{d}\mu_{n,\alpha_j}}{\mathrm{d}\mu_{n-1}}\big(z_{n-1}^{(\beta_j)}\big)}{\sum_{i,k} \frac{\mathrm{d}\mu_{n,k}}{\mathrm{d}\mu_{n-1}}\big(z_{n-1}^{(i)}\big)} \\ &\propto \frac{\frac{\mathrm{d}\mu_{n,\alpha_j}}{\mathrm{d}\mu_{n-1}}\big(z_{n-1}^{(\beta_j)}\big)}{\sum_{i=1}^N \frac{\mathrm{d}\bar{\mu}_n}{\mathrm{d}\mu_{n-1}}\big(z_{n-1}^{(i)}\big)},\end{aligned}$$

and the first statement follows from conditional independence of $(b_1, a_1), \dots, (b_N, a_N)$ and the fact that $\mathbb{P}_2\big(z_n^{(1:N)} \in A \mid \bar{z}_{n-1}^{(1:N)}\big) = E_A(\alpha, \beta, \bar{z}_{n-1}^{(1:N)})$. Since by construction $\mathbb{P}_3\big((z_0^{(1)}, \dots, z_0^{(N)}) \in A\big) = \mathbb{P}_2\big((z_0^{(1)}, \dots, z_0^{(N)}) \in A\big)$ for any $A \in \mathscr{Z}$ the second statement follows from a standard Markov chain argument. Now for $g \colon \mathsf{Z} \to \mathbb{R}$ and $\iota \in [\![N]\!]$

$$\begin{aligned}\mathbb{E}_3\big(g(\check{z}_n^{(\iota)}) \mid z_{n-1}^{(j)}, j \in [\![N]\!]\big) =& \mathbb{E}_3\big(g(z_{n-1}^{(b_\iota)}) \mid z_{n-1}^{(j)}, j \in [\![N]\!]\big) \\ =& \sum_{i=1}^N g(z_{n-1}^{(i)}) \frac{\frac{\mathrm{d}\bar{\mu}_n}{\mathrm{d}\mu_{n-1}}\big(z_{n-1}^{(i)}\big)}{\sum_{j=1}^N \frac{\mathrm{d}\bar{\mu}_n}{\mathrm{d}\mu_{n-1}}\big(z_{n-1}^{(j)}\big)}.\end{aligned}$$

Therefore for any $k \in [\![0, T]\!]$,

$$\mathbb{E}_3\left(\frac{\mathrm{d}\mu_{n,k}}{\mathrm{d}\bar{\mu}_n}(\check{z}_n^{(\iota)}) f \circ \psi_{n,k}(\check{z}_n^{(\iota)}) \mid z_{n-1}^{(j)}, j \in [\![N]\!]\right) = \sum_{i=1}^{N} \frac{\mathrm{d}\mu_{n,k}}{\mathrm{d}\bar{\mu}_n}(z_{n-1}^{(i)}) f \circ \psi_{n,k}(z_{n-1}^{(i)}) \frac{\frac{\mathrm{d}\bar{\mu}_n}{\mathrm{d}\mu_{n-1}}(z_{n-1}^{(i)})}{\sum_{j=1}^{N} \frac{\mathrm{d}\bar{\mu}_n}{\mathrm{d}\mu_{n-1}}(z_{n-1}^{(j)})}$$
$$= \sum_{i=1}^{N} \frac{\frac{\mathrm{d}\mu_{n,k}}{\mathrm{d}\mu_{n-1}}(z_{n-1}^{(i)})}{\sum_{j=1}^{N}\sum_{l=0}^{T} \frac{\mathrm{d}\mu_{n,l}}{\mathrm{d}\mu_{n-1}}(z_{n-1}^{(j)})} f \circ \psi_{n,k}(z_{n-1}^{(i)}) .$$

The third statement follows. □

Since the justification of the latter interpretation of the algorithm is straightforward, as a standard SMC sampler targeting instrumental distributions $\{\bar{\mu}_n, n \in [\![0, P]\!]\}$, and allows for further easy generalisations we will adopt this perspective in the remainder of the manuscript for simplicity.

This reinterpretation also allows the use of known facts about SMC sampler algorithms. For example it is well known that the output of SMC samplers can be used to estimate unbiasedly unknown normalising constants by virtue of the fact that, in the present scenario,

$$\prod_{n=0}^{P-1}\left[\frac{1}{N}\sum_{i=1}^{N}\frac{\mathrm{d}\bar{\mu}_{n+1}}{\mathrm{d}\mu_n}(z_n^{(i)})\right]$$

has expectation 1 under $\mathbb{P}_3$. Now assume that the densities of $\{\mu_n, n \in [\![0, P]\!]\}$ are known up to a constant only, say $\mathrm{d}\mu_n/\mathrm{d}\upsilon(z) = \tilde{\mu}_n(z)/Z_n$ and $\upsilon \gg \upsilon^{\psi_n^{-k}}$ for $k \in [\![0, T]\!]$, then

$$Z_n = Z_n \int \mu_n^{\psi_{n,k}^{-1}}(\mathrm{d}z) = \int \tilde{\mu}_n \circ \psi_n^k(z) \frac{\mathrm{d}\upsilon^{\psi_{n,k}^{-1}}}{\mathrm{d}\upsilon}(z)\, \upsilon(\mathrm{d}z) ,$$

and the measure $Z_n\bar{\mu}_n(\mathrm{d}z)$ shares the same normalising constant as $\tilde{\mu}_n(z)\upsilon(\mathrm{d}z)$,

$$Z_n = \int Z_n \bar{\mu}_n(\mathrm{d}z) = \sum_{k=1}^{T} \frac{1}{T} \int \tilde{\mu}_n \circ \psi_n^k(z) \frac{\mathrm{d}\upsilon^{\psi_{n,k}^{-1}}}{\mathrm{d}\upsilon}(z)\, \upsilon(\mathrm{d}z) .$$

Consequently

$$\prod_{n=0}^{P-1}\left[\frac{1}{N(T+1)}\sum_{i=1}^{N}\sum_{k=0}^{N}\frac{\tilde{\mu}_{n+1} \circ \psi_{n+1,k}}{\tilde{\mu}_n}(z_n^{(i)}) \frac{\mathrm{d}\upsilon^{\psi_{n+1,k}^{-1}}}{\mathrm{d}\upsilon}(z_n^{(i)})\right],$$

is an unbiased estimator of Z_P/Z_0.

The results of the following lemma can be deduced from Lemma 8 but we provide more direct arguments for the present scenario.

Lemma 4. *Assume $\mu_{n-1} \gg \bar{\mu}_n$, $\mu_{n-1}R_n = \mu_{n-1}$ and let $\bar{M}_n \colon \mathsf{Z} \times \mathscr{Z} \to [0, 1]$ be as in (12). Then for $n \in [\![P]\!]$,*

1. *$\bar{\mu}_{n-1}\bar{M}_n = \mu_{n-1}$,*
2. *there exists $\bar{L}_{n-1} \colon \mathsf{Z} \times \mathscr{Z} \to [0, 1]$ in (13) such that* for any $A, B \in \mathscr{Z}$

$$\bar{\mu}_{n-1} \otimes \bar{M}_n(A \times B) = \int_B \bar{\mu}_{n-1}\bar{M}_n(\mathrm{d}z)\bar{L}_{n-1}(z, A) ,$$

3. the importance weight in (14) is well defined and

$$\bar{w}_n(z, z') := \frac{\mathrm{d}\bar{\mu}_n}{\mathrm{d}\mu_{n-1}}(z')\,.$$

Proof. The first statement, follows from the definition in (12) of $\bar{M}_n$ and the identity (11). The second statement is a consequence of the following classical arguments, justifying Bayes' rule. For $A \in \mathscr{Z}$ fixed, consider the measure

$$\mathscr{Z} \ni B \mapsto \bar{\mu}_{n-1} \otimes \bar{M}_n\big(A \times B\big) \leq \bar{\mu}_{n-1} \otimes \bar{M}_n\big(\mathsf{Z} \times B\big) = \bar{\mu}_{n-1}\bar{M}_n(B)\,,$$

implying $\bar{\mu}_{n-1}\bar{M}_n \gg \bar{\mu}_{n-1} \otimes \bar{M}_n\big(A \times \cdot\big)$ from which we deduce the existence of a Radon-Nikodym derivative such that

$$\bar{\mu}_{n-1} \otimes \bar{M}_n\big(A \times B\big) = \int_B \frac{\mathrm{d}\bar{\mu}_{n-1} \otimes \bar{M}_n\big(A \times \cdot\big)}{\mathrm{d}\bar{\mu}_{n-1}\bar{M}_n}(z')\,\bar{\mu}_{n-1}\bar{M}_n(\mathrm{d}z')\,.$$

For $(z, A) \in \mathsf{Z} \times \mathscr{Z}$, we let

$$\bar{L}_{n-1}(z, A) := \frac{\mathrm{d}\bar{\mu}_{n-1} \otimes \bar{M}_n\big(A \times \cdot\big)}{\mathrm{d}\bar{\mu}_{n-1}\bar{M}_n}(z)\,,$$

and note that almost surely $\bar{L}_{n-1}(z, A) \in [0, 1]$ and $\bar{L}_{n-1}(z, \mathsf{Z}) = 1$. For the third statement note that from the second statement, for $A, B \in \mathscr{Z}$ we have $\bar{\mu}_{n-1} \otimes \bar{M}_n\big(A \times B\big) = \bar{\mu}_{n-1}\bar{M}_n \hat{\otimes} \bar{L}_{n-1}\big(A \times B\big)$ and from Fubini's and the $\pi - \lambda$ theorem [7, Theorems 3.1 and 3.2] $\bar{\mu}_{n-1} \otimes \bar{M}_n$ and $\bar{\mu}_{n-1}\bar{M}_n \hat{\otimes} \bar{L}_{n-1}$ are probability distributions coinciding on $\mathscr{Z} \otimes \mathscr{Z}$. To conclude proof of the third statement, for $f\colon \mathsf{Z}^2 \to [0, 1]$ measurable, we successively apply the definition of the Radon-Nikodym derivative, Fubini's theorem, use that $\mu_{n-1} \gg \bar{\mu}_n$, the first statement of this lemma, Fubini again, the second statement

$$\begin{aligned}
\int f(z, z')\frac{\mathrm{d}\bar{\mu}_n}{\mathrm{d}\mu_{n-1}}(z, z')\bar{\mu}_{n-1} \otimes \bar{M}_n\big(\mathrm{d}(z, z')\big) &= \int f(z, z')\frac{\mathrm{d}\bar{\mu}_n}{\mathrm{d}\mu_{n-1}}(z')\,\bar{\mu}_{n-1}\bar{M}_n \hat{\otimes} \bar{L}_n\big(\mathrm{d}(z, z')\big) \\
&= \int f(z, z')\frac{\mathrm{d}\bar{\mu}_n}{\mathrm{d}\mu_{n-1}}(z')\,\mu_{n-1}(\mathrm{d}z')\bar{L}_n\big(z', \mathrm{d}z\big) \\
&= \int f(z, z')\bar{\mu}_n(\mathrm{d}z')\bar{L}_n\big(z', \mathrm{d}z\big)
\end{aligned}$$

therefore establishing that $\bar{\mu}_{n-1} \otimes \bar{M}_n$-almost surely

$$\frac{\mathrm{d}\bar{\mu}_{n-1}\bar{M}_n \hat{\otimes} \bar{L}_{n-1}}{\mathrm{d}\bar{\mu}_{n-1} \otimes \bar{M}_n}(z, z') = \frac{\mathrm{d}\bar{\mu}_n}{\mathrm{d}\mu_{n-1}}(z, z')\,.$$

□

2.3 Direct extensions

It should be clear that the algorithm we have described lends itself to numerous generalizations, which we briefly discuss below.

There is no reason to limit the number of snippets arising from a seed particle to one, therefore offering the possibility to take advantage of parallel machines. For example the velocity of a given seed particle can be refreshed multiple times, resulting in partial copies of the seed particle from which integrator snippets can be grown.

The main scenario motivating this work, corresponds to the choice, for $n \in [\![0, P]\!]$, of$\{\psi_{n,k} = \psi_n^k, k \in [\![0, T]\!]\}$ for a given $\psi_n \colon \mathsf{Z} \to \mathsf{Z}$. As should be apparent from the theoretical justification this can be replaced by a general family of invertible mappings $\{\psi_{n,k} \colon \mathsf{Z} \to \mathsf{Z}, k \in [\![0, T]\!]\}$ where the $\psi_{n,k}$'s are now not required to be measure preserving in general, in which case the expression for $w_{n,k}(z)$ may involve additional terms of the "Jacobian" type. These mappings may correspond to integrators other than those of Hamilton's equations but may be more general deterministic mappings and T may have no temporal meaning and only represent the number of deterministic transformations used in the algorithm. More specifically, asssuming that $\upsilon \gg \mu_n$ and $\upsilon \gg \upsilon^{\psi_{n,k}^{-1}}$ for some σ-finite dominating measure υ and letting $\mu_n(z) := \mathrm{d}\mu_n/\mathrm{d}\upsilon(z)$ the required weights are now of the form (see Lemmas 32-34)

$$\bar{w}_{n,k}(z) := \frac{1}{T+1} \frac{\mu_n \circ \psi_{n,k}(z)}{\mu_{n-1}(z)} \frac{\mathrm{d}\upsilon^{\psi_{n,k}^{-1}}}{\mathrm{d}\upsilon}(z) \,.$$

Again when υ is the Lebesgue measure and $\psi_{n,k}$ a diffeomorphism, then $\mathrm{d}\upsilon^{\psi_{n,k}^{-1}}/\mathrm{d}\upsilon$ is the absolute value of the determinant of the Jacobian of $\psi_{n,k}$. Non uniform weights may be ascribed to each of these transformations in the definition of $\bar{\mu}$ (24). A more useful application of this generality is in the situation where it is believed that using multiple integrators, specialised in capturing different geometric features of the target density, could be beneficial. Hereafter we simplify notation by setting $\nu \leftarrow \mu_n$ and $\mu \leftarrow \mu_{n+1}$. For the purpose of illustration consider two distinct integrators ψ_i, $i \in [\![2]\!]$ (again n disappears from the notation) we wish to use, each for $T_i \in \mathbb{N}$ time steps with proportions $\gamma_i \geq 0$ such that $\gamma_1 + \gamma_2 = 1$. Again with $\mu = \pi \otimes \varpi$ define the mixture

$$\bar{\mu}(i, k, \mathrm{d}z) := \frac{\gamma_i}{T_i + 1} \mu^{\psi_{i,k}^{-1}}(\mathrm{d}z) \mathbf{1}\{k \in [\![0, T_i]\!]\} \,,$$

which still possesses the fundamental property that if $z \sim \bar{\mu}$ (resp. $(i, z) \sim \bar{\mu}$), then with $(i, k) \sim \bar{\nu}(k, i \mid z)$ (resp. $k \sim \bar{\nu}(k \mid i, z)$) we have $\psi_{i,k}(z) \sim \mu$. It is possible to aim to sample from $\bar{\mu}(\mathrm{d}z)$, in which case the pair (i, k) plays the rôle played by k in the earlier simpler scenario. However, in order to introduce the sought persistency, that is use either ψ_1 or ψ_2 when constructing a snippet, we focus on the scenario where the pair (i, z) plays the rôle played by z earlier. In other words the target distribution is now $\bar{\mu}(i, \mathrm{d}z)$, which is to $\mu(i, \mathrm{d}z) := \gamma_i \cdot \mu(\mathrm{d}z)$ what $\bar{\mu}_n(\mathrm{d}z)$ in (24) was to $\mu_n(\mathrm{d}z)$. The mutation kernel corresponding to (24) is given by

$$\bar{M}_{\nu,\mu}(i, z; j, \mathrm{d}z') := \sum_{k=0}^{T_i} \bar{\nu}(k \mid i, z) R(i, \psi_{i,k}(z); j, \mathrm{d}z') \,,$$

with now the requirement that $\bar{\nu} R(i, \mathrm{d}z) = \nu(i, \mathrm{d}z) = \gamma_i \cdot \nu(\mathrm{d}z)$. A sensible choice seems to be $R(i, z; j, \mathrm{d}z') = \gamma_j \cdot R_0(z, \mathrm{d}z')$ with $\nu R_0(\mathrm{d}z') = \nu(\mathrm{d}z')$. With these choices using the near-optimal backward kernel simple substitutions $(i, z) \leftarrow z$ and $(j, z') \leftarrow z'$ yields

$$\frac{\mathrm{d}\bar{\mu} \widehat{\otimes} \bar{L}_{\nu,\mu}}{\mathrm{d}\bar{\nu} \otimes \bar{M}_{\nu,\mu}}(i, z; j, z') = \frac{\mathrm{d}\bar{\mu}}{\mathrm{d}\nu}(j, z') \,,$$

and justifies the earlier abstract presentation.

Another direct extension consists of generalizing the definition of $\bar{\mu}_n$ by assigning non-uniform and possibly state-dependent weights to the integrator snippet particles as follows

$$\bar{\mu}_n(k, \mathrm{d}z) = \omega_{n,k} \circ \psi_{n,k}(z) \, \mu_{n,k}(\mathrm{d}z) \,,$$

with $\omega_{n,k} \colon \mathsf{Z} \to \mathbb{R}_+$ for $k \in [\![0, T]\!]$ and such that $\sum_{k=0}^{T} \omega_{n,k}(z) = 1$ for any $z \in \mathsf{Z}$; this should be contrasted with

(8). Such choices ensure that the central identity (11) can be generalized as follows

$$\sum_{k=0}^{T}\int f\circ\psi_n^k(z)\frac{\omega_{n,k}\circ\psi_{n,k}(z)\cdot w_{n,k}(z)}{\sum_{l=0}^{T}\omega_{n,l}\circ\psi_{n,l}(z)w_{n,l}(z)}\bar{\mu}_n(\mathrm{d}z)=\sum_{k=0}^{T}\int f\circ\psi_{n,k}(z)\,\bar{\mu}_n(k\mid z)\bar{\mu}_n(\mathrm{d}z)$$
$$=\int f(z)\mu_n(\mathrm{d}z)\,.$$

Note the analogy with the identity behind umbrella sampling [37, 39]. In the light of the developments of Subsection 4.3.2, a potentially useful choice could be $\omega_{n,k}(z)\propto\|z-\psi_{n,k}^{-1}(z)\|$, which however requires additional computations as

$$\omega_{n,k}\circ\psi_{n,k}(z)=\frac{\|\psi_{n,k}(z)-z\|}{\sum_{l=0}^{T}\|\psi_{n,k}(z)-\psi_{n,k}\circ\psi_{n,l}^{-1}(z)\|}$$

which may require computation of additional states for $l>k$.

We leave exploration of some of these generalisations for future work, although we think that the choice of the leapfrog integrator is particularly natural and attractive here.

2.4 Rational and computational considerations

Our initial motivation for this work was that computation of $\{z_{n,k},k\in[\![T]\!]\}$ and $\{w_{n,k},k\in[\![T]\!]\}$ most often involve common quantities, leading to negligible computational overhead, and reweighting offers the possibility to use all the states of a snippet in a simulation algorithm rather than the endpoint only. There are other reasons for which this approach may be beneficial.

The benefit of using all states along integrator snippets in sampling algorithms has been noted in the literature. For example, in the context of Hamiltonian integrators, with $H_n(z):=-\log\mu_n(z)$ and $\psi_{n,k}=\psi_n^k$, it is known that for $z\in\mathsf{Z}$ the mapping $k\mapsto H_n\circ\psi_n^k(z)$ is typically oscillatory, motivating for example the x-tra chance algorithm of [11]. The "windows of state" strategy of [28, 29] effectively makes use of the mixture $\bar{\mu}$ as an instrumental distribution and improved performance is noted, with performance seemingly improving with dimension on particular problems; see Section 5 for a more detailed discussion. Further averaging is well known to address scenarios where components of x_t evolve on different scales and no choice of a unique integration time $\tau:=T\times\varepsilon$ can accommodate all scales [29, 22, section 3.2]; averaging addresses this issue effectively, see Example 27. We also note that, keeping $T\times\varepsilon$ constant, in the limit as $\varepsilon\to 0$ the average in (14) corresponds to the numerical integration, Riemann like, along the contour of $H_n(z)$, effectively leading to some form of Rao-Blackwellization of this contour; this is discussed in detail in Section 4.

Another benefit we have observed with the SMC context is the following. Use of a WF-SMC strategy involves comparing particles within each Markov snippet arising from a single seed particle, while our strategy involves comparing all particles across snippets, which proves highly beneficial in practice. This seems to bring particular robustness to the choice of the integrator parameters ε and T and can be combined with highly robust adaptation schemes taking advantage of the population of samples, in the spirit of [21, 22]; see Subsections 2.5 and 2.6.

At a computational level we note that, in contrast with standard SMC or WF-SMC implementations relying on an MH mechanism, the construction of integrator snippets does not involve an accept reject mechanism, therefore removing control flow operations and enabling lockstep computations on GPUs – actual implementation of our algorithms on such architectures is, however, left to future work.

Finally, when using integrators of Hamilton's equations we naturally expect similar benefits to those enjoyed by HMC. Let $d\in\mathbb{N}$ and let $\mathsf{X}=\mathbb{R}^d$. We know that in certain scenarios [4, 10], the distributions $\{\mu_{n,d},d\in\mathbb{N},n\in[\![T(d)]\!]\}$ are such that for $n\in\mathbb{N}$, $\log\left({}^{\mu_{n,d}\circ\psi_{n,d}^k(z)}/_{\mu_{n,d}(z)}\right)\to_{d\to\infty}\mathcal{N}(\mu_n,\sigma_n^2)$, that is the weights do not degenerate to

zero or one: in the context of SMC this means that the part of the importance weight (14) arising from the mutation mechanism does not degenerate. Further, with an appropriate choice of schedule, i.e. sequence $\{\mu_{n,d}, n \in [\![T(d)]\!]\}$ for $d \in \mathbb{N}$, ensures that the contribution ${}^{\mu_{n,d}(z)}/_{\mu_{n-1,d}(z)}$ to the importance weights (14) is also stable as $d \to \infty$. As shown in [4, 2, 10], while direct important sampling may require an exponential number of samples as d grows, the use of such a schedule may reduce complexity to a polynomial order.

2.5 Numerical illustration: logistic regression

In this section, we consider sampling from the posterior distribution of a logistic regression model, focusing on the compution of the normalising constant. We follow [14] and consider the sonar dataset, previously used in [13]. With intercept terms, the dataset has responses $y_i \in \{-1, 1\}$ and covariates $z_i \in \mathbb{R}^p$, where $p = 61$. The likelihood of the parameters $x \in \mathsf{X} := \mathbb{R}^p$ is then given by

$$L(x) = \prod_i^n F(z_i^\top x \cdot y_i), \tag{17}$$

where $F(x) := 1/(1 + \exp(-x))$. We ascribe x a product of independent normal distributions of zero mean as a prior, with standard deviation equal to 20 for the intercept and 5 for the other parameters. Denote $p(dx)$ the prior distribution of x, we define a sequence of tempered distributions of densities of the form $\pi_n(x) \propto p(dx)L(x)^{\lambda_n}$ for $\lambda_n \colon [\![0, P]\!] \to [0, 1]$ non-decreasing and such that $\lambda_0 = 0$ and $\lambda_P = 1$. We apply both Hamiltonian Snippet-SMC and the implementation of waste-free SMC of [14] and compare their performance.

For both algorithms, we set the total number of particles at each SMC step to be $N(T + 1) = 10,000$. For the waste-free SMC, the Markov kernel is chosen to be a random-walk Metropolis-Hastings kernel with covariances adaptively computed as $2.38^2/d\ \hat{\Sigma}$, where $\hat{\Sigma}$ is the empirical covariance matrix obtained from the particles in the previous SMC step. For the Hamiltonian Snippet-SMC algorithm, we set ψ_n to be the one-step leap-frog integrator with stepsize ε, $U_n(x) = -\log(\pi_n(x))$ and ϖ a $\mathcal{N}(0, \mathrm{Id})$. To investigate the stability of our algorithm, we ran Hamiltonian Snippet SMC with $\varepsilon = 0.05, 0.1, 0.2$ and 0.3. For both algorithms, the temperatures λ_n are adaptively chosen so that the effective sample size (ESS) of the current SMC step will be αESS_{max}, where ESS_{max} is the maximum ESS achievable at the current step. In our experiments, we have chosen $\alpha = 0.3, 0.5$ and 0.7 for both algorithms.

Performance comparison

Figure 2 shows the boxplots of estimates of the logarithm of the normalising constant obtained from both algorithms, for different choices of N and ε for the Hamiltonian Snippet SMC algorithm. The boxplots are obtained by running both algorithms 100 times for different of the algorithm parameters, with $\alpha = 0.5$ in all setups. Several points are worth observing. For a suitably choice of ε, the Hamiltonian Snippet SMC algorithm can produce stable and consistent estimates of the normalising constant with $10,000$ particles at each iteration. On the other hand, however, the waste-free SMC algorithm fails to produce accurate results for the same computational budget. It is also clear that with larger values of N (meaning smaller value of T and hence shorter snippets), the waste-free SMC algorithm produces results with larger biases and variability. For Hamiltonian Snippet SMC algorithm, the results are stable both for short and long snippets when ε is equal to 0.1 or 0.2. Another point is that when $\varepsilon = 0.05$ or 0.3, the Hamiltonian Snippet SMC algorithm becomes unstable with short (i.e. $\varepsilon = 0.05$) and long (i.e. $\varepsilon = 0.3$). Possible reasons are for too small a stepsize the algorithm is not able to explore the target distribution efficiently, resulting in unstable performances. On the other hand, when the stepsize is too large, the leapfrog integrator becomes unstable, therefore affecting the variability of the weights; this is a common problem of HMC algorithms. Hence, a long trajectory will result in detoriate estimations. Hence, to obtain the best performance, one should find a way of tuning the stepsize and trajectory length along with the SMC steps.

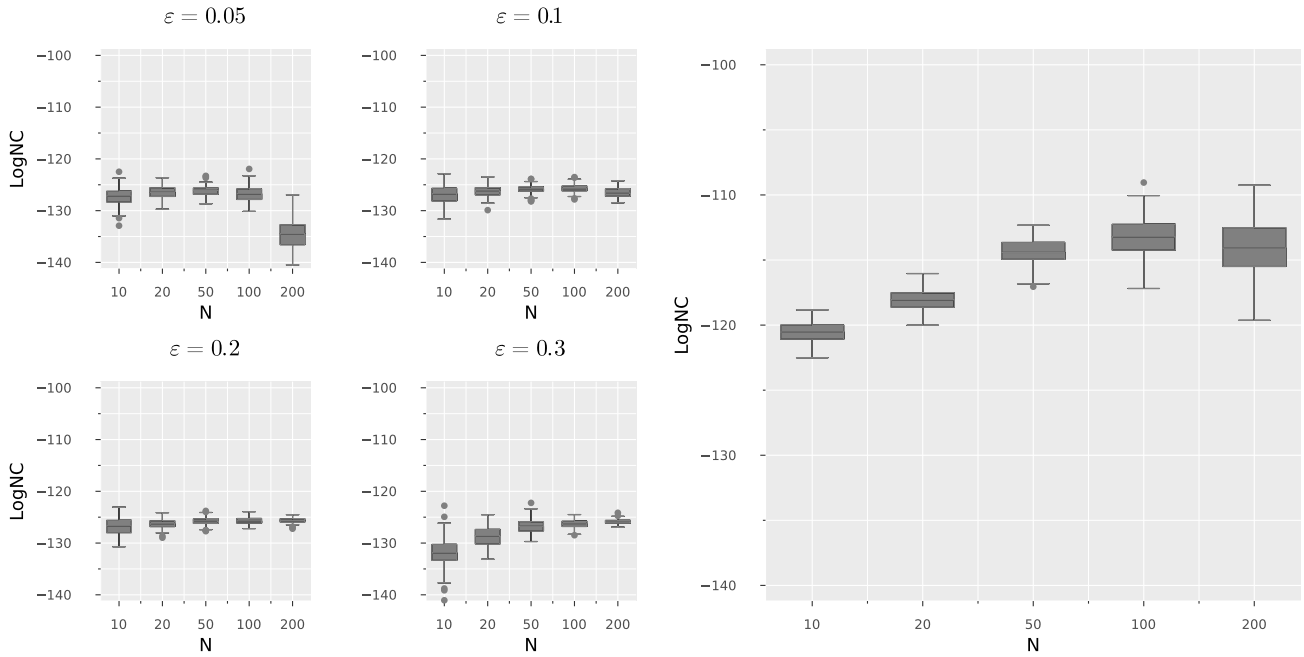

Figure 2: Estimates of the normalising constant (in log scale) obtained from both Hamiltonian Snippet SMC and waste-free SMC algorithm. Left: Estimate obtained from Hamiltonian Snippet SMC algorithm with different values of ε. Right: Estimates obtained from waste-free SMC algorithm.

In Figure 3 we display boxplots of the estimates of the posterior expectations of the mean of all coefficients, i.e. $\mathbb{E}_{\pi_P}(d^{-1}\sum_{i=1}^{d} x_i)$. This quantity is referred to as the *mean of marginals* in [14] and we use this terminology. One can see that the same properties can be seen from the estimations of the mean of marginals, with the unstability problems exacerbated with small and large stepsizes.

Computational Cost

In this section, we compare the running time of both algorithms. Since the calculations of the potential energy and its gradient often share common intermediate steps, we can recycle these to save computational cost. As the waste-free SMC also requires density evaluations, the Hamiltonian Snippet SMC algorithm will not require significant additional computations. Figure 4 shows boxplots of the simulation time of both algorithms from 100 runs. The simulations were run on an Apple M1-Pro CPU with 16G of memory. One can see that in comparison to the waste-free SMC the additional computational time is only marginal for the Hamiltonian Snippet SMC algorithm and mostly due to the larger memory needed to store the intermediate values.

2.6 Numerical illustration: simulating from filamentary distributions

We now illustrate the interest of integrator snippets in a scenario where the target distribution possesses specific geometric features. Specifically, we focus here on distributions concentrated around a manifold $\mathcal{M} \subset \mathsf{X} = \mathbb{R}^d$ defined as the zero level set of a smooth function $\ell : \mathbb{R}^d \to \mathbb{R}^m$

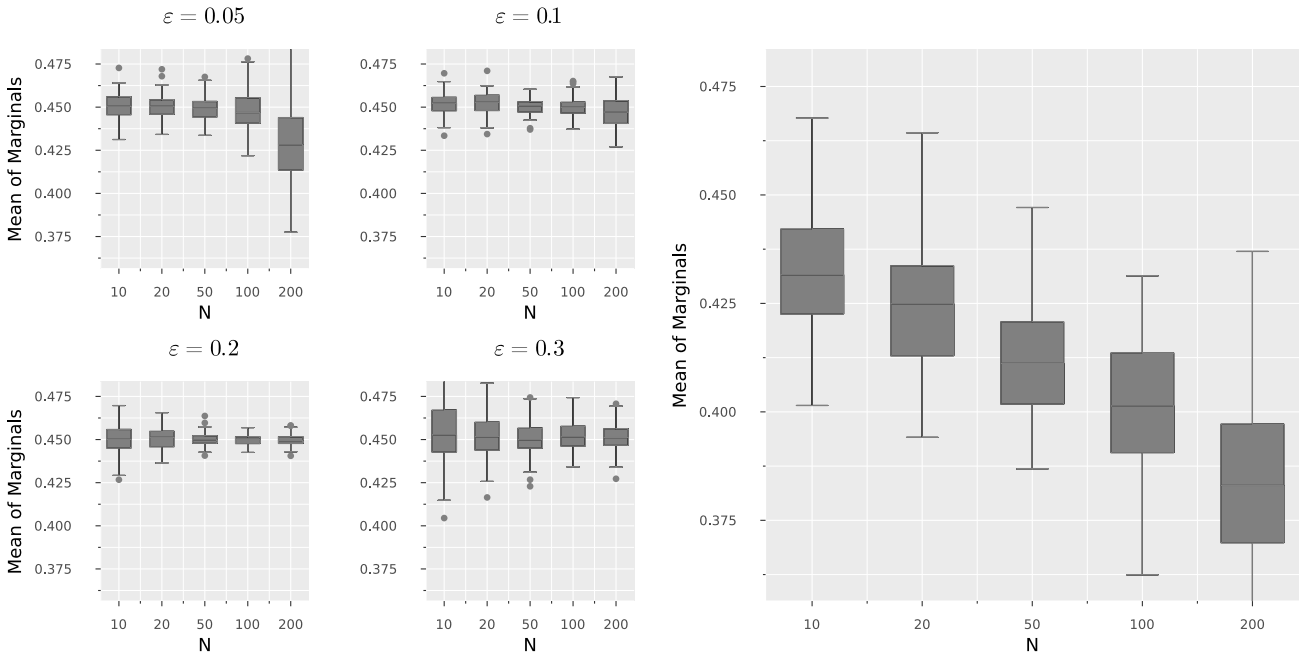

Figure 3: Estimates of the mean of marginals obtained from both Hamiltonian Snippet SMC and the waste-free SMC algorithms. Left: Estimate obtained from the Hamiltonian Snippet SMC algorithm with difference values of ε. Right: Estimates obtained from the waste-free SMC algorithm.

$$\mathcal{M} := \{x \in \mathsf{X} : \ell(x) = 0\},$$

sometimes referred to as filamentary distributions [12, 24, 25]. Such distributions arise naturally in various scenarios, including inverse problems or generalisations of the Gibbs sampler [12, 24, 25] or as a relaxation of problems where the support of the distribution of interest is included in $\mathcal{M}$ or for example generalisation of the Gibbs sampler through disintegration [12]. Such is the case for ABC methods in Bayesian statistics [3]. Assume that π is a probability density with respect to the Lebesgue measure defined on X and for $\epsilon > 0$ consider a "kernel" function $k_\epsilon(u) := \epsilon^{-m} k(u/\epsilon)$ where $k : \mathbb{R}^m \to \mathbb{R}_+$ and define the probability density

$$\pi_\epsilon(x) \propto k_\epsilon \circ \ell(x)\, \pi(x)\,.$$

The corresponding probability distribution can be thought of as an approximation of the probability distribution of density $\pi_0(x) \propto \pi(x)\mathbf{1}\{x \in \mathcal{M}\}$ with respect to the Hausdorff measure on $\mathcal{M}$. Typical choices for the kernel are $k(u) = \mathbf{1}\{\|u\| \leq 1\}$ or $k(u) = \mathcal{N}(u; 0, \mathbf{I}_m)$. Strong anisotropy may result from such constraints and make exploration of the support of such distributions awkward for standard Monte Carlo algorithms. This is illustrated in Fig. 2.6 where a standard MH-HMC algorithms is used to sample from π_ϵ defined on $\mathbb{R}^2$, for three values $\epsilon = 0.5, 0.1, 0.05$, and performance is observed to deteriorate as ϵ decreases. The samples produced are displayed in blue: for $\epsilon = 0.5$ HMC-MH mixes well and explores the support rapidly, but for $\epsilon = 0.1$ the chain gets stuck in a narrow region of π_ϵ at the bottom, near initialisation, while for $\epsilon = 0.05$, no proposal is ever accepted in the experiment.

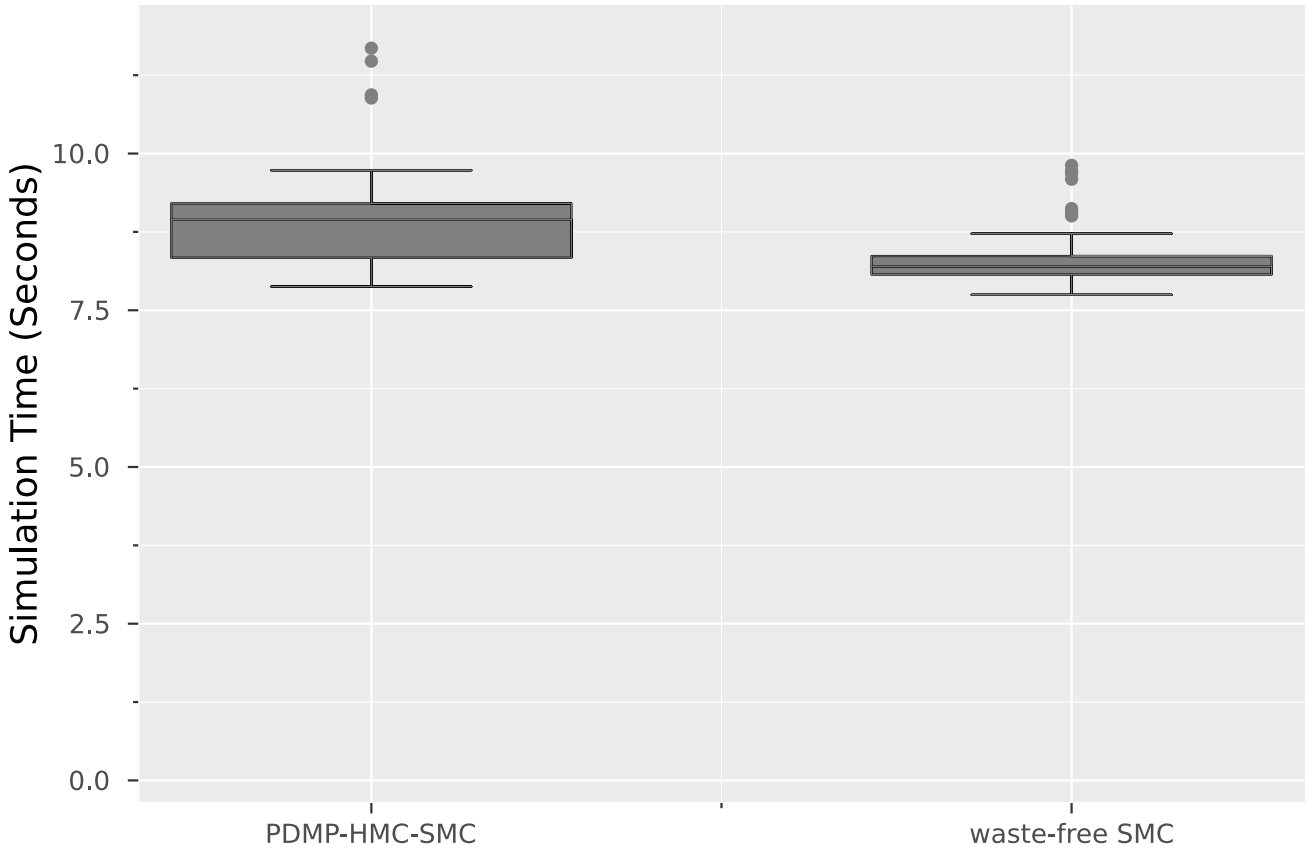

Figure 4: Simulation times for both algorithms. Left: Simulation time for Hamiltonian Snippet SMC algorithm, with $N = 100, \varepsilon = 0.2, \alpha = 0.5$. Right:Simulation times for the waste-free SMC with $N = 100$

To illustrate the properties of integrator snippet techniques we consider the following toy example. For $\epsilon > 0$ and $d \in \mathbb{N}_*$ let

$$\pi_\epsilon(x) \propto \frac{1}{\epsilon^m} \mathbf{1}\{\|\ell(x)\| \leq \epsilon\} \mathcal{N}(x; 0, \mathbf{I}_d) \quad \text{with} \quad \ell(x) = x^\top \Sigma^{-1} x - c\,,$$

for Σ a $d \times d$ symmetric positive matrix and $c \in \mathbb{R}$, that is we consider the restriction of a standard normal distribution around an ellipsoid defined by $\ell(x) = 0$.

In order to explore the support of the target distribution we use two mechanisms based on reflections either through tangent hyperplanes of equicontours of $\ell(x)$ or through the corresponding orthogonal complements. More specifically for $x \in \mathsf{X}$ such that $\nabla \ell(x) \neq 0$ let $n(x) := \nabla \ell(x) / \|\nabla \ell(x)\|$ and define the tangential HUG (THUG) and symmetrically-normal HUG (SNUG) as $\psi_{/\!/} := {}_{\mathrm{A}}\psi \circ \mathrm{b} \circ {}_{\mathrm{A}}\psi$, , and $\psi_\perp := {}_{\mathrm{A}}\psi \circ (-\mathrm{b}) \circ {}_{\mathrm{A}}\psi$ respectively, with ${}_{\mathrm{A}}\psi$ and b as in Example 1 and (25). Both updates can be understood as discretisations of ODEs and are volume preserving. Intuitively for $(x, v) \in \mathsf{Z}$ with $v_{/\!/} = v_{/\!/}(x + \varepsilon v) := n(x + \varepsilon v) n(x + \varepsilon v)^\top v$ for $\varepsilon > 0$ and $v_\perp = v - v_{/\!/}$ we have $\psi_{/\!/}(x, v) = (x + 2\varepsilon v_{/\!/}, v_{/\!/} - v_\perp)$ and $\psi_\perp(x, v) = (x + 2\varepsilon v_\perp, v_\perp - v_{/\!/})$. This is illustrated in Fig. 6 for $d = 2$. Further, for an initial state $z_0 = (x_0, v_0) \in \mathsf{Z}$, trajectories of the first component of $k \mapsto \psi_{/\!/}^k(z_0)$ remain close to $\ell^{-1}(\{\ell(x_0)\})$ while

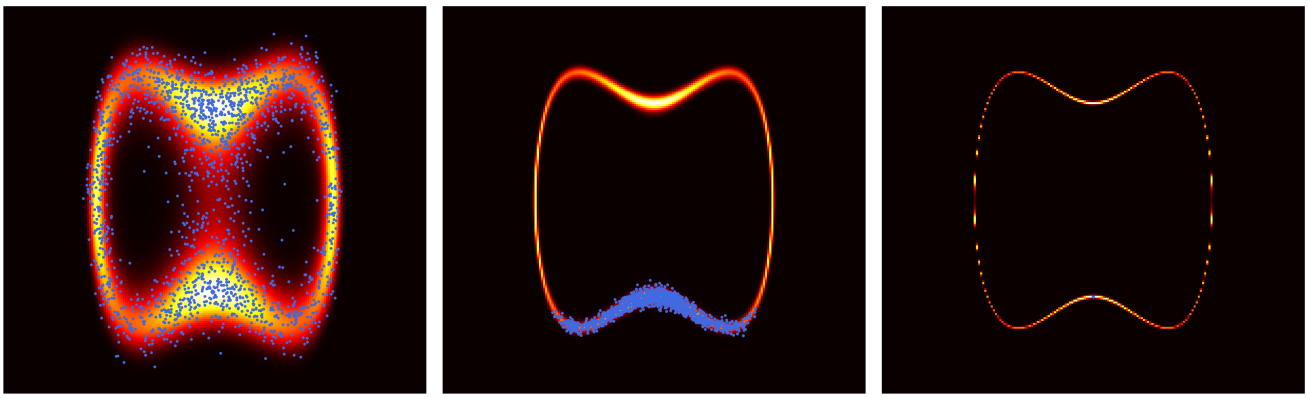

Figure 5: Heat map of $\pi_\epsilon(x) \propto \mathcal{N}(y \mid F(x), \epsilon^2)\mathcal{N}(x \mid 0, \mathbf{I}_2)$ with $F(x_1, x_2) := x_2^2 + x_1^2\left(x_1^2 - \frac{1}{2}\right)$, defined on $\mathbb{R}^2$, for $y \in \mathbb{R}$ and $\epsilon = 0.5, 0.1, 0.05$ (left to right) superimposed with 2000 samples (in blue) obtained from an MH-HMC with step size 0.05 and involving $T = 20$ leapfrog steps.

$k \mapsto \psi_\perp^k(z_0)$ follows the gradient field $x \mapsto \nabla\ell(x)$ and leads to hops across equicontours.

The sequence of target distributions on $(\mathsf{Z}, \mathcal{Z})$ we consider is of the form

$$\mu_n(\mathrm{d}z) \propto \pi_{\epsilon_n}(\mathrm{d}x)\mathcal{N}(\mathrm{d}v; 0, \mathbf{I}_d), \qquad \epsilon_n > 0, n \in [\![0, P]\!],$$

and the Integrator Snippet SMC is defined through the mixture, for $\alpha \in [0, 1]$,

$$\bar{\mu}_n(\mathrm{d}z) = \frac{\alpha}{T+1}\sum_{k=0}^{T} \mu_n^{\psi_{/\!/}^{-k}}(\mathrm{d}z) + \frac{1-\alpha}{T+1}\sum_{k=0}^{T} \mu_n^{\psi_\perp^{-k}}(\mathrm{d}z)\,.$$

We compare performance of integrator snippet with an SMC Sampler relying on a mutation kernel M_n consisting of a mixture of two updates targetting μ_{n-1} each iterating T times a MH kernel applying integrator $\psi_{/\!/}$ or $\psi_\perp$ once after refreshing the velocity; the backward kernels are chosen to correspond to the default choices discussed earlier. In the experiments below we set $d = 50$, $c = 12$, and Σ is the diagonal matrix alternating 1's and 0.1's along the diagonal. We used $N = 5000$ particles, sampled from $\mathcal{N}(0, \mathbf{I}_d)$ at time zero and ϵ_0 is set to the maximum distance of these particles in the ℓ domain from 0 and $\alpha = 0.8$ for both algorithms. We compared the two samplers across three metrics; all results are averaged over 20 runs.

Robustness and Accuracy: we fix the step size for SNUG to $\varepsilon_\perp = 0.1$ and run both algorithms for a grid of values of T and $\varepsilon_{/\!/}$ using a standard adaptive scheme based on ESS to determine $\{\epsilon_n, n \in [\![P]\!]\}$ until a criterion described below is satisfied and retain the final tolerance achieved, recorded in Fig. 7. As a result the terminal value ϵ_P and computational costs are different for both algorithms: the point of this experiment is only to demonstrate robustness and potential accuracy of Integrator Snippets. Both the SMC sampler and Integrator Snippet are stopped when the average probability of leaving the seed particle drops below 0.01.

Our proposed algorithm consistently achieves a two order magnitude smaller final value of ϵ and is more robust to the choice of stepsize.

Variance: for this experiment we set $T = 50$ steps, $\varepsilon_{/\!/} = 0.01$ and determine and compare the variances of the estimates of the mean of π_{ϵ_P} for the final SMC step, for which $\epsilon_P = 1 \times 10^{-7}$. To improve comparison, and in particular ensure comparable computational costs, both algorithms share the same schedule $\{\epsilon_n, n \in [\![P]\!]\}$, determined adaptively by the SMC algorithm in a separate pre-run.

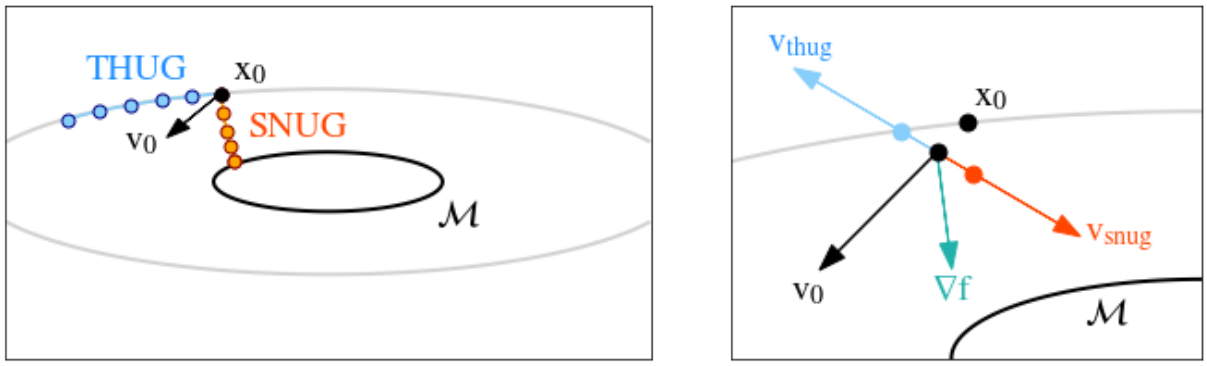

Figure 6: Black line: manifold of interest $\mathcal{M} = \ell^{-1}(0)$, grey line: level set of $\ell^{-1}(\{\ell(x_0)\})$ the ℓ-level set x_0 belongs to. Left: THUG trajectory (blue) remains close to $\ell^{-1}(\{\ell(x_0)\})$, SNUG trajectory (orange) explores different contours of $\ell(x)$. Right: $v_{\text{thug}} = v_{/\!/} - v_{\perp}$ and $v_{\text{snug}} = -(v_{/\!/} - v_{\perp})$.

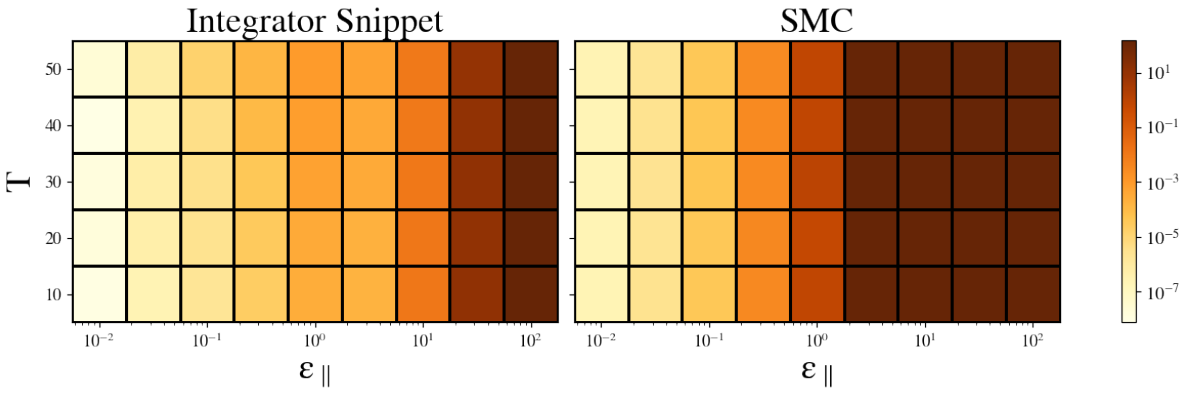

Figure 7: Final tolerances achieved by the two algorithms, averaged over 20 runs.

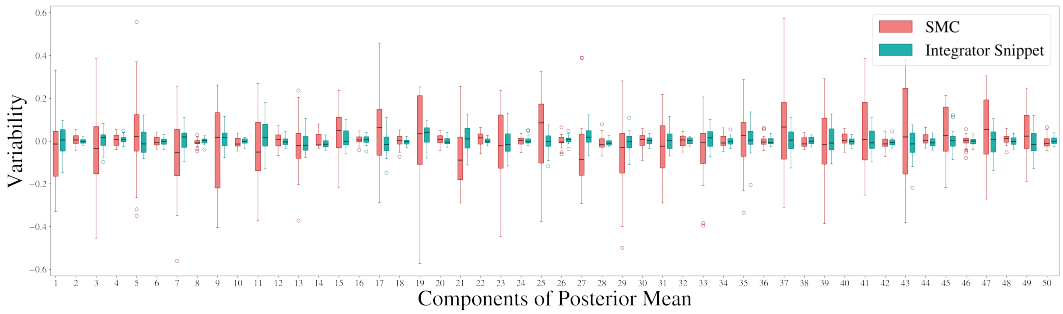

Figure 8: Variability of the estimates of the mean of π_ϵ for the SMC sampler and the Integrator Snippet, using estimator $\hat{\mu}_P(f)$, for the test function $f(x) = x$. Integrator snippet is able uniformly achieve smaller variance, in particular on the odd components corresponding to larger variances in Σ.

	Integrator Snippet	SMC
ESJD/s	$\mathbf{5.3 \times 10^{-5}}$ $(\pm 5.9 \times 10^{-6})$	1.2×10^{-7} $(\pm 3.7 \times 10^{-9})$
ESJD-THUG/s	$\mathbf{6.6 \times 10^{-5}}$ $(\pm 7.5 \times 10^{-6})$	1.7×10^{-7} $(\pm 4.5 \times 10^{-9})$
ESJD-SNUG/s	$\mathbf{2.7 \times 10^{-4}}$ $(\pm 3.2 \times 10^{-5})$	1.6×10^{-20} $(\pm 6.0 \times 10^{-21})$

Table 1: $d^{-1} \sum_{i=1}^{d} \mathrm{ESJD}_n(f_i)$ normalised by time for Integrator Snippet and an SMC sampler.

The results are reported as componentwise boxplots in Fig. 2.6 where we observe a significant variance reduction for comparable computational cost.

Efficiency: we report the Expected Squared Jump Distance (ESJD) as a proxy of distance travelled by the two algorithms. For Integrator Snippets, it is possible to estimate this quantity as follows for a function $f : \mathsf{Z} \to \mathbb{R}$

$$\mathrm{ESJD}_n(f) \approx \sum_{i=1}^{N} \sum_{k=0}^{T} \sum_{l=k+1}^{T} \left(f(Z_{n-1,l}^{(i)}) \bar{W}_{n,l}^{(i)} - f(Z_{n-1,k}^{(i)}) \bar{W}_{n,k}^{(i)} \right)^2, \qquad \bar{W}_{n,k}^{(i)} = \frac{\bar{w}_{n,k}(Z_{n-1}^{(i)})}{\sum_{l=0}^{T} \bar{w}_{n,l}(Z_{n-1,l}^{(i)})}.$$

We report the average of this metric in Table 1 for the functions $f_i(x) = x_i$ $i \in [\![d]\!]$, normalised by total runtime in seconds for all particles (first row), for the particles using the THUG update (second row) and those using the SNUG update (third row), with standard deviation shown in parenthesis. Our proposed algorithm is several orders of magnitude more efficient in the exploration of π_ϵ than its SMC counterpart which, thanks to its ability to take full advantage of all the intermediary states of the snippet. This is in contrast with the SMC sampler which creates trajectories of random length.

2.7 Links to the literature

Alg. 2, and its reinterpretation Alg. 3, are reminiscent of various earlier contributions and we discuss here parallels and differences. This work was initiated in [42] and pursued and extended in [18].

Readers familiar with the "waste-free SMC" algorithm (WF-SMC) of [14] where, similarly to Alg. 2, seed particles are extended with an MCMC kernel leaving μ_{n-1} invariant at iteration n, yielding $N \times (T+1)$ particles subsequently

whittled down to N new seed particles. A first difference is that while generation of an integrator snippet can be interpreted as applying a sequence of deterministic Markov kernels (see Subsection 3.3 for a detailed discussion) what we show is that the mutation kernels involved do not have to leave μ_{n-1} invariant; in fact we show in Section 3 that this indeed is not a requirement, therefore offering more freedom. Further, it is instructive to compare our procedure with the following two implementations of WF-SMC using an HMC kernel (4) for the mutation. A first possibility for the mutation stage is to run T steps of an HMC update in sequence, where each update uses one integrator step. Assuming no velocity refreshment along the trajectory this would lead to the exploration of a random number of states of our integrator snippet due to the accept/reject mechanism involved; incorporating refreshment or partial refreshment would similarly lead to a random number of useful samples. Alternatively one could consider a mutation mechanism consisting of an HMC build around T integration steps and where the endpoint of the trajectory would be the mutated particle; this would obviously mean discarding $T-1$ potentially useful candidates. To avoid possible reader confusion, we note apparent typos in [14, Proposition 1], provided without proof, where the intermediate target distribution of the SMC algorithms, $\bar{\mu}_n$ in our notation (see (18)), seems improperly stated. The statement is however not used further in the paper.

Our work shares, at first sight, similarities with [33, 38], but differs in several respects. In the discrete time setup considered in [38] a mixture of distributions similar to ours is also introduced. Specifically for a sequence $\{\omega_k \geq 0, k \in \mathbb{Z}\}$ such that $\#\{\omega_k \neq 0, k \in \mathbb{Z}\} < \infty$ and $\sum_{l\in\mathbb{Z}} \omega_k = 1$, the following mixture is considered (in the notation of [38] their transformation is $T = \psi^{-1}$ and we stick to our notation for the sake of comparison and avoid confusion with our T),

$$\bar{\mu}(\mathrm{d}z) = \sum_{k\in\mathbb{Z}} \omega_k \mu_k(\mathrm{d}z)\,.$$

The intention is to use the distribution $\bar{\mu}$ as an importance sampling proposal to estimate expectations with respect to μ,

$$\begin{aligned}\int f(z)\frac{\mathrm{d}\mu}{\mathrm{d}\bar{\mu}}(z)\bar{\mu}(\mathrm{d}z) &= \sum_{k\in\mathbb{Z}} \omega_k \int f(z)\frac{\mathrm{d}\mu}{\mathrm{d}\bar{\mu}}(z)\mu^{\psi^{-k}}(\mathrm{d}z)\\ &= \sum_{k\in\mathbb{Z}} \omega_k \int f\circ\psi^{-k}(z)\frac{\mathrm{d}\mu}{\mathrm{d}\bar{\mu}}\circ\psi^{-k}(z)\mu(\mathrm{d}z)\\ &= \sum_{k\in\mathbb{Z}} f\circ\psi^{-k}(z)\omega_k\frac{\mu\circ\psi^{-k}(z)}{\sum_{l\in\mathbb{Z}}\omega_l\,\mu\circ\psi^{l-k}(z)}\mu(\mathrm{d}z)\,,\end{aligned}$$

where the last line holds when $\upsilon \gg \mu$, $\upsilon^{\psi} = \upsilon$ and $\mu(z) = \mathrm{d}\mu/\mathrm{d}\upsilon(z)$. The rearrangement on the second and third line simply capture the fact noted earlier in the present paper that with $z \sim \mu$ and $k \sim \mathrm{Cat}\big(\{\omega_l : l \in \mathbb{Z}\}\big)$ then $\psi^{-k}(z) \sim \bar{\mu}$. As such NEO relies on generating samples from μ first, typically exact samples, which then undergo a transformation and are then properly reweighted. In contrast we aim to sample from a mixture of the type $\bar{\mu}$ directly, typically using an iterative method, and then exploit the mixture structure to estimate expectations with respect to μ. Note also that some of the $\psi^{l-k}(z)$ terms involved in the renormalization may require additional computations beyond terms for which $\omega_{l-k} \neq 0$. Our approach relies on a different important sampling identity

$$\int f\circ\psi^{k}(z)\frac{\mathrm{d}\mu^{\psi^{-k}}}{\mathrm{d}\bar{\mu}}(z)\bar{\mu}(\mathrm{d}z) = \mu(f)\,.$$

We note however that the conformal Hamiltonian integrator used in [33, 38] could be used in our framework, which we leave for future investigations. Their NEO-MCMC targets a "posterior distribution" $\mu'(\mathrm{d}z) = Z^{-1}\mu(\mathrm{d}z)L(z)$

and inspired by the identity

$$\int f(z)L(z)\frac{\mathrm{d}\mu}{\mathrm{d}\bar{\mu}}(z)\bar{\mu}(\mathrm{d}z) = \sum_{k\in\mathbb{Z}}\omega_k \int f\circ\psi^{-k}(z)L\circ\psi_k^{-1}(z)\frac{\mathrm{d}\mu}{\mathrm{d}\bar{\mu}}\circ\psi^{-k}(z)\mu(\mathrm{d}z)\,,$$

which is an expection of $\bar{f}(k,z) = f\circ\psi^{-k}(z)$ with respect to

$$\bar{\mu}'(k,\mathrm{d}z) \propto \mu(\mathrm{d}z)\cdot\omega_k\frac{\mathrm{d}\mu}{\mathrm{d}\bar{\mu}}\circ\psi^{-k}(z)\cdot L\circ\psi^{-k}(z)$$

and they consider algorithms targetting $\bar{\mu}'(\mathrm{d}z)$ which relying on exact samples from μ to construct proposals, resulting in a strategy which shares the weaknesses of an independent MH algorithm.

Much closer in spirit to our work is the "window of states" idea proposed in [28, 29] in the context of HMC algorithms. Although the MCMC algorithms of [28, 29] ultimately target μ and are not reversible, they involve a reversible MH update with respect to $\bar{\mu}$ and additional transitions permitting transitions from μ to $\bar{\mu}$ and $\bar{\mu}$ to μ; see Section 5 for full details.

A link we have not explored in the present manuscript is that to normalizing flows [36, 26] and related literature. In particular the ability to tune the parameter of the leapfrog integrator suggests that our methodology lends itself learning normalising flows. We finally note an analogy of such approaches with umbrella sampling [37, 39], although the targetted mixture is not of the same form, and the "Warp Bridge Sampling" idea of [27].

3 Sampling Markov snippets

In this section we develop the Markov snippet framework, largely inspired by the WF-SMC framework of [14] but provide here a detailed derivation following the standard SMC sampler framework [16] which allows us to consider much more general mutation kernels; integrator snippet SMC is recovered as a particular case. Importantly we provide recipes to compute some of the quantities involved using simple criteria (see Lemma 9 and Corollary 10) which allow us to consider unusual scenarios such as in Subsection 3.4.

3.1 Markov snippet SMC sampler or waste free SMC with a difference

Given a sequence $\{\mu_n, n\in[\![0,P]\!]\}$ of probability distributions defined on a measurable space $(\mathsf{Z},\mathscr{Z})$ introduce the sequence of distributions defined on $\left([\![0,T]\!]\times\mathsf{Z}^{T+1}, \mathscr{P}([\![0,T]\!])\otimes\mathscr{Z}^{\otimes(T+1)}\right)$ such that for any $(n,k,\mathsf{z})\in[\![0,P]\!]\times[\![0,T]\!]\times\mathsf{Z}$

$$\bar{\mu}_n(k,\mathrm{d}\mathsf{z}) := \frac{1}{T+1}w_{n,k}(\mathsf{z})\mu_n\otimes M_n^{\otimes T}(\mathrm{d}\mathsf{z})$$

where for $M_n, L_{n-1,k}\colon \mathsf{Z}\times\mathscr{Z}\to[0,1]$, $k\in[\![0,P]\!]$ and any $\mathsf{z}=(z_0,z_1,\ldots,z_T)\in\mathsf{Z}^{T+1}$,

$$w_{n,k}(\mathsf{z}) := \frac{\mathrm{d}\mu_n\widehat{\otimes}L_{n-1,k}}{\mathrm{d}\mu_n\otimes M_n^k}(z_0,z_k)\,,$$

is assumed to exist for now and $w_{n,0}(\mathsf{z}) := 1$. This yields the marginals

$$\bar{\mu}_n(\mathrm{d}\mathsf{z}) := \sum_{k=0}^{T}\bar{\mu}_n(k,\mathrm{d}\mathsf{z}) = \frac{1}{T+1}\sum_{k=1}^{n}w_{n,k}(\mathsf{z})\,\mu_n\otimes M_n^{\otimes T}(\mathrm{d}\mathsf{z})\,. \tag{18}$$

Further, consider $R_n \colon \mathsf{Z} \times \mathscr{Z} \to [0,1]$ such that $\mu_{n-1} R_n = \mu_{n-1}$ and define $\bar{M}_n \colon \mathsf{Z}^{T+1} \times \mathscr{Z}^{\otimes(T+1)} \to [0,1]$

$$\bar{M}_n(\mathsf{z}, \mathrm{d}\mathsf{z}') := \sum_{k=0}^{T} \bar{\mu}_{n-1}(k \mid \mathsf{z}) R_n(z_k, \mathrm{d}z'_0) M_n^{\otimes T}\left(z'_0, \mathrm{d}\mathsf{z}'_{-0}\right).$$

Note that [14] set $R_n = M_n$, which we do not want in our later application and further assume that M_n is $\mu_{n-1}-$invariant, which is not necessary and too constraining for our application in Section 3.3 (corresponding to the application in the introductory Subsection 2.2). We only require the condition $\mu_{n-1} R_n = \mu_{n-1}$ is required. As in Subsection 2.2 we consider the optimal backward kernel $\bar{L}_n \colon \mathsf{Z}^{T+1} \times \mathscr{Z}^{\otimes(T+1)} \to [0,1]$, given for $(\mathsf{z}, A) \in \mathsf{Z}^{T+1} \times \mathscr{Z}^{\otimes(T+1)}$ by

$$\bar{L}_{n-1}(\mathsf{z}, A) = \frac{\mathrm{d}\bar{\mu}_{n-1} \otimes \bar{M}_n(A \times \cdot)}{\mathrm{d}\bar{\mu}_{n-1}\bar{M}_n}(\mathsf{z}),$$

and as established in Lemma 6 and Lemma 8, one obtains

$$\bar{w}_n(\mathsf{z}') = \frac{\mathrm{d}\bar{\mu}_n \hat{\otimes} \bar{L}_{n-1}}{\mathrm{d}\bar{\mu}_{n-1} \otimes \bar{M}_n}(\mathsf{z}, \mathsf{z}') = \frac{\mathrm{d}\mu_n}{\mathrm{d}\mu_{n-1}}(z'_0) \frac{1}{T+1} \sum_{k=0}^{T} w_{n,k}(\mathsf{z}'). \tag{19}$$

The corresponding folded version of the algorithm is given in Alg. 4, with $\check{\mathsf{z}}_n := (\check{z}_{n,0}, \check{z}_{n,1}, \ldots, \check{z}_{n,T})$...:

- $\left\{(\check{\mathsf{z}}_n^{(i)}, 1), i \in [\![N]\!]\right\}$ represents $\bar{\mu}_n(\mathrm{d}\mathsf{z})$,
- $\left\{(z_{n,k}^{(i)}, w_{n+1,k}), (i,k) \in [\![N]\!] \times [\![0,T]\!]\right\}$ represent $\mu_n(\mathrm{d}z)$,
- $\left\{\left(\mathsf{z}_n^{(i)}, \bar{w}_{n+1}\left(\mathsf{z}_n^{(i)}\right)\right), i \in [\![N]\!]\right\}$ represents $\bar{\mu}_{n+1}(\mathrm{d}\mathsf{z})$ and so do $\left\{(\check{\mathsf{z}}_{n+1}^{(i)}, 1), i \in [\![N]\!]\right\}$.

Remark 5. Other choices are possible and we detail here an alternative, related to the Waste-free framework of [14]. With notation as above here we take

$$\bar{\mu}_n(k, \mathrm{d}\mathsf{z}) := \frac{1}{T+1} w_{n,k}(\mathsf{z}) \mu_{n-1} \otimes M_n^{\otimes T}(\mathrm{d}\mathsf{z})$$

with the weights now defined as

$$w_{n,k}(\mathsf{z}) := \frac{\mathrm{d}\mu_n \hat{\otimes} L_{n-1,k}}{\mathrm{d}\mu_{n-1} \otimes M_n^k}(z_0, z_k).$$

With these choices we retain the fundamental property that for $f \colon \mathsf{Z} \to \mathbb{R}$, with $\bar{f}(k, \mathsf{z}) := f(z_k)$ then $\bar{\mu}_n\left(\bar{f}\right) = \mu_n(f)$. Now with

$$\bar{M}_n(\mathsf{z}, \mathrm{d}\mathsf{z}') := \sum_{k=0}^{T} \bar{\mu}_{n-1}(k \mid \mathsf{z}) R_n(z_k, \mathrm{d}z'_0) M_n^{\otimes T}\left(z'_0, \mathrm{d}\mathsf{z}'_{-0}\right),$$

assuming $\mu_{n-1} R_n = \mu_{n-1}$, we have the property that for any $A \in \mathscr{Z}^{\otimes(T+1)}$ $\bar{\mu}_{n-1}\bar{M}_n(A) = \mu_{n-1} \otimes M_n^{\otimes T}\left(A\right)$, yielding

$$\bar{w}_n(\mathsf{z}') = \frac{1}{T+1} \sum_{k=0}^{T} w_{n,k}(\mathsf{z}').$$

Finally choosing $L_{n,k}$ to be the optimized backward kernel, the importance weight of the algorithm is,

$$w_{n,k}(\mathsf{z}') = \frac{\mathrm{d}\mu_n \hat{\otimes} L_{n-1,k}}{\mathrm{d}\mu_{n-1} \otimes M_n^k}(z_0, z_k) = \frac{\mathrm{d}\mu_n}{\mathrm{d}\mu_{n-1}}(z_k).$$

sample $\check{\mathsf{z}}_0^{(i)} \overset{\text{iid}}{\sim} \mu_0^{\otimes(T+1)}$, set $w_{0,k}(\check{\mathsf{z}}_0^{(i)}) = 1$ for $(i,k) \in [\![N]\!] \times [\![T]\!]$.
for $n = 0, \dots, P-1$ **do**
  for $i \in [\![N]\!]$ **do**
    sample $a_i \sim \mathrm{Cat}\left(1, w_{n,1}(\check{\mathsf{z}}_n^{(i)}), w_{n,2}(\check{\mathsf{z}}_n^{(i)}), \dots, w_{n,T}(\check{\mathsf{z}}_n^{(i)})\right)$
    sample $z_{n,0}^{(i)} = z_n^{(i)} \sim R_{n+1}(\check{z}_{n,a_i}, \cdot)$
    for $k \in [\![T]\!]$ **do**
      sample $z_{n,k}^{(i)} \sim M_{n+1}(z_{n,k-1}^{(i)}, \cdot)$
      compute

$$w_{n+1,k}\left(\mathsf{z}_n^{(i)}\right) = \frac{\mathrm{d}\mu_n \hat{\otimes} L_n^k}{\mathrm{d}\mu_n \otimes M_{n+1}^k}(z_{n,0}^{(i)}, z_{n,k}^{(i)})$$

    end
    compute

$$\bar{w}_{n+1}\left(\mathsf{z}_n^{(i)}\right) = \frac{\mathrm{d}\mu_{n+1}}{\mathrm{d}\mu_n}\left(z_{n,0}^{(i)}\right) \frac{1}{T+1} \sum_{k \in [\![T]\!]} w_{n+1,k}\left(\mathsf{z}_n^{(i)}\right),$$

  end
  for $j \in \mathbb{N}$ **do**
    sample $b_j \sim \mathrm{Cat}\left(\bar{w}_{n+1}\left(\mathsf{z}_n^{(1)}\right), \bar{w}_{n+1}\left(\mathsf{z}_n^{(2)}\right), \dots, \bar{w}_{n+1}\left(\mathsf{z}_n^{(N)}\right)\right)$
    set $\check{\mathsf{z}}_{n+1}^{(j)} = \mathsf{z}_n^{(b_j)}$
  end
end

Algorithm 4: (Folded) Markov Snippet SMC algorithm

We note that continuous time Markov process snippets could also be used. For example piecewise deterministic Markov processes such as the Zig-Zag process [6] or the Bouncy Particle Sampler [8] could be used in practice since finite time horizon trajectories can be parametrized in terms of a finite number of parameters. We do not pursue this here.

3.2 Theoretical justification

In this section we provide the theoretical justification for the correctness of Alg. 4 and an alternative proof for Alg. 3, seen as a particular case of Alg. 4.

Throughout this section we use the following notation where μ (resp. ν) plays the rôle of μ_{n+1} (resp. μ_n), for notational simplicity. Let $\mu \in \mathcal{P}(\mathsf{Z}, \mathscr{Z})$ and $M, L \colon \mathsf{Z} \times \mathscr{Z} \to [0,1]$ be two Markov kernels such that the following condition holds. For $T \in \mathbb{N} \setminus \{0\}$ let $\mathsf{z} := (z_0, z_1, \ldots, z_T) \in \mathsf{Z}^{T+1}$ and for $k \in [\![0, T]\!]$ assume that $\mu \otimes M^k \gg \mu \widehat{\otimes}(L^k)$, implying the existence of the Radon-Nikodym derivatives

$$w_k(\mathsf{z}) := \frac{\mathrm{d}\mu \widehat{\otimes}(L^k)}{\mathrm{d}\mu \otimes M^k}(z_0, z_k)\,, \tag{20}$$

with the convention $w_0(\mathsf{z}) = 1$. We let $w(\mathsf{z}) := (T+1)^{-1} \sum_{k=0}^{T} w_k(\mathsf{z})$. For $\mathsf{z} \in \mathsf{Z}^{T+1}$, define

$$M^{\otimes T}(z_0, \mathrm{d}\mathsf{z}_{-0}) := \prod_{i=1}^{T} M(z_{i-1}, \mathrm{d}z_i)\,,$$

and introduce the mixture of distributions, defined on $\left(\mathsf{Z}^{T+1}, \mathscr{Z}^{\otimes(T+1)}\right)$,

$$\bar{\mu}(\mathrm{d}\mathsf{z}) = \sum_{k=0}^{T} \bar{\mu}(k, \mathrm{d}\mathsf{z})\,, \tag{21}$$

where for $(k, \mathsf{z}) \in [\![0, T]\!] \times \mathsf{Z}^{T+1}$

$$\bar{\mu}(k, \mathrm{d}\mathsf{z}) := \frac{1}{T+1} w_k(\mathsf{z}) \mu \otimes M^{\otimes T}(\mathrm{d}\mathsf{z})\,,$$

so that one can write $\bar{\mu}(\mathrm{d}\mathsf{z}) = w(\mathsf{z}) \mu \otimes M^{\otimes T}(\mathrm{d}\mathsf{z})$ with $w(\mathsf{z}) := (T+1)^{-1} \sum_{k=0}^{T} w_k(\mathsf{z})$.

We first establish properties of $\bar{\mu}$, showing how samples from $\bar{\mu}$ can be used to estimate expectations with respect to μ.

Lemma 6. *For any $f \colon \mathsf{Z} \to \mathbb{R}$ such that $\mu(|f|) < \infty$*

1. we have for $k \in [\![0, T]\!]$
$$\int f(z') \frac{\mathrm{d}\mu \widehat{\otimes}(L^k)}{\mathrm{d}\mu \otimes M^k}(z, z') \mu(\mathrm{d}z) M^k(z, \mathrm{d}z') = \mu(f)\,,$$
2. with $\bar{f}(k, \mathsf{z}) := f(z_k)$,
   (a) then
   $$\bar{\mu}(\bar{f}) = \mu(f)\,,$$

(b) in particular,

$$\int \sum_{k=0}^{T} \bar{f}(k, \mathsf{z}) \bar{\mu}(k \mid \mathsf{z}) \bar{\mu}(\mathrm{d}\mathsf{z}) = \mu(f) \,. \tag{22}$$

Proof. The first statement follows directly from the definition of the Radon-Nikodym derivative

$$\begin{aligned} \int f(z') \frac{\mathrm{d}\mu \hat{\otimes} L^k}{\mathrm{d}\mu \otimes M^k}(z, z') \mu \otimes M^k\big(\mathrm{d}(z, z')\big) &= \int f(z') \mu(\mathrm{d}z') L^k(z', \mathrm{d}z) \\ &= \int f(z') \mu(\mathrm{d}z') \,. \end{aligned}$$

The second statement follows from the definition of $\bar{\mu}$ and

$$\frac{1}{T+1} \sum_{k=0}^{T} \int \bar{f}(k, \mathsf{z}) w_k(\mathsf{z}) \mu \otimes M^{\otimes T}(\mathrm{d}\mathsf{z}) = \frac{1}{T+1} \sum_{k=0}^{T} \int f(z_k) \frac{\mathrm{d}\mu \hat{\otimes}(L^k)}{\mathrm{d}\mu \otimes M^k}(z_0, z_k) \mu \otimes M^k\big(\mathrm{d}(z_0, z_k)\big) \,,$$

and the first statement. The last statement follows from the tower property for expectations. □

Corollary 7. *Assume that* $\mathsf{z} \sim \bar{\mu}$*, then*

$$\sum_{k=0}^{T} f(z_k) \frac{w_k(\mathsf{z})}{\sum_{l=0}^{T} w_l(\mathsf{z})}$$

is an unbiased estimator of $\mu(f)$ *since we notice that for* $k \in [\![0, T]\!]$*,*

$$\bar{\mu}(k \mid \mathsf{z}) = \frac{w_k(\mathsf{z})}{\sum_{l=0}^{T} w_l(\mathsf{z})} \,.$$

This justifies algorithms which sample from the mixture $\bar{\mu}$ *directly in order to estimate expectations with respect to* μ*.*

Let $\nu \in \mathcal{P}\big(\mathsf{Z}, \mathscr{Z}\big)$ and $\bar{\nu} \in \mathcal{P}\big(\mathsf{Z}^{T+1}, \mathscr{Z}^{\otimes(T+1)}\big)$ be derived from ν in the same way $\bar{\mu}$ is from μ in (21), but for possibly different Markov kernels $M_\nu, L_\nu \colon \mathsf{Z} \times \mathscr{Z} \to [0, 1]$. Define now the kernel $\bar{M} \colon \mathsf{Z}^{T+1} \times \mathscr{Z}^{\otimes(T+1)} \to [0, 1]$ be the Markov kernel

$$\bar{M}(\mathsf{z}, \mathrm{d}\mathsf{z}') := \sum_{k=0}^{T} \bar{\nu}(k \mid \mathsf{z}) R(z_k, \mathrm{d}z'_0) M^{\otimes T}\big(z'_0, \mathrm{d}\mathsf{z}'_{-0}\big) \,, \tag{23}$$

with $R \colon \mathsf{Z} \times \mathscr{Z} \to [0, 1]$. Remark that M, L, M_ν, L_ν and R can be made dependent on both μ and ν, provided they satisfy all the conditions stated above and below.

The following justifies the existence of $\bar{w}_{n+1}$ in (19) for a particular choice of backward kernel and provides a simplified expression for a particular choices of R.

Lemma 8. *With the notation (20)-(21) and (23),*

1. *there exists* $\bar{M}^* \colon \mathsf{Z}^{T+1} \times \mathscr{Z}^{\otimes(T+1)} \to [0, 1]$ *such that for any* $A, B \in \mathscr{Z}^{\otimes(T+1)}$

$$\bar{\nu} \otimes \bar{M}(A \times B) = (\bar{\nu}\bar{M}) \otimes \bar{M}^*(B \times A)$$

2. *we have*

$$\bar{\nu}\bar{M} = (\nu R) \otimes M^{\otimes T} \,,$$

3. assuming $\nu R \gg \mu$ *then with the choice* $\bar{L} = \bar{M}^*$ *we have* $\bar{\nu} \otimes \bar{M} \gg \bar{\mu} \widehat{\otimes} \bar{L}$ *and for* $\mathsf{z}, \mathsf{z}' \in \mathsf{Z}^{T+1}$ *almost surely*

$$\bar{w}(\mathsf{z}, \mathsf{z}') = \frac{\mathrm{d}\bar{\mu} \widehat{\otimes} \bar{L}}{\mathrm{d}\bar{\nu} \otimes \bar{M}}(\mathsf{z}, \mathsf{z}') = \frac{\mathrm{d}\mu}{\mathrm{d}\nu R}(z_0') \frac{1}{T+1} \sum_{k=0}^{T} w_k(\mathsf{z}') = \frac{\mathrm{d}\mu}{\mathrm{d}\nu R}(z_0') w(\mathsf{z}') =: \bar{w}(\mathsf{z}') \,,$$

that is we introduce notation reflecting this last simplication.

4. Notice the additional simplification when $\nu R = \nu$.

Proof. The first statement is a standard result and $\bar{M}^*$ is a conditional expectation. We however provide the short argument taking advantage of the specific scenario. For a fixed $A \in \mathscr{Z}^{\otimes(T+1)}$ consider the finite measure

$$B \mapsto \bar{\nu} \otimes \bar{M}(A \times B) \leq \bar{\nu}\bar{M}(B) \,,$$

such that $\bar{\nu}\bar{M}(B) = 0 \implies \bar{\nu} \otimes \bar{M}(A \times B) = 0$, that is $\bar{\nu}\bar{M} \gg \bar{\nu} \otimes \bar{M}(A \times \cdot)$. Consequently we have the existence of a Radon-Nikodym derivative such that

$$\bar{\nu} \otimes \bar{M}(A \times B) = \int_B \frac{\mathrm{d}\bar{\nu} \otimes \bar{M}(A \times \cdot)}{\mathrm{d}(\bar{\nu}\bar{M})}(\mathsf{z}) \bar{\nu}\bar{M}(\mathrm{d}\mathsf{z})$$

which indeed has the sought property. This is a kernel since for any fixed $A \in \mathscr{Z}$, for any $B \in \mathscr{Z}$, $\bar{\nu} \otimes \bar{M}(A \times B) \leq 1$ and therefore $\bar{M}^*(\mathsf{z}, A) \leq 1$ $\bar{\nu}\bar{M}$-a.s., with equality for $A = \mathsf{X}$. For the second statement we have, for any $(\mathsf{z}, A) \in \mathsf{Z}^{T+1} \times \mathscr{Z}^{\otimes(T+1)}$,

$$\bar{M}(\mathsf{z}, A) = \sum_{k=0}^{T} \bar{\nu}(k \mid \mathsf{z}) R \otimes M^{\otimes T}\big(z_k, A\big),$$

and we can apply (22) in Lemma 6 with the substitutions $\mu \leftarrow \nu$ and $M \leftarrow M_\nu, L \leftarrow L_\nu$ to conclude. For the third statement, since $\nu R \gg \mu$ then $\bar{\nu}\bar{M} = (\nu R) \otimes M^{\otimes T} \gg \mu \otimes M^{\otimes T}$ because for any $A \in \mathscr{Z}^{\otimes(T+1)}$,

$$\int \mathbf{1}\{\mathsf{z} \in A\} \nu R(\mathrm{d}z_0) M^{\otimes T}(z_0, \mathrm{d}\mathsf{z}_{-0}) = 0 \implies \begin{cases} \int \mathbf{1}\{\mathsf{z} \in A\} M^{\otimes T}(z_0, \mathrm{d}\mathsf{z}_{-0}) = 0 \quad \nu R - a.s. \\ \text{or} \\ \nu R\big(\mathfrak{P}(A)\big) = 0 \end{cases}$$

where $\mathfrak{P} \colon \mathsf{Z}^{T+1} \to \mathsf{Z}$ is such that $\mathfrak{P}(\mathsf{z}) = z_0$. In either case, since $\nu R \gg \mu$, this implies that

$$\int \mathbf{1}\{\mathsf{z} \in A\} \mu(\mathrm{d}z_0) M^{\otimes T}(z_0, \mathrm{d}\mathsf{z}_{-0}) = 0 \,,$$

that is $\bar{\nu}\bar{M} \gg \bar{\mu}$ from the definition of $\bar{\mu}$. Consequently

$$\begin{aligned} \bar{\mu} \otimes \bar{M}^*(A \times B) &= \int_A \bar{M}^*(\mathsf{z}, B) \frac{\mathrm{d}\bar{\mu}}{\mathrm{d}(\bar{\nu}\bar{M})}(\mathsf{z}) \bar{\nu}\bar{M}(\mathrm{d}\mathsf{z}) \\ &= \int_A \frac{\mathrm{d}\bar{\mu}}{\mathrm{d}(\bar{\nu}\bar{M})}(\mathsf{z}) \bar{M}^*(\mathsf{z}, B) \bar{\nu}\bar{M}(\mathrm{d}\mathsf{z}) \\ &= \int \mathbf{1}\{\mathsf{z} \in A, \mathsf{z}' \in B\} \frac{\mathrm{d}\bar{\mu}}{\mathrm{d}(\bar{\nu}\bar{M})}(\mathsf{z}) \bar{\nu}(\mathrm{d}\mathsf{z}') \bar{M}(\mathsf{z}', \mathrm{d}\mathsf{z}) \end{aligned}$$

and indeed from Fubini's and Dynkin's $\pi - \lambda$ theorems $\bar{\nu} \otimes \bar{M} \gg \bar{\mu} \widehat{\otimes} \bar{L}$ and

$$\frac{\mathrm{d}\bar{\mu}\widehat{\otimes}\bar{L}}{\mathrm{d}\bar{\nu} \otimes \bar{M}}(\mathsf{z}, \mathsf{z}') = \frac{\mathrm{d}\bar{\mu}}{\mathrm{d}(\bar{\nu}\bar{M})}(\mathsf{z}') \,.$$

Now for $f \colon \mathsf{Z}^{T+1} \to [0,1]$ and using the second statement and $\nu R \gg \mu$,

$$\begin{aligned}
\int f(\mathsf{z}) \frac{\mathrm{d}\bar{\mu}}{\mathrm{d}(\bar{\nu}\bar{M})}(\mathsf{z})\, (\bar{\nu}\bar{M})(\mathrm{d}\mathsf{z}) &= \int f(\mathsf{z}) w(\mathsf{z}) \mu \otimes M^{\otimes T}(\mathrm{d}\mathsf{z}) \\
&= \int f(\mathsf{z}) w(\mathsf{z}) \frac{\mathrm{d}\mu}{\mathrm{d}(\nu R)}(z_0)(\nu R)(\mathrm{d}z_0) M^{\otimes T}(z_0, \mathrm{d}\mathsf{z}_{-0}) \\
&= \int f(\mathsf{z}) w(\mathsf{z}) \frac{\mathrm{d}\mu}{\mathrm{d}(\nu R)}(z_0) \bar{\nu}\bar{M}(\mathrm{d}\mathsf{z}) \,.
\end{aligned}$$

We therefore conclude that

$$\frac{\mathrm{d}\bar{\mu}\widehat{\otimes}\bar{L}}{\mathrm{d}\bar{\nu} \otimes \bar{M}}(\mathsf{z}, \mathsf{z}') = w(\mathsf{z}') \frac{\mathrm{d}\mu}{\mathrm{d}(\nu R)}(z_0') = w(\mathsf{z}') \frac{\mathrm{d}\mu}{\mathrm{d}\nu}(z_0')$$

where the last inequality holds only when $\nu R = \nu$. □

The following result is important in two respects. First it establishes that if M, L satisfy a simple property then w_k always have a simple expresssion in terms of certain densities of μ and ν – this implies in particular that in Subsection 3.1 the kernel M_n is not required to leave μ_{n-1} invariant to make the method implementable [14]. Second it provides a direct justification of the validity of advanced schemes – see Example 14. This therefore establishes that generic and widely applicable sufficient conditions for $\bar{w}(\mathsf{z}, \mathsf{z}')$ to be tractable are $\nu R = \nu$ and the notion of (υ, M, M^*)-reversibility.

Lemma 9. *Let $\mu, \nu \in \mathcal{P}(\mathsf{Z}, \mathscr{Z})$, υ be a σ-finite measure on $(\mathsf{Z}, \mathscr{Z})$ such that $\upsilon \gg \mu, \nu$ and assume that we have a pair of Markov kernels $M, M^* \colon \mathsf{Z} \times \mathscr{Z} \to [0,1]$ such that*

$$\upsilon(\mathrm{d}z) M(z, \mathrm{d}z') = \upsilon(\mathrm{d}z') M^*(z', \mathrm{d}z) \,. \tag{24}$$

We call this property (υ, M, M^*)-reversibility. Then for $z, z' \in \mathsf{Z}$ such that $\nu(z) := \mathrm{d}\nu/\mathrm{d}\upsilon(z) > 0$ we have

$$\frac{\mathrm{d}\mu\widehat{\otimes}M^*}{\mathrm{d}\nu \otimes M}(z, z') = \frac{\mathrm{d}\mu/\mathrm{d}\upsilon(z')}{\mathrm{d}\nu/\mathrm{d}\upsilon(z)} \,.$$

Proof. For $z, z' \in \mathsf{Z}$ such that $\mathrm{d}\nu/\mathrm{d}\upsilon(z) > 0$ we have

$$\begin{aligned}
\mu(\mathrm{d}z') M^*(z', \mathrm{d}z) &= \frac{\mathrm{d}\mu}{\mathrm{d}\upsilon}(z') \upsilon(\mathrm{d}z') M^*(z', \mathrm{d}z) \\
&= \frac{\mathrm{d}\mu}{\mathrm{d}\upsilon}(z') \upsilon(\mathrm{d}z) M(z, \mathrm{d}z') \\
&= \frac{\mathrm{d}\mu/\mathrm{d}\upsilon(z')}{\mathrm{d}\nu/\mathrm{d}\upsilon(z)} \frac{\mathrm{d}\nu}{\mathrm{d}\upsilon}(z) \upsilon(\mathrm{d}z) M(z, \mathrm{d}z') \\
&= \frac{\mathrm{d}\mu/\mathrm{d}\upsilon(z')}{\mathrm{d}\nu/\mathrm{d}\upsilon(z)} \nu(\mathrm{d}z) M(z, \mathrm{d}z') \,.
\end{aligned}$$

□

Corollary 10. *Let $\mu_{n-1}, \mu_n \in \mathcal{P}(\mathsf{Z}, \mathscr{Z})$ and υ be a σ-finite measure such that $\upsilon \gg \mu_{n-1}, \mu_n$, let $M_n, L_{n-1} \colon (\mathsf{Z}, \mathscr{Z}) \to [0,1]$ such that*

$$\upsilon(\mathrm{d}z) M_n(z, \mathrm{d}z') = \upsilon(\mathrm{d}z') L_{n-1}(z', \mathrm{d}z)\,,$$

then for any $z, z' \in \mathsf{Z}$ such that $\mathrm{d}\mu_n/\mathrm{d}\upsilon(z) > 0$ and $k \in \mathbb{N}$

$$\frac{\mu_n \widehat{\otimes} L_{n-1}^k}{\mu_n \otimes M_n^k}(z, z') = \frac{\mathrm{d}\mu_n/\mathrm{d}\upsilon(z')}{\mathrm{d}\mu_n/\mathrm{d}\upsilon(z)}\,.$$

and provided $\mu_{n-1} R_n = \mu_{n-1}$ we can deduce

$$\begin{aligned}\bar{w}_n(\mathsf{z}) &= \frac{\mathrm{d}\mu_n/\mathrm{d}\upsilon(z_0)}{\mathrm{d}\mu_{n-1}/\mathrm{d}\upsilon(z_0)} \frac{1}{T+1} \sum_{k=0}^{T} \frac{\mathrm{d}\mu_n/\mathrm{d}\upsilon(z_k)}{\mathrm{d}\mu_n/\mathrm{d}\upsilon(z_0)} \\ &= \frac{1}{T+1} \sum_{k=0}^{T} \frac{\mathrm{d}\mu_n/\mathrm{d}\upsilon(z_k)}{\mathrm{d}\mu_{n-1}/\mathrm{d}\upsilon(z_0)}\end{aligned}$$

where we have used Lemma 34.

We have shown earlier that standard integrator based mutation kernels used in the context of Monte Carlo method satisfy (24) with υ the Lebesgue measure but other scenarios involved that of the preconditioned Crank–Nicolson (pCN) algorithm where υ is the distribution of a Gaussian process.

3.3 Revisiting sampling with integrator snippets

In this scenario we have $\mu(\mathrm{d}z) = \pi(\mathrm{d}x)\varpi(\mathrm{d}v)$ assumed to have a density w.r.t. a σ-finite measure υ, and ψ is an invertible mapping $\psi \colon \mathsf{Z} \to \mathsf{Z}$ such that $\upsilon^{\psi} = \upsilon$ and $\psi^{-1} = \sigma \circ \psi \circ \sigma$ with $\sigma \colon \mathsf{Z} \to \mathsf{Z}$ such that $\mu \circ \sigma(z) = \mu(z)$. In his manuscript we focus primarily on the scenario where ψ is a discretization of Hamilton's equations for a potential $U \colon \mathsf{X} \to \mathbb{R}$ e.g. a leapfrog integrator. We consider now the scenario where, in the framework developed in Section 3.1, we let $M(z, \mathrm{d}z') := \delta_{\psi(z)}(\mathrm{d}z')$ be the deterministic kernel which maps the current state $z \in \mathsf{Z}$ to $\psi(z)$. Define $\Psi(z, \mathrm{d}z') := \delta_{\psi(z)}(\mathrm{d}z')$ and $\Psi^*(z, \mathrm{d}z') := \delta_{\psi^{-1}(z)}(\mathrm{d}z')$; we exploit the ideas of [1, Proposition 4] to establish that Ψ^* is the $\upsilon-$adjoint of Ψ if υ is invariant under ψ.

Lemma 11. *Let μ be a probability measure and υ a σ-finite measure, on $(\mathsf{Z}, \mathscr{Z})$ such that $\upsilon \gg \mu$. Denote $\mu(z) := \mathrm{d}\mu/\mathrm{d}\upsilon(z)$ for any $z \in \mathsf{Z}$. Let $\psi \colon \mathsf{Z} \to \mathsf{Z}$ be an invertible and volume preserving mapping, i.e. such that $\upsilon^{\psi}(A) = \upsilon\big(\psi^{-1}(A)\big) = \upsilon(A)$ for all $A \in \mathscr{Z}$, then*

1. *(υ, Ψ, Ψ^*) form a reversible triplet, that is for all $z, z' \in \mathsf{Z}$,*

$$\upsilon(\mathrm{d}z)\delta_{\psi(z)}(\mathrm{d}z') = \upsilon(\mathrm{d}z')\delta_{\psi^{-1}(z')}(\mathrm{d}z),$$

2. *for all $z, z' \in \mathsf{Z}$ such that $\mu(z) > 0$*

$$\mu(\mathrm{d}z')\delta_{\psi^{-1}(z')}(\mathrm{d}z) = \frac{\mu \circ \psi(z)}{\mu(z)} \mu(\mathrm{d}z)\delta_{\psi(z)}(\mathrm{d}z')\,.$$

Proof. For the first statement

$$\begin{aligned}\int f(z)g\circ\psi(z)\upsilon(\mathrm{d}z) &= \int f\circ\psi^{-1}\circ\psi(z)g\circ\psi(z)\upsilon(\mathrm{d}z)\\ &= \int f\circ\psi^{-1}(z)g(z)\upsilon^{\psi}(\mathrm{d}z)\\ &= \int f\circ\psi^{-1}(z)g(z)\upsilon(\mathrm{d}z).\end{aligned}$$

We have

$$\begin{aligned}\mu(\mathrm{d}z')\delta_{\psi^{-1}(z')}(\mathrm{d}z) &= \mu(z')\upsilon(\mathrm{d}z')\delta_{\psi^{-1}(z')}(\mathrm{d}z)\\ &= \mu(z')\upsilon(\mathrm{d}z)\delta_{\psi(z)}(\mathrm{d}z')\\ &= \mu\circ\psi(z)\upsilon(\mathrm{d}z)\delta_{\psi(z)}(\mathrm{d}z')\\ &= \frac{\mu\circ\psi(z)}{\mu(z)}\mu(z)\upsilon(\mathrm{d}z)\delta_{\psi(z)}(\mathrm{d}z')\\ &= \frac{\mu\circ\psi(z)}{\mu(z)}\mu(\mathrm{d}z)\delta_{\psi(z)}(\mathrm{d}z')\end{aligned}$$

$$\begin{aligned}\int f(z')g(z)\mu(\mathrm{d}z')\delta_{\psi^{-1}(z')}(\mathrm{d}z) &= \int f(z')g(z)\mu(z')\upsilon(\mathrm{d}z')\delta_{\psi^{-1}(z')}(\mathrm{d}z)\\ &= \int f(z')g(z)\mu(z')\upsilon(\mathrm{d}z)\delta_{\psi(z)}(\mathrm{d}z')\\ &= \int f\circ\psi(z)g(z)\mu\circ\psi(z)\upsilon(\mathrm{d}z)\\ &= \int f\circ\psi(z)g(z)\frac{\mu\circ\psi(z)}{\mu(z)}\mu(z)\upsilon(\mathrm{d}z)\\ &= \int f(z')g(z)\frac{\mu\circ\psi(z)}{\mu(z)}\mu(\mathrm{d}z)\delta_{\psi(z)}(\mathrm{d}z')\end{aligned}$$

As a result

$$\begin{aligned}\int f(z')\mu(\mathrm{d}z') &= \int f(z')\mu(\mathrm{d}z')\delta_{\psi^{-1}(z')}(\mathrm{d}z)\\ &= \int f(z')\frac{\mu\circ\psi(z)}{\mu(z)}\mu(\mathrm{d}z)\delta_{\psi(z)}(\mathrm{d}z')\\ &= \int f\circ\psi(z)\frac{\mu\circ\psi(z)}{\mu(z)}\mu(\mathrm{d}z)\end{aligned}$$

□

Corollary 12. *With the assumptions of Lemma 11 above for $k\in[\![0,T]\!]$ the weight (20) for $M_\mu=\Psi^k, L_\mu=\Psi^{-k}$ admits the expression*

$$w_k(\mathsf{z}) = \frac{\mathrm{d}\mu\hat{\otimes}\Psi^{-k}}{\mathrm{d}\mu\otimes\Psi^k}(z_0,z_k) = \frac{\mu\circ\psi^k(z_0)}{\mu(z_0)},$$

Further, for R a ν-invariant Markov kernel the expression for the weight (19) becomes

$$\bar{w}(\mathsf{z}, \mathsf{z}') = \frac{\mu(z_0')}{\nu(z_0')} \frac{1}{T+1} \sum_{k=0}^{T} \frac{\mu \circ \psi^k(z_0')}{\mu(z_0')},$$

hence recovering the expression used in Section 2. This together with the results of Section 3.1 provides an alternative justification of correctness of Alg. 3 and hence Alg. 2. Note that this choice for L_μ corresponds to the so-called optimal scenario; this can be deduced from Lemma 11 or by noting that

$$\begin{aligned}\int f(z, z')\mu\Psi(\mathrm{d}z')\Psi^{-1}(z', \mathrm{d}z) &= \int f(\psi^{-1}(z'), z')\mu\Psi(\mathrm{d}z') \\ &= \int f(z, \psi(z))\mu(\mathrm{d}z) \\ &\int f(z, z')\mu(\mathrm{d}z)\Psi(z, \mathrm{d}z')\,.\end{aligned}$$

3.4 More complex integrator snippets

Here we provide examples of more complex integrator snippets to which earlier theory immediately applies thanks to the abstract point of view adopted throughout.

Example 13. Let $\upsilon \gg \mu$, so that $\mu(z) := \mathrm{d}\mu/\mathrm{d}\upsilon(z)$ is well defined. Let, for $i \in \{1, 2\}$, $\psi_i \colon \mathsf{Z} \to \mathsf{Z}$ be invertible and such that $\upsilon^{\psi_i} = \upsilon$, $\psi_i^{-1} = \sigma \circ \psi_i \circ \sigma$ for $\sigma \colon \mathsf{Z} \to \mathsf{Z}$ such that $\sigma^2 = \mathrm{Id}$ and $\upsilon^\sigma = \upsilon$. Then one can consider the delayed rejection MH transition probability

$$M(z, \mathrm{d}z') = \alpha_1(z)\delta_{\psi_1(z)}(\mathrm{d}z') + \bar{\alpha}_1(z)\left[\alpha_2(z)\delta_{\psi_2(z)}(\mathrm{d}z') + \bar{\alpha}_2(z)\delta_z(\mathrm{d}z')\right],$$

with $\alpha_i(z) = 1 \wedge r_i(z)$ $\bar{\alpha}_i(z) = 1 - \alpha_i(z)$ and $z \in S_1 \cap S_2$ given below

$$r_1(z) = \frac{\mathrm{d}\upsilon^{\psi_1^{-1}}}{\mathrm{d}\upsilon}(z), \quad r_2(z) = \frac{\bar{\alpha}_1 \circ \sigma \circ \psi_2(z)}{\bar{\alpha}_1(z)} \frac{\mathrm{d}\upsilon^{\psi_2^{-1}}}{\mathrm{d}\upsilon}(z)\,.$$

A particular example is when υ is the Lebesgue measure on $\mathsf{X} \times \mathsf{V}$ and ψ_1 is volume preserving, for instance the leapfrog integrator for Hamilton's equations for some potential U or the bounce update (25). Following [1] notice that υ has density $\upsilon(z) = 1/2$ with respect to the measure $\upsilon + \upsilon^{\psi_i} = 2\upsilon$. Now define $S_1 := \left\{z \in \mathsf{Z} \colon \upsilon(z) \wedge \upsilon \circ \psi_1(z) > 0\right\}$, we have

$$r_1(z) := \begin{cases} \frac{\upsilon \circ \psi_1(z)}{\upsilon(z)} = 1 & \text{for } z \in S_1 \\ 0 & \text{otherwise} \end{cases},$$

and with $S_2 := \left\{z \in \mathsf{Z} \colon \left[\bar{\alpha}_1(z)\,\upsilon(z)\right] \wedge \left[\bar{\alpha}_1 \circ \sigma \circ \psi_2(z)\,\upsilon \circ \psi_2(z)\right] > 0\right\}$

$$r_2(z) := \begin{cases} \frac{\bar{\alpha}_1 \circ \sigma \circ \psi_2(z)\,\upsilon \circ \psi_2(z)}{\bar{\alpha}_1(z)\,\upsilon(z)} = 1 & \text{for } z \in S_2 \\ 0 & \text{otherwise} \end{cases}.$$

For instance ψ_1 can be a HMC update, while

$$\psi_2 = \psi_1 \circ \mathrm{b} \circ \psi_1 \text{ with } \mathrm{b}(x, v) = \left(x, v - 2\langle v, n(x)\rangle n(x)\right) \tag{25}$$

for some unit length vector field $n \colon \mathsf{X} \to \mathsf{X}$. This can therefore be used as part of Alg. 4; care must be taken when compute the weights $w_{n,k}$ and $\bar{w}_n$, see Section 3-3.3 provide the tools to achieve this. For example, in the situation where υ is the Lebesgue measure on $\mathsf{Z} = \mathbb{R}^d$ and $\psi_2 = \mathrm{Id}$ then we recover Alg. 3 or equivalently Alg. 2.

Example 14. An interesting instance of Example 13 is concerned with the scenario where interest is in sampling μ constrained to some set $C \subset \mathsf{Z}$ such that $\upsilon(C) < \infty$. Define μ constrainted to C, $\mu_C(\cdot) := \mu(C \cap \cdot)/\mu(C)$ and similarly $\upsilon_C(\cdot) := \upsilon(C \cap \cdot)/\upsilon(C)$. We let M be defined as above but targeting υ_C. Naturally υ_C has a density w.r.t. υ, $\upsilon_C(z) = \mathbf{1}_C(z)/\upsilon(C)$ for $z \in \mathsf{Z}$ and for $i \in \{1, 2\}$ we have $\upsilon_C \circ \psi_i(z) = \mathbf{1}_{\psi_i^{-1}(C)}(z)$, $\upsilon \circ \psi_1 \circ \psi_2(z) = \mathbf{1}_{\psi_2^{-1} \circ \psi_1^{-1}(C)}(z)$. Consequently $S_1 := \left\{z \in \mathsf{Z} \colon \mathbf{1}_{C \cap \psi_1^{-1}(C)}(z) > 0\right\}$ and $S_2 = \left\{z \in \mathsf{Z} \colon \mathbf{1}_{C \cap \psi_1^{-1}(C^{\complement})}(z)\mathbf{1}_{\psi_2^{-1}(C) \cap \psi_2^{-1} \circ \psi_1^{-1}(C^{\complement})}(z) > 0\right\}$ and as a result

$$\begin{aligned}\alpha_1(z) &:= \mathbf{1}_{A \cap \psi_1^{-1}(C)}(z) \\ \alpha_2(z) &:= \mathbf{1}_{A \cap \psi_1^{-1}(C^{\complement}) \cap \psi_2^{-1}(C) \cap \psi_2^{-1} \circ \psi_1^{-1}(C^{\complement})}(z)\,.\end{aligned}$$

The corresponding kernel $M^{\otimes T}$ is described algorithmically in Alg. 5. In the situation where $C := \{x \in \mathsf{X} \colon c(x) = 0\}$ for a continuously differentiable function $c \colon \mathsf{X} \to \mathbb{R}$, the bounces described in (25) can be defined in terms of the field $x \mapsto n(x)$such that

$$n(x) := \begin{cases} \nabla c(x)/|\nabla c(x)| & \text{for } \nabla c(x) \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

This justifies the ideas of [5], where a process of the type given in Alg. 5 is used as a proposal within a MH update, although the possibility of a rejection after the second stage seems to have been overlooked in that reference.

Given $z_0 = z \in C \subset \mathsf{Z}$
for $k = 1, \ldots, T$ **do**
  if $\psi_1(z_{k-1}) \in C$ **then**
    $z_k = \psi_1(z_{k-1})$
  else if $\psi_2(z_{k-1}) \in C$ *and* $\psi_1 \circ \psi_2(z_{k-1}) \notin C$ **then**
    $z_k = \psi_2(z_{k-1})$
  else
    $z_k = z_{k-1}$.
  end
end

Algorithm 5: $M^{\otimes T}$ for M the delayed rejection algorithm targetting the uniform distribution on C

Naturally a rejection of both transformations ψ_1 and ψ_2 of the current state means that the algorithm gets stuck. We note that it is also possible to replace the third update case with a full refreshment of the velocity, which can be interpreted as a third delayed rejection update, of acceptance probability one.

3.5 Numerical illustration: orthant probabilities

In this section, we consider the problem of calculating the Gaussian orthant probabilities, which is given by

$$p(\mathbf{a}, \mathbf{b}, \Sigma) := \mathbb{P}(\mathbf{a} \leq \mathbf{X} \leq \mathbf{b}),$$

where $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ are known vectors of dimension d and $\mathbf{X} \sim \mathcal{N}_d(\mathbf{0}, \Sigma)$ with Σ a covariance matrix of size $d \times d$. Consider the Cholesky decomposition of Σ which is given by $\Sigma := LL^{\top}$, where $L := (l_{ij}, i, j \in [\![d]\!])$ is a lower

triangular matrix with positive diagonal entries. It is clear that $\mathbf{X}$ can be viewed as $\mathbf{X} := L\eta$, where $\eta \sim \mathcal{N}_d(\mathbf{0}, \mathrm{Id}_d)$. Consequently, one can instead rewrite $p(\mathbf{a}, \mathbf{b}, \Sigma)$ as a product of d probabilities given by

$$p_1 = \mathbb{P}(a_1 \leq l_{11}\eta_1 \leq b_1) = \mathbb{P}\left(a_1/l_{11} \leq \eta_1 \leq b_1/l_{11}\right), \tag{26}$$

and

$$p_n = \mathbb{P}\left(a_t \leq \sum_{j=1}^{n} l_{nj}\eta_j \leq b_t\right) = \mathbb{P}\left(\left(a_t - \sum_{j=1}^{n-1} l_{nj}\eta_j\right)/l_{nn} \leq \eta_t \leq \left(b_t - \sum_{j=1}^{n-1} l_{nj}\eta_j\right)/l_{nn}\right), \tag{27}$$

for $n = 2, \ldots, d$. For notational simplicity, we let $\mathcal{B}_n(\eta_{1:n-1})$ denote the interval $\left[\left(a_t - \sum_{j=1}^{n-1} l_{nj}\eta_j\right)/l_{nn}, \left(b_t - \sum_{j=1}^{n-1} l_{nj}\eta_j\right)/l_{nn}\right]$, with the convention $\mathcal{B}_1(\eta_{1:0}) := [a_1/l_{11}, b_1/l_{11}]$. Then, $p(\mathbf{a}, \mathbf{b}, \Sigma)$ can be written as the product of p_n s for $n = 1, 2, \ldots, d$. Moreover, one could see that p_n is also the normalising constant of the conditional distribution of η_n given $\eta_{1:n-1}$. To calculate the orthant probability, [32] have proposed an SMC algorithm targetting the sequence of distributions $\pi_n(\eta_{1:n}) := \pi_1(\eta_n)\prod_{k=2}^{n}\pi_n(\eta_k|\eta_{1:k-1})$ for $n \in [\![1, d]\!]$, given by

$$\pi_1(\eta_1) \propto \phi(\eta_1)\mathbf{1}\{\eta_1 \in \mathcal{B}_1\} = \gamma_1(\eta_1) \tag{28}$$

$$\pi_n(\eta_n|\eta_{1:n-1}) \propto \phi(\eta_n)\mathbf{1}\{\eta_n \in \mathcal{B}_n(\eta_{1:n-1})\} = \gamma_n(\eta_n|\eta_{1:n-1}). \tag{29}$$

where ϕ denotes the probability density of a $\mathcal{N}(0, 1)$. One could also note that

$$\pi_n(\eta_n|\eta_{1:n-1}) = 1/\Phi(\mathcal{B}_n(\eta_{1:n-1}))\phi(\eta_n)\mathbf{1}\{\eta_n \in \mathcal{B}_n(\eta_{1:n-1})\}$$

and $\gamma_n(\eta_{1:n}) = \phi(\eta_{1:n})\prod_{k=1}^{n}\mathbf{1}\{\eta_k \in \mathcal{B}_k(\eta_{1:k-1})\}$, where $\Phi(\mathcal{B}_n(\eta_{1:n-1}))$ represents the probability of a standard Normal random variable being in the region $\mathcal{B}_n(\eta_{1:n-1})$. Therefore, the SMC algorithm proposed by [32] then proceeds as follows. (1) At time t, particles $\eta_{1:t-1}^n$ are extended by sampling $\eta_n^n \sim \pi_n(\mathrm{d}\eta_t|\eta_{1:t-1}^{(i)})$. (2) Particles $\eta_{1:t}^{(i)}$ are then reweighted by multiplying the incremental weights $\Phi(\mathcal{B}_n(\eta_{1:n-1}^{(i)}))$ to $w_{n-1}^{(i)}$. (3) If the ESS is below a certain threshold, resample the particles and move them through an MCMC kernel that leaves π_t invariant for k iterations. For the MCMC kernel, [32] recommended using Gibbs sampler that leaves π_t invariant to move the particles at step (3). The orthant probability we are interested in can then be viewed as the normalising constant of π_d and this can be estimated as a by-product of the SMC algorithm.

Since we are trying to sample from the constrained Gaussian distributions, the Hamiltonian equation can be solved exactly and $w_{n,k}$ is always 1. As a result, the incremental weights for the trajectories simplify to $\Phi(\mathcal{B}_n(u_{n-1}))$ and each particle on the trajectory starting from z_t will have an incremental weight proportional to $\Phi(\mathcal{B}_n(u_{n-1}))$. To obtain a trajectory, we follow [30] who perform HMC with reflections to sample. As the dimension increases, the number of reflections performed under ψ_n will also increase given a fixed integration time. We adaptively tuned the integrating time ε to ensure that the everage number of reflections at each SMC step does not exceed a given threshold. In our experiment we set this threshold to be 5. To show that the waste-recycling RSMC algorithm scales well in high dimension, we set $d = 150$, $a = (1.5, 1.5, ...)$ and $b = (\infty, \infty, ...)$. Also, we use the same covariance matrix in [14] and perform variable re-ordering as suggested in [32] before the simulation.

Figures 9 and 10 show the results obtained with $N \times (T + 1) = 50,000$ and various values for N. With a quarter of the number of particles used in [14], the waste-recycling HSMC algorithm achieves comparable performance when estimating the normalising constant (i.e. the orthant probability). Moreover, the estimates are stable for different choices of N values, although one observes that the algorithm achieves best performance when $N = 500$ (i.e. each trajectory contains 100 particles). This also suggests that the integrating time should be adaptively tuned in a different way to achieve the best performance given a fixed computational budget. Estimates of the function $\varphi(x_{0:d}) = \mathbb{E}(1/d\sum_{i=1}^{d} x_i)$ with respect to the Gaussian distribution $\mathcal{N}_d(\mathbf{0}, \Sigma)$ truncated between $\mathbf{a}$ and $\mathbf{b}$ are also stable for different choices of N, although they are more variable than those obtained in [14]. This indicates that

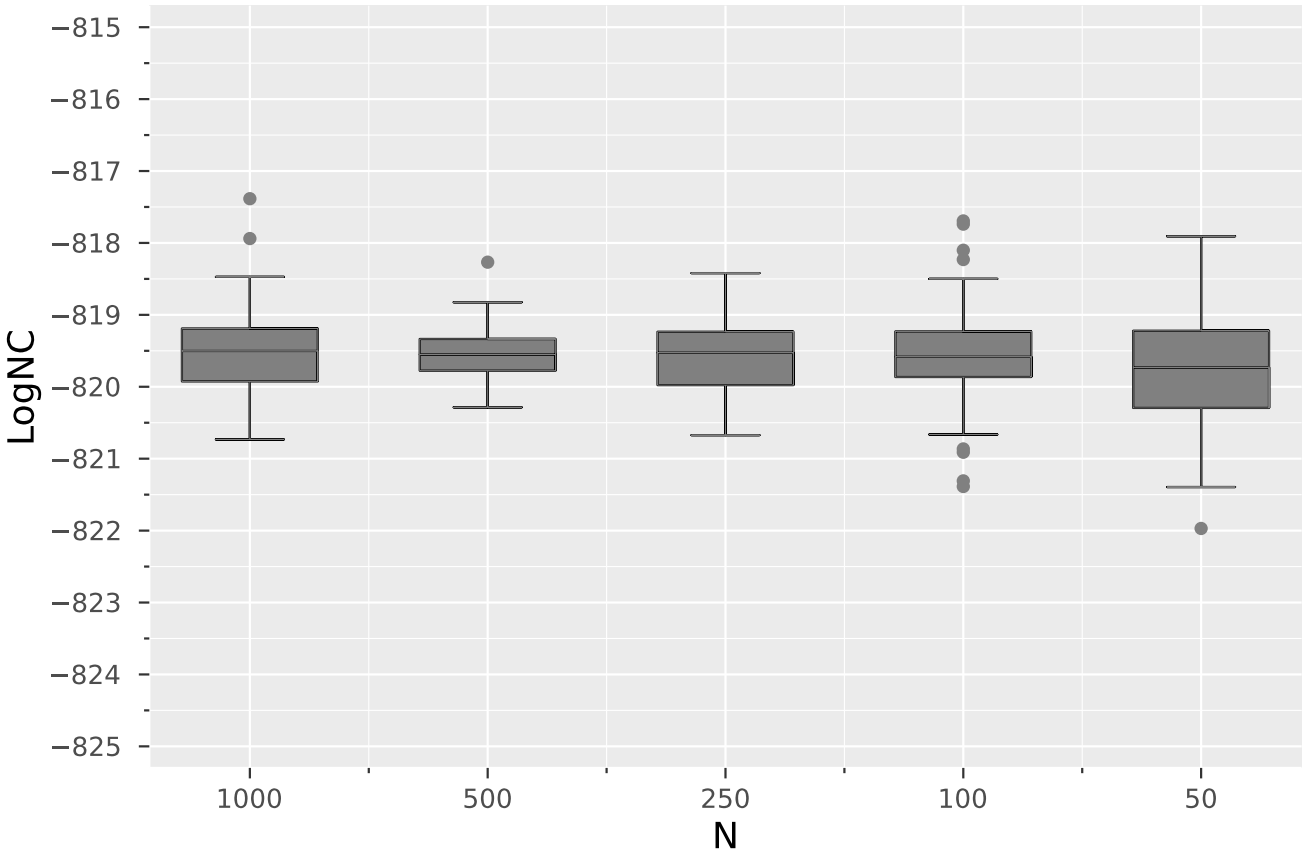

Figure 9: Orthant probability example: estimates of the normalising constant (i.e. the orthant probability) obtained from the waste-recycling HSMC algorithm with $N = 50,000$.

the waste-recycling HSMC algorithm does scale well in high dimension. We note that this higher variance compared to the waste-free SMC of [14], is obtained in a scenario where they are able to exploit the particular structure of the problem and implement an exact Gibbs sampler to move the particles. The integrator snippet SMC we propose is however more general and applicable to scenarios where such a structure is not present.

4 Preliminary theoretical exploration

In this section we omit the index $n \in [\![0, P]\!]$ and provide precise elements to understand the properties of the algorithms and estimators considered in this manuscript.

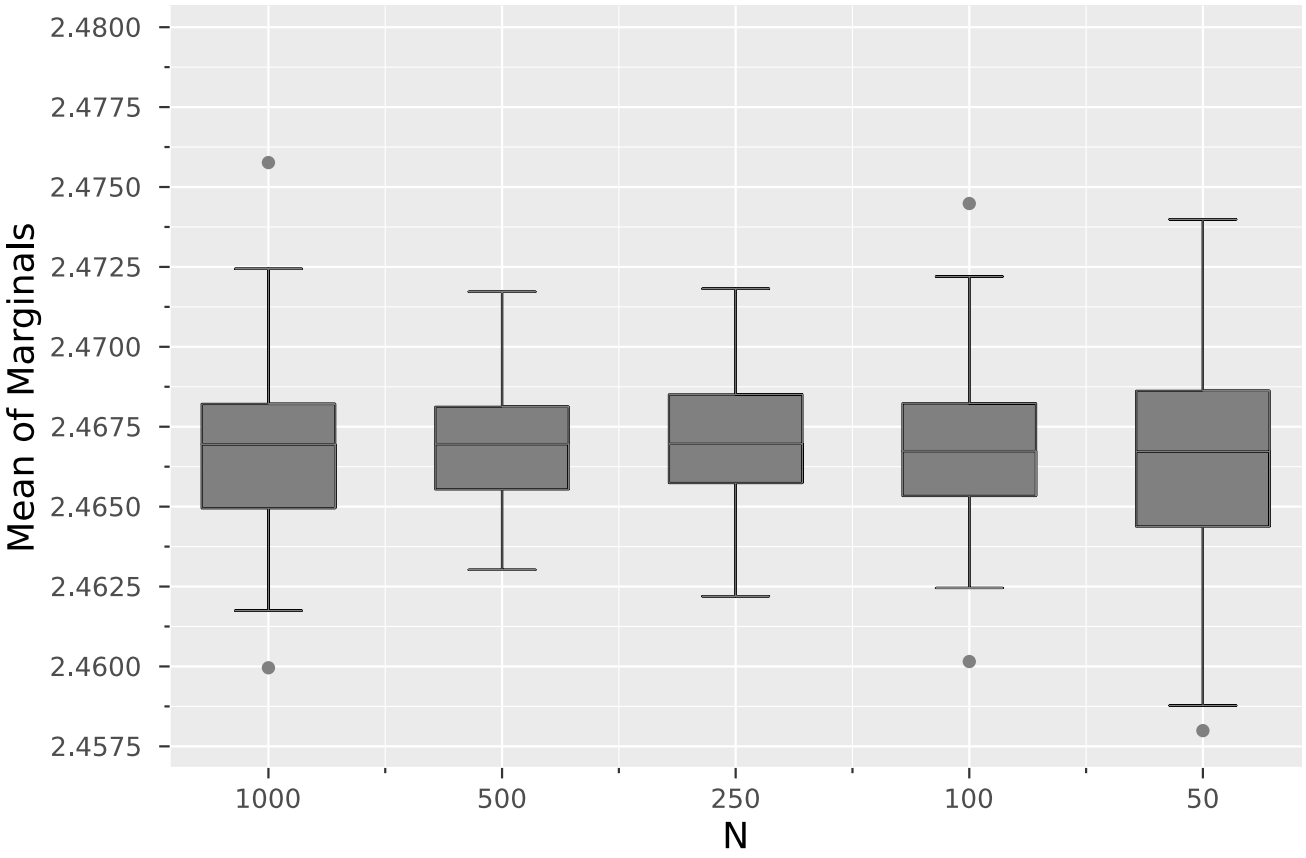

Figure 10: Orthant probability example: estimates of the mean of marginals obtained from the integrator snippet SMC algorithm with $N = 50,000$.

4.1 Variance using folded mixture samples

This section establishes variance reduction for an estimator of $\mu(f)$ of the type (15), but where it is assumed that $\{\check{z}_n^{(i)}, i \in [\![N]\!]\}$ are iid distributed according to $\bar{\mu}_n$. While this is not the exact distribution in the algorithm this provides a proxy representative of the quantities one needs to control when analysing SMC samplers [15, Chapter 9]: variance of estimators in an SMC can be decomposed as the sum of local variances at each iterations, which convergence to the variance terms considered hereafter as $N \to \infty$. The main message is that $\{f \circ \psi^k(z), k \in [\![P]\!]\}$ can be understood as playing the rôle of control variates, with the potential of reducing variance.

We use the following simplified notation throughout. Let

$$\bar{\mu}(k, \mathrm{d}z) := \frac{1}{T+1} \mu_k(\mathrm{d}z)\,,$$

with $\mu_k := \mu^{\psi^{-k}}$. The fundamental property exploited throughout is (11), which can be rephrased as, for $f \colon \mathsf{Z} \to \mathbb{R}$

μ-integrable,

$$\int \Big\{ \sum_{k=0}^{T} f \circ \psi^k(z) \bar{\mu}(k \mid z) \Big\} \bar{\mu}(\mathrm{d}z) = \mu(f) \,,$$

or written differently with $\bar{f}(k, z) := f \circ \psi^k(z)$

$$\mathbb{E}_{\bar{\mu}} \left(\mathbb{E}_{\bar{\mu}}(\bar{f}(\mathrm{T}, \check{Z}) \mid \check{Z}) \right) = \mathbb{E}_{\bar{\mu}} \left(\bar{f}(\mathrm{T}, \check{Z}) \right) = \mathbb{E}_{\mu} \left(f(Z) \right).$$

The estimator we use is therefore a so-called "Rao-Blackwellized estimator", assuming $\check{Z}^{(1)}, \ldots, \check{Z}^{(N)} \overset{\text{iid}}{\sim} \bar{\mu}$,

$$\check{\mu}(f) = N^{-1} \sum_{i=1}^{N} \mathbb{E}_{\bar{\mu}} \big(\bar{f}(\mathrm{T}, \check{Z}^{(i)}) \mid \check{Z}^{(i)} \big) \,,$$

of variance $\mathrm{var}_{\bar{\mu}} \left(\mathbb{E}_{\bar{\mu}} \big(\bar{f}(\mathrm{T}, \check{Z}) \mid \check{Z} \big) \right) / N$. The following relates this variance to that of the standard estimator using N iid samples from μ, which would likewise play a role in the asymptotic properties of SMC algorithms.

Proposition 15. *We have*

$$\mathrm{var}_{\bar{\mu}} \left(\mathbb{E}_{\bar{\mu}} \big(\bar{f}(\mathrm{T}, \check{Z}) \mid \check{Z} \big) \right) = \mathrm{var}_{\mu} \left(f(\check{Z}) \right) - \mathbb{E}_{\bar{\mu}} \big(\mathrm{var}_{\bar{\mu}}(\bar{f}(\mathrm{T}, \check{Z}) \mid \check{Z}) \big)$$

Proof. The variance decomposition identity yields

$$\mathrm{var}_{\bar{\mu}} \left(\bar{f}(\mathrm{T}, \check{Z}) \right) = \mathrm{var}_{\bar{\mu}} \left(\mathbb{E}_{\bar{\mu}} \big(\bar{f}(\mathrm{T}, \check{Z}) \mid \check{Z} \big) \right) + \mathbb{E}_{\bar{\mu}} \big(\mathrm{var}_{\bar{\mu}}(\bar{f}(\mathrm{T}, \check{Z}) \mid \check{Z}) \big) \,,$$

but from the fundamental property

$$\begin{aligned} \mathrm{var}_{\bar{\mu}} \left(\bar{f}(\mathrm{T}, \check{Z}) \right) &= \mathbb{E}_{\bar{\mu}} \left(\bar{f}(\mathrm{T}, \check{Z})^2 \right) - \mathbb{E}_{\bar{\mu}} \left(\bar{f}(\mathrm{T}, \check{Z}) \right)^2 \\ &= \mathbb{E}_{\mu} \left(f(\check{Z})^2 \right) - \mathbb{E}_{\mu} \left(f(\check{Z}) \right)^2 \\ &= \mathrm{var}_{\mu} \left(f(\check{Z}) \right). \end{aligned}$$

□

The following provides us with a notion of effective sample size for estimators of the form $\check{\mu}(f)$.

Theorem 16. *Assume that $\upsilon \gg \mu^{\psi^{-k}}$ for $k \in [\![T]\!]$, then with $\check{Z} \sim \bar{\mu}$,*

1. *for $f \colon \mathsf{Z} \to \mathbb{R}$,*

$$\mathrm{var} \left(\check{\mu}(f) \right) \leq \frac{2}{N} \frac{\mathrm{var} \big(\sum_{k=0}^{T} \mu \circ \psi^k(\check{Z}) f \circ \psi^k(\check{Z}) \big) + \|f\|_{\infty}^2 \mathrm{var} \big(\sum_{k=0}^{T} \mu \circ \psi^k(\check{Z}) \big)}{\mathbb{E}_{\bar{\mu}} \big(\sum_{k=0}^{T} \mu \circ \psi^k(\check{Z}) \big)^2} \,,$$

2. *further*

$$\mathrm{var} \left(\check{\mu}(f) \right) \leq \frac{2 \|f\|_{\infty}^2}{N} \left\{ \frac{2 \mathbb{E}_{\bar{\mu}} \big((\sum_{k=0}^{T} \mu \circ \psi^k(\check{Z}))^2 \big)}{\mathbb{E}_{\bar{\mu}} \big(\sum_{k=0}^{T} \mu \circ \psi^k(\check{Z}) \big)^2} - 1 \right\} .$$

Remark 17. Multiplying μ by any constant $C_\mu > 0$ does not alter the values of the upper bounds, making the results relevant to the scenario where μ is known up to a constant only. From the second statement we can define a notion of effective sample size (ESS)

$$\frac{N}{2}\frac{\mathbb{E}_{\bar{\mu}}\left(\sum_{k=0}^{T}\mu\circ\psi^k(\check{Z})\right)^2}{2\mathbb{E}_{\bar{\mu}}\left((\sum_{k=0}^{T}\mu\circ\psi^k(z))^2\right)-\mathbb{E}_{\bar{\mu}}\left(\sum_{k=0}^{T}\mu\circ\psi^k(\check{Z})\right)^2},$$

which could be compared to $N(T+1)$ or N.

Proof. We have

$$\mathrm{var}\left(\check{\mu}(f)\right)=\frac{1}{N}\mathrm{var}_{\bar{\mu}}\left(\frac{\sum_{k=0}^{T}\mu\circ\psi^k(\check{Z})f\circ\psi^k(\check{Z})}{\sum_{l=0}^{T}\mu\circ\psi^l(\check{Z})}\right)$$

and we apply Lemma 21

$$\begin{aligned}\mathrm{var}_{\bar{\mu}}\left(\frac{\sum_{k=0}^{T}\mu\circ\psi^k(\check{Z})f\circ\psi^k(\check{Z})}{\sum_{l=0}^{T}\mu\circ\psi^l(\check{Z})}\right)&\leq 2\frac{\mathrm{var}\left(\sum_{k=0}^{T}\mu\circ\psi^k(\check{Z})f\circ\psi^k(\check{Z})\right)+\|f\|_\infty^2\mathrm{var}\left(\sum_{k=0}^{T}\mu\circ\psi^k(\check{Z})\right)}{\mathbb{E}_{\bar{\mu}}\left(\sum_{k=0}^{T}\mu\circ\psi^k(z)\right)^2}\\&\leq 2\|f\|_\infty^2\left\{\frac{2\mathbb{E}_{\bar{\mu}}\left((\sum_{k=0}^{T}\mu\circ\psi^k(\check{Z}))^2\right)}{\mathbb{E}_{\bar{\mu}}\left(\sum_{k=0}^{T}\mu\circ\psi^k(\check{Z})\right)^2}-1\right\}.\end{aligned}$$

□

4.2 Variance using unfolded mixture samples

For $\psi\colon \mathsf{Z}\to\mathbb{R}$ invertible, ν a probability distribution such that $\nu \gg \mu^{\psi^{-1}}$ and $f\colon \mathsf{Z}\to\mathbb{R}$ such that $\mu(f)$ exists we have the identity

$$\mu(f)=\int f\circ\psi(z)\mu^{\psi^{-1}}(\mathrm{d}z)=\int f\circ\psi(z)\frac{\mathrm{d}\mu^{\psi^{-1}}}{\mathrm{d}\nu}(z)\nu(\mathrm{d}z)\,.$$

As a result for $\{\psi_k\colon \mathsf{Z}\to\mathbb{R}, k\in[\![K]\!]\}$, all invertible and such that $\nu\gg\mu^{\psi_k^{-1}}$ for $k\in[\![K]\!]$, we have

$$\mu(f)=\int\left\{\frac{1}{K}\sum_{k\in[\![K]\!]}f\circ\psi_k(z)\frac{\mathrm{d}\mu^{\psi_k^{-1}}}{\mathrm{d}\nu}(z)\right\}\nu(\mathrm{d}z)\,, \tag{30}$$

which suggests the Pushforward Importance Sampling (PISA) estimator, for $Z^i\overset{\text{iid}}{\sim}\nu$, $i\in[\![N]\!]$ and with $\bar{\mu}_{/\nu}(f\mid z):=K^{-1}\sum_{k\in[\![K]\!]}f\circ\psi_k(z)\frac{\mathrm{d}\mu^{\psi_k^{-1}}}{\mathrm{d}\nu}(z)$

$$\hat{\mu}(f)=\sum_{i=1}^{N}\frac{\bar{\mu}_{/\nu}(f\mid Z^i)}{\sum_{j=1}^{N}\bar{\mu}_{/\nu}(1\mid Z^j)}=\frac{\frac{1}{N}\sum_{i=1}^{N}\bar{\mu}_{/\nu}(f\mid Z^i)}{\frac{1}{N}\sum_{j=1}^{N}\bar{\mu}_{/\nu}(1\mid Z^j)}\,, \tag{31}$$

This is the estimator in (7) when $\nu=\mu_{n-1}$ and $\mu=\mu_n$.

4.2.1 Relative efficiency for unfolded estimators

In order to define the notion of relative efficiency for the estimator $\hat{\mu}(f)$ in (31) we first establish the following bounds.

Theorem 18. *With the notation above and $Z \sim \nu$ throughout. For any $f\colon \mathsf{Z} \to \mathbb{R}$,*

1.

$$\mathrm{var}\big(\hat{\mu}(f)\big) \leq \mathbb{E}\Big(|\hat{\mu}(f) - \mu(f)|^2\Big) \leq \frac{2}{N}\big\{\mathrm{var}\big(\bar{\mu}_{/\nu}(f \mid Z)\big) + \|f\|_\infty^2 \mathrm{var}\big(\bar{\mu}_{/\nu}(1 \mid Z)\big)\big\}\,,$$

2. more simply

$$\mathbb{E}\Big(|\hat{\mu}(f) - \mu(f)|^2\Big) \leq \frac{2\|f\|_\infty^2}{KN}\Big\{\frac{2\mathbb{E}\Big[\big(\sum_{k\in[\![K]\!]} \mathrm{d}\mu^{\psi_k^{-1}}/\mathrm{d}\nu(Z)\big)^2\Big]}{K} - K\Big\}\,, \tag{32}$$

3. and

$$|\mathbb{E}[\hat{\mu}(f)] - \mu(f)| \leq \frac{2\|f\|_\infty^2}{KN}\frac{\mathbb{E}\Big[\big(\sum_{k\in[\![K]\!]} \mathrm{d}\mu^{\psi_k^{-1}}/\mathrm{d}\nu(Z)\big)^2\Big]}{K}\,.$$

Remark 19. The upper bound in the first statement confirms the control variate nature of integrator snippets, even when using the unfolded perspective, a property missed by the rougher bounds of the last two statements.

Remark 20 (ESS for PISA). The notion of efficiency is usually defined relative to the "perfect" Monte Carlo scenario that is the standard estimator $\hat{\mu}_0$ of $\mu(f)$ relying on KN iid samples from μ for which we have

$$\mathrm{var}\Big(\hat{\mu}_0(f)\Big) = \frac{\mathrm{var}_\mu(f)}{KN} \leq \frac{\|f\|_\infty^2}{KN}\,. \tag{33}$$

The RE_0, is determined by the ratio of the upper bound in (32) by (33). Our point below is that the notion of efficiency can be defined relative to any competing algorithm, virtual or not, in order to characterize particular properties. For example we can compute the efficiency relative to that of the "ideal" PISA estimator i.e. for which $\bar{\mu}_{/\nu}(\bar{f} \mid Z^i)$ is replaced with $K^{-1}\sum_{k\in[\![K]\!]} f \circ \psi_k(Z^{i,k})\frac{\mathrm{d}\mu^{\psi_k^{-1}}}{\mathrm{d}\nu}(Z^{i,k})$, $Z^{i,k} \overset{\mathrm{iid}}{\sim} \nu$ and

$$\mathrm{var}\Big(\hat{\mu}_1(f)\Big) \leq \frac{2\|f\|_\infty^2}{KN}\left\{\frac{2\sum_{k\in[\![K]\!]}\mathbb{E}\Big[\big(\mathrm{d}\mu^{\psi_k^{-1}}/\mathrm{d}\nu(Z)\big)^2\Big]}{K} - K\right\}\,. \tag{34}$$

The corresponding RE_1 captures the loss incurred because of dependence along a snippet. However, given our initial motivation of recycling the computation of a standard HMC based SMC algorithm we opt to define the RE_2 relative to the estimator relying on both ends of the snippet only, i.e.

$$\hat{\mu}_2(f) = \frac{1}{N}\sum_{i=1}^N \frac{\frac{1}{2}\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(z^i)f(Z^i) + \frac{1}{2}\frac{\mathrm{d}\mu^{\psi_K^{-1}}}{\mathrm{d}\nu}(Z^i)f\circ\psi_K(Z^i)}{\sum_{j=1}^N \frac{1}{2}\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(Z^j) + \frac{1}{2}\frac{\mathrm{d}\mu^{\psi_K^{-1}}}{\mathrm{d}\nu}(Z^j)}\,.$$

In the SMC scenario considered in this manuscript (see Section 1) the above can be thought of as a proxy for estimators obtained by a "Rao-Blackwellized" SMC algorithm using $P_{n-1,T}$ in (6), where N particles $\{z_{n-1}^{(i)}, i \in [\![N]\!]\}$ in Alg. 1 give rise to $2N$ weighted particles

$$\{\big(z^i, \bar{\alpha}_{n-1}(z_i;T)\tfrac{\mu_n}{\mu_{n-1}}(z_i)\big); \big(z^i, \alpha_{n-1}(z_i;T)\tfrac{\mu_n}{\mu_{n-1}}\circ\psi_n^T(z_i)\big), i \in [\![N]\!]\}\,,$$

with $\alpha_{n-1}(\cdot;T)$ defined in (5). Resampling with these weights is then applied to obtain N particles and followed by an update of velocities to yield $\{z_n^{(i)}, i \in [\![N]\!]\}$. Now, we observe the similarity between

$$\alpha_{n-1}(z;T)\tfrac{\mu_n}{\mu_{n-1}} \circ \psi_n^T(z) = \min\left\{1, \frac{\mu_{n-1}\circ\psi_{n-1}^T}{\mu_{n-1}}(z)\right\} \times \frac{\mu_n \circ \psi_n^T}{\mu_{n-1}\circ\psi_n^T}(z)\,, \tag{35}$$

and the corresponding weight in Alg. 2, in particular when μ_{n-1} and μ_n are similar, and hence ψ_{n-1} and ψ_n. This motivates our choice of reference to define ESS_2 which has a clear computational advantage since it involves ignoring $T-1$ terms only. In the present scenario, following Lemma 21, we have

$$\mathrm{var}\big(\hat{\mu}_2(f)\big) \leq \frac{\|f\|_\infty^2}{N}\left\{\mathbb{E}\left[\Big(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(Z) + \frac{\mathrm{d}\mu^{\psi_K^{-1}}}{\mathrm{d}\nu}(Z)\Big)^2\right] - \frac{1}{2}\right\},$$

which leads to the relative efficiency for PISA,

$$\mathrm{RE}_2 = \frac{\frac{4\mathbb{E}\left[\left(\sum_{k\in[\![K]\!]}\mathrm{d}\mu^{\psi_k^{-1}}/\mathrm{d}\nu(Z)\right)^2\right]}{K^2} - 2}{\mathbb{E}\left[\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(Z) + \frac{\mathrm{d}\mu^{\psi_K^{-1}}}{\mathrm{d}\nu}(Z)\right)^2\right] - 2},$$

which can be estimated using empirical averages.

Proof of Theorem 18. We apply Lemma 21 with

$$A(Z^1,\ldots,Z^N) = \frac{1}{N}\sum_{i=1}^N \bar{\mu}_{/\nu}(f \mid Z^i) \quad B(Z^1,\ldots,Z^N) = \frac{1}{N}\sum_{j=1}^N \bar{\mu}_{/\nu}(1 \mid Z^j)\,.$$

With $Z^i \overset{\text{iid}}{\sim} \nu$, we have directly

$$a = \mathbb{E}(A) = \mathbb{E}\big(\bar{\mu}_{/\nu}(f \mid Z)\big) = \mu(f) \quad b = \mathbb{E}(B) = 1\,,$$

and

$$\begin{aligned}\mathrm{var}(B) &= \frac{1}{N}\mathrm{var}\big(\bar{\mu}_{/\nu}(1 \mid Z^j)\big)\\ &= \frac{1}{K^2 N}\left\{\mathbb{E}\Big[\big(\textstyle\sum_{k\in[\![K]\!]}\mathrm{d}\mu^{\psi_k^{-1}}/\mathrm{d}\nu(Z)\big)^2\Big] - K^2\right\}.\end{aligned}$$

Now with $\|f\|_\infty < \infty$

$$\begin{aligned}\mathrm{var}(A) &= \frac{1}{N}\mathrm{var}\big(\bar{\mu}_{/\nu}(f \mid Z)\big)\\ &= \frac{1}{K^2 N}\left\{\|f\|_\infty^2\mathbb{E}\Big[\big(\textstyle\sum_{k\in[\![K]\!]}\mathrm{d}\mu^{\psi_k^{-1}}/\mathrm{d}\nu(Z)\big)^2\Big] - K^2\mu(f)^2\right\}\\ &\leq \frac{\|f\|_\infty^2}{K^2 N}\mathbb{E}\Big[\big(\textstyle\sum_{k\in[\![K]\!]}\mathrm{d}\mu^{\psi_k^{-1}}/\mathrm{d}\nu(Z)\big)^2\Big].\end{aligned}$$

We conclude noting that we have $|A/B| \leq \|f\|_\infty$. □

For clarity we reproduce the very useful lemma [38, Lemma 6], correcting a couple of minor typos along the way.

Lemma 21. *Let A, B be two integrable random variables satisfying $|A/B| \leq M$ almost surely for some $M > 0$ and let $a = \mathbb{E}(A)$ and $b = \mathbb{E}(B) \neq 0$. Then*

$$\mathrm{var}(A/B) \leq \mathbb{E}[|A/B - a/b|^2] \leq \frac{2}{b^2}\Big\{\mathrm{var}(A) + M^2\mathrm{var}(B)\Big\}\,,$$
$$|\mathbb{E}[A/B] - a/b| \leq \frac{\sqrt{\mathrm{var}(A/B)\mathrm{var}(B)}}{b}\,.$$

4.2.2 Optimal weights

As mentioned in Subsection 2.3 it is possible to consider the more general scenario where unequal probability weights $\omega = \{\omega_k \in \mathbb{R}, k \in [\![0, T]\!] \colon \sum_{k=0}^T \omega_k = 1\}$ are ascribed to the elements of the snippets, yielding the estimator

$$\hat{\mu}_\omega(f) := \frac{\sum_{i=1}^N \sum_{k=0}^T \omega_k \frac{\mathrm{d}\mu^{\psi_k^{-1}}}{\mathrm{d}\nu}(Z^i) f \circ \psi_k(Z^i)}{\sum_{j=1}^N \sum_{l=0}^T \omega_l \frac{\mathrm{d}\mu^{\psi_l^{-1}}}{\mathrm{d}\nu}(Z^j)}\,,$$

and a natural question is that of the optimal choice of ω. Note that in the context of PISA the condition $\omega_k \geq 0$ is not required, as suggested by the justification of the identity (30). However it should be clear that this condition should be enforced in the context of integrator snippet SMC, since the probabilistic interpretation is otherwise lost, or if the expectation is known to be non-negative. Here we discuss optimization of the variance upperbound provided by Lemma 21,

$$\frac{2\|f\|_\infty^2}{N}\Big\{\mathrm{var}\Big(\sum_{k\in[\![K]\!]}\omega_k \frac{\mathrm{d}\mu^{\psi_k^{-1}}}{\mathrm{d}\nu}(Z)\frac{f\circ\psi_k}{\|f\|_\infty}(Z)\Big) + \mathrm{var}\Big(\sum_{k\in[\![K]\!]}\omega_k \frac{\mathrm{d}\mu^{\psi_k^{-1}}}{\mathrm{d}\nu}(Z)\Big)\Big\}$$
$$\leq \frac{2\|f\|_\infty^2}{N}\omega^\top\big(\Sigma(f;\psi) + \Sigma_\psi(1;\psi)\big)\omega$$

where for $k, l \in [\![K]\!]$,

$$\Sigma_{kl}(f;\psi) = \mathbb{E}\left[\frac{\mathrm{d}\mu^{\psi_k^{-1}}}{\mathrm{d}\nu}(Z)\frac{f\circ\psi_k}{\|f\|_\infty}(Z)\frac{\mathrm{d}\mu^{\psi_l^{-1}}}{\mathrm{d}\nu}(Z)\frac{f\circ\psi_l}{\|f\|_\infty}(Z)\right] - \mu(f/\|f\|_\infty)^2\,.$$

It is a classical result that $\omega^\top\big(\Sigma(f;\psi) + \Sigma_\psi(1;\psi)\big)\omega \geq \lambda_{\min}\big(\Sigma(f;\psi) + \Sigma_\psi(1;\psi)\big)\omega^\top\omega$ and that minimum is reached for the eigenvector(s) $\omega_{\min}$ corresponding to the smallest eigenvalue $\lambda_{\min}\big(\Sigma(f;\psi) + \Sigma_\psi(1;\psi)\big)$ of $\Sigma(f;\psi) + \Sigma_\psi(1;\psi)$. If the constraint of non-negative entries is to be enforced then a standard quadratic programming procedure should be used. The same ideas can be applied to the function independent upperbounds used in our definitions of efficiency.

4.3 More on variance reduction and optimal flow

We now focus on some properties of the estimator $\hat{\mu}(f)$ of $\mu(f)$ in (16). To facilitate the analysis and later developments we consider the scenario where, with $t \mapsto \psi_t(z)$ the flow solution of an ODE, assume the dominating measure $\upsilon \gg \mu$ and υ is invariant by the flow. We consider

$$\bar{\mu}(\mathrm{d}t, \mathrm{d}z) = \frac{1}{T}\mu^{\psi_{-t}}(\mathrm{d}z)\mathbf{1}\{0 \leq t \leq T\}\mathrm{d}t\,,$$

and notice that similarly to the integrator scenario, for any $f\colon \mathsf{Z} \to \mathbb{R}$, $\bar{f}(t,z) := f \circ \psi_t(z)$ we have $\mathbb{E}_{\bar{\mu}}\left(\bar{\mu}(\bar{f} \mid \check{Z})\right) = \bar{\mu}(\bar{f}) = \mu(f)$. where for any $z \in \mathsf{Z}$, where

$$\bar{\mu}(\bar{f} \mid z) = \frac{1}{T}\int_0^T f \circ \psi_t(z) \frac{\mu \circ \psi_t(z)}{\int \mu \circ \psi_u(z) \mathrm{d}u} \mathrm{d}t$$

the following estimator of $\mu(f)$ is considered, with $\check{Z}_i \overset{\text{iid}}{\sim} \bar{\mu}$

$$\hat{\mu}(f) = \frac{1}{N}\sum_{i=1}^{N} \bar{\mu}\big(\bar{f}(\mathrm{T}, \check{Z}_i) \mid \check{Z}_i\big)\,.$$

We now show that in the examples considered in this paper our approach can be understood as implementing unbiased Riemann sum approximations of line integrals along contours. Adopting a continuous time approach can be justified as follows. For numerous integrators, the conditions of the following proposition are satisfied; this is the case for the leapfrog integrator of Hamilton's equation e.g. [20, 38, Theorem 3.4] and [38, Appendix 3.1, Theorem 9] for detailed results.

Proposition 22. *Let $\tau > 0$, for any $z \in \mathsf{Z}$ let $[0,\tau] \ni t \mapsto \psi_t(z)$ be a flow and for $\epsilon > 0$ let $\{\psi^k(z;\epsilon), 0 \leq k\epsilon \leq \tau\}$ be a discretization of $t \mapsto \psi_t$ such that that for any $z \in \mathsf{Z}$ there exists $C > 0$ such that for any $(k,\epsilon) \in \mathbb{N} \times \mathbb{R}_+$*

$$|\psi^k(z;\epsilon) - \psi_{k\epsilon}(z)| \leq C\epsilon^2\,.$$

Then for any continuous $g\colon \mathsf{Z} \to \mathbb{R}$ such that the Riemann integral

$$I(g) := \int_0^\tau g \circ \psi_t(z) \mathrm{d}t\,,$$

exists we have

$$\lim_{T\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} g \circ \psi_{k\tau/n}(z;\tau/n) = I(g)\,.$$

Proof. We have

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} g \circ \psi_{k\tau/n}(z) = \int_0^\tau g \circ \psi_t(z) \mathrm{d}t$$

$$\lim_{n\to\infty} \frac{1}{n}\left|\sum_{k=0}^{n-1} g \circ \psi_{k\tau/n}(z) - g \circ \psi^k(z;\tau/n)\right| = 0\,,$$

and we can conclude. □

Remark 23. Naturally convergence is uniform in $z \in \mathsf{Z}$ with additional assumptions and we note that in some scenarios dependence on τ can be characterized, allowing in principle τ to grow with n .

4.3.1 Hamiltonian contour decomposition

Assume μ has a probability density with respect to the Lebesgue measure and let $\zeta\colon \mathsf{Z} \subset \mathbb{R}^d \to \mathbb{R}$ be Lipschitz continuous such that $\nabla\zeta(z) \neq 0$ for all $z \in \mathsf{Z}$, then the co-area formula states that

$$\int_{\mathsf{Z}} f(z)\mu(z)\mathrm{d}z = \int \Big[\int_{\zeta^{-1}(s)} f(z)|\nabla\zeta(z)|^{-1}\mu(z)\mathcal{H}_{d-1}(\mathrm{d}z)\Big]\mathrm{d}s,$$

where $\mathcal{H}_{d-1}$ is the $(d-1)$-dimensional Hausdorff measure, used here to measure length along the contours $\zeta^{-1}(s)$. For example in the HMC setup where $\mu(z) \propto \exp\big(-H(z)\big)$ and $z = (x, v)$ one may choose $\zeta(z) = H(z) = -\log\big(\mu(z)\big)$, leading to a decomposition of the expectation $\mu(f)$ according to equi-energy contours of $H(z)$

$$\mu(f) = \int \exp(-s)\big[\int_{\zeta^{-1}(s)} f(z)|\nabla H(z)|^{-1}\mathcal{H}_{d-1}(\mathrm{d}z)\big]\mathrm{d}s.$$

We now show how the solution of Hamilton's equation could be used as the basis for estimators of $\mu(f)$ mixing Riemanian sum-like and Monte Carlo estimation techniques.

Favourable scenario where $d = 2$. Let $s \in \mathbb{R}$ such that $H^{-1}(s) \neq \emptyset$ and assume that for some $(x_0, v_0) \in H^{-1}(s)$ Hamilton's equations $\dot{x} = -\nabla_v H(x, v)$ and $\dot{v} = \nabla_x H(x, v)$ have solutions $t \mapsto (x_t, v_t) = \psi_t(z_0) \in H^{-1}(s)$ with $\big(x_{\tau(s)}, v_{\tau(s)}\big) = (x_0, v_0) = z_0$ for some $\tau(s) > 0$, that is the contours $H^{-1}(s)$ can be parametrised with the solutions of Hamilton's equation at corresponding level.

Proposition 24. *We have*

$$\int_{\mathsf{Z}} f(z)\mu(\mathrm{d}z) = \int_{\mathsf{Z}} \left[[\tau \circ H(z_0)]^{-1} \int_0^{\tau\circ H(z_0)} f(z_t)\mathrm{d}t\right] \mu(\mathrm{d}z_0)\,,$$

implying that, assuming the integral along the path $t \mapsto z_t$ is tractable,

$$[\tau \circ H(z_0)]^{-1} \int_0^{\tau\circ H(z_0)} f(z_t)\mathrm{d}t \,\textit{for}\, z_0 \sim \mu\,, \tag{36}$$

is an unbiased estimator of $\mu(f)$.

Proof. Since $H^{-1}(s)$ is rectifiable we have that [9, Theorem 2.6.2] for $g\colon \mathsf{Z} \to \mathbb{R}$,

$$\int_{H^{-1}(s)} g(z)\mathcal{H}_{d-1}(\mathrm{d}z) = \int_0^{\tau(s)} g(z_t)|\nabla H(z_t)|\mathrm{d}t\,.$$

Consequently

$$\int_{\mathsf{Z}} f(z)\mu(\mathrm{d}z) = \int \tau(s)\exp(-s)\left[\tau(s)^{-1}\int_0^{\tau(s)} f(z_{s,t})\mathrm{d}t\right]\mathrm{d}s\,, \tag{37}$$

where the notation now reflects dependence on s of the flow and notice that since

$$H^{-1}(s) \ni z_0 \mapsto [\tau \circ H(z_0)]^{-1}\int_0^{\tau\circ H(z_0)} f(z_t)$$

is constant, $z_{0,s} \in H^{-1}(s)$ can be arbitrary. since indeed from (37)

$$\begin{aligned}\int_{\mathsf{Z}} \mathbf{1}\{H(z) \in A\}\mu(\mathrm{d}z) &= \int \tau(s)\exp(-s)\left[\tau(s)^{-1}\int_0^{\tau(s)} \mathbf{1}\{s \in A\}\mathrm{d}t\right]\mathrm{d}s \\ &= \int \tau(s)\exp(-s)\mathbf{1}\{s \in A\}\mathrm{d}s.\end{aligned}$$

Note that we can write, since for $s \in \mathbb{R}$, $H^{-1}(s) \ni z_0 \mapsto [\tau \circ H(z_0)]^{-1} \int_0^{\tau \circ H(z_0)} f(z_t)$ is constant,

$$\begin{aligned}
&\int_{\mathsf{Z}} \left[[\tau \circ H(z_0)]^{-1} \int_0^{\tau \circ H(z_0)} f(z_t) \mathrm{d}t \right] \mu(\mathrm{d}z_0) \\
&\qquad = \int \tau(s) \exp(-s) \left\{ \tau(s)^{-1} \int_0^{\tau(s)} \left[\left(\tau \circ H(z_{s,0})\right)^{-1} \int_0^{\tau \circ H(z_0)} f(z_{s,t}) \mathrm{d}t \right] \right\} \mathrm{d}t\mathrm{d}s \\
&\qquad = \int \tau(s) \exp(-s) \left[\tau(s)^{-1} \int_0^{\tau(s)} f(z_{s,t}) \mathrm{d}t \right] \mathrm{d}t\mathrm{d}s \,.
\end{aligned}$$

□

A remarkable point is that the strategy developed in this manuscript provides a general methodology to implement numerically the ideas underpinning the decomposition of Proposition 24 by using the estimator $\check{\mu}(f)$ in (15) and invoking Proposition 22, assuming $s \mapsto \tau(s)$ to be known. This point is valid outside of the SMC framework and it is worth pointing out that $\check{\mu}(f)$ in (15) is unbiased if the samples used are marginally from $\bar{\mu}$.

Remark the promise of dimension free estimators if the one dimensional line integrals in (36) were easily computed and sampling from the one-dimensional energy distribution was routine – however the scenario $d \geq 3$ is more subtle.

General scenario In the scenario where $d \geq 3$ the co-area decomposition still holds, but the solution to Hamilton's equation can in general not be used to compute integrals over the hypersurface $H^{-1}(s)$. This would require a form of ergodicity [35, 40] of the form, for $z_0 \in H^{-1}(s)$,

$$\lim_{\tau \to \infty} \frac{1}{\tau} \int_0^{\tau} f \circ \psi_t(z_0) \mathrm{d}t = \bar{f}(z_0) = \frac{\int_{\zeta^{-1}(s)} f(z) |\nabla H(z)|^{-1} \mathcal{H}_{d-1}(\mathrm{d}z)}{\int_{\zeta^{-1}(s)} |\nabla H(z)|^{-1} \mathcal{H}_{d-1}(\mathrm{d}z)} \,,$$

where the limit always exists in the $\mathrm{L}^2(\mu)$ sense and constitutes Von Neumann's mean ergodic theorem [41, 35], and the rightmost equality forms Boltzman's conjecture. An interesting property in the present context is that $\mathbb{E}_{\bar{\mu}}\big(\bar{f}(Z)\big) = \mathbb{E}_{\mu}\big(f(Z)\big)$ for $f \in \mathrm{L}^2(\mu)$ and one could replicate the variance reduction properties developed earlier. Boltzman's conjecture has long been disproved, as numerous Hamiltonians can be shown not to lead to ergodic systems, although some sub-classes do. However a weaker, or local, form or ergodicity can hold on sets of a partition of Z

Example 25 (Double well potential)**.** Consider the scenario where $\mathsf{X} = \mathsf{V} = \mathbb{R}$, $U(x) = (x^2 - 1)^2$ and kinetic energy v^2 [34]. Elementary manipulations show that satisfying Hamilton's equations (2) imposes $t \mapsto H(x_t, \dot{x}_t) = (x_t^2 - 1)^2 + \dot{x}_t^2 = s > 0$ and therefore requires $t \mapsto x_t \in [-\sqrt{1+\sqrt{s}}, -\sqrt{1-\sqrt{s}}] \cup [\sqrt{1-\sqrt{s}}, \sqrt{1+\sqrt{s}}]$ – importantly the intervals are not connected for $s < 1$. Rearranging terms any solution of (2) must satisfy

$$\dot{x}_t = \pm \sqrt{s - (x_t^2 - 1)^2} \,,$$

that is the velocity is a function of the position in the double well, maximal for $x_t^2 = 1$, vanishing as $x_t^2 \to 1 \pm \sqrt{s}$ and a sign flip at the endpoints of the intervals. Therefore the system is not ergodic, but ergodicity trivially occurs in each well.

In general this partitioning of Z can be intricate but it should be clear that in principle this could reduce variability of an estimator. In the toy Example 25, a purely mathematical algorithm inspired by the discussions

above would choose the right or left well with probability $1/2$ and then integrate deterministically, producing samples taking at most two values which could be averaged to compute $\mu(f)$. We further note that in our context a relevant result would be concerned with the limit, for any $x_0 \in \mathsf{X}$,

$$\lim_{\tau\to\infty} \int \big[\frac{1}{\tau}\int_0^\tau f\circ\psi_t(x_0,\rho\mathbf{e})\mathrm{d}t\big]\varpi_{\mathbf{e}}(\mathrm{d}\mathbf{e})\,,$$

where we have used a polar reparametrization $v = \rho\mathbf{e}$ $(\rho,\mathbf{e}) \in \mathbb{R}_+ \times \mathcal{S}_{d-1}$, and whether a marginal version of Boltzman's conjecture holds for this limit. Example 25 indicates that this is not true in general but the cardinality of the partition of X may be reduced.

4.3.2 Advantage of averaging and control variate interpretation

Consider the scenario where $(\mathsf{T},\check{Z}) \sim \bar{\mu}(t,\mathrm{d}z) = \frac{1}{\tau}\mu^{\psi_{-t}}(\mathrm{d}z)\mathbf{1}\{0\le t\le\tau\}$, $[0,\tau] \ni t \mapsto \psi_t$ for some $\tau > 0$ is the flow solution of an ODE of the form $\dot{z}_t = F\circ z_t$ for some field $F\colon \mathsf{Z}\to\mathsf{Z}$. For $f\colon \mathsf{Z}\to\mathbb{R}$ we have

$$\mathrm{var}_\mu\left(f(Z)\right) = \mathrm{var}_{\bar\mu}\left(\mathbb{E}_{\bar\mu}\big(\bar f(\mathsf{T},\check{Z})\mid\check{Z}\big)\right) + \mathbb{E}_{\bar\mu}\big(\mathrm{var}_{\bar\mu}(\bar f(\mathsf{T},\check{Z})\mid\check{Z})\big)\,.$$

and we are interested in determining $t\mapsto\psi_t$ (or F) in order to minimize $\mathrm{var}_{\bar\mu}\left(\mathbb{E}_{\bar\mu}\big(\bar f(\mathsf{T},\check{Z})\mid\check{Z}\big)\right)$ and maximize improvement over simple Monte Carlo. We recall that the Jabobian determinant of the flow $t\mapsto\psi_t(z)$ is given by [19, p. 174108-5]

$$J_t(z) = |\det\big(\nabla\otimes\psi_t(z)\big)| = \mathcal{J}\circ\psi_t(z) \text{ with } \mathcal{J}(z) := \exp\left(\int_0^t(\nabla\cdot F)(z)\right)\,.$$

Lemma 26. *Let $t\mapsto\psi_t$ be a flow solution of $\dot{z}_t = F\circ z_t$ and assume that with υ the Lebesgue measure, $\upsilon \gg \mu$. Then*

$$\begin{aligned}\mathbb{E}_{\bar\mu}\left(\mathbb{E}_{\bar\mu}\big(\bar f(\mathsf{T},\check{Z})\mid\check{Z}\big)^2\right) &= 2\int_0^\tau\left\{\int_0^{\tau-u}[\bar\mu\circ\psi_{-t}(z)]^{-1}\mathrm{d}t\right\}\left\{\int\mu(\mathrm{d}z)f(z)f\circ\psi_u(z)\mu\circ\psi_u(z)\mathcal{J}\circ\psi_u(z)\right\}\mathrm{d}u\\ &\le 2\left\{\int_0^\tau[\bar\mu\circ\psi_{-t}(z)]^{-1}\mathrm{d}t\right\}\int_0^\tau\left\{\int\mu(\mathrm{d}z)f(z)f\circ\psi_u(z)\mu\circ\psi_u(z)\mathcal{J}\circ\psi_u(z)\right\}\mathrm{d}u\,,\end{aligned}$$

with

$$\bar\mu\circ\psi_{-t}(z) = \int_{-t}^{\tau-t}\mu\circ\psi_u(z)\,\mathcal{J}\circ\psi_u(z)\mathrm{d}u\,.$$

In particular for $t\mapsto\psi_t$ the flow solution of Hamilton's equations associated with μ we have

$$\mathbb{E}_{\bar\mu}\left(\mathbb{E}_{\bar\mu}\big(\bar f(\mathsf{T},\check{Z})\mid\check{Z}\big)^2\right) = 2\int(1-t/\tau)\int\mu(\mathrm{d}z)f(z)f\circ\psi_t(z)\mathrm{d}t\,.$$

Proof. Using Fubini's theorem we have

$$\begin{aligned}&\int\left\{\int_0^\tau f\circ\psi_t(z)\frac{\mathrm{d}\mu^{\psi_{-t}}}{\mathrm{d}\bar\mu}(z)\mathrm{d}t\right\}^2\bar\mu(\mathrm{d}z) =\\ &\quad= 2\int_0^\tau\int_0^\tau\mathbf{1}\{t'\ge t\}\int\bar\mu(\mathrm{d}z)f\circ\psi_t(z)\frac{\mathrm{d}\mu^{\psi_{-t}}}{\mathrm{d}\bar\mu}(z)f\circ\psi_{t'-t+t}(z)\frac{\mathrm{d}\mu^{\psi_{-t'}}}{\mathrm{d}\bar\mu}\circ\psi_{-t}\circ\psi_t(z)\mathrm{d}t'\mathrm{d}t\\ &\quad= 2\int_0^\tau\int_0^\tau\mathbf{1}\{t'\ge t\}\int\mu(\mathrm{d}z)f(z)f\circ\psi_{t'-t}(z)\frac{\mathrm{d}\mu^{\psi_{-t'}}}{\mathrm{d}\bar\mu}\circ\psi_{-t}(z)\mathrm{d}t'\mathrm{d}t\,.\end{aligned}$$

Further, we have

$$\frac{\mathrm{d}\mu^{\psi_{-t'}}}{\mathrm{d}\bar{\mu}}(z) = \frac{\mu \circ \psi_{t'}(z)\, \mathcal{J} \circ \psi_{t'}(z)}{\bar{\mu}(z)}\,,$$

with $\bar{\mu}(z) = \int_0^\tau \mu \circ \psi_u(z)\, \mathcal{J} \circ \psi_u(z) \mathrm{d}u$. It is straightforward that

$$\begin{aligned}\bar{\mu} \circ \psi_{-t}(z) &= \int_0^\tau \mu \circ \psi_{u-t}(z)\, \mathcal{J} \circ \psi_{u-t}(z) \mathrm{d}u \\ &= \int_{-t}^{\tau-t} \mu \circ \psi_{u'}(z)\, \mathcal{J} \circ \psi_{u'}(z) \mathrm{d}u'\,.\end{aligned}$$

Consequently

$$\begin{aligned}\int \left\{ \int_0^\tau f \circ \psi_t(z) \frac{\mathrm{d}\mu^{\psi_{-t}}}{\mathrm{d}\bar{\mu}}(z) \mathrm{d}t \right\}^2 \bar{\mu}(\mathrm{d}z) = \\ = 2 \int_0^\tau \int_0^\tau \frac{\mathbf{1}\{\tau - t \geq u \geq 0\}}{\bar{\mu} \circ \psi_{-t}(z)} \int \mu(\mathrm{d}z) f(z) f \circ \psi_u(z) \mu \circ \psi_u(z) \mathcal{J} \circ \psi_u(z) \mathrm{d}u \mathrm{d}t \\ = 2 \int_0^\tau \left\{ \int_0^{\tau-u} [\bar{\mu} \circ \psi_{-t}(z)]^{-1} \mathrm{d}t \right\} \int \mu(\mathrm{d}z) f(z) f \circ \psi_u(z) \mu \circ \psi_u(z) \mathcal{J} \circ \psi_u(z) \mathrm{d}u\end{aligned}$$

For the second statement we have $\bar{\mu} \circ \psi_{-t}(z) = \tau\, \mu(z)$ and

$$\begin{aligned}\int \left\{ \int_0^\tau f \circ \psi_t(z) \frac{\mathrm{d}\mu^{\psi_{-t}}}{\mathrm{d}\bar{\mu}}(z) \mathrm{d}t \right\}^2 \bar{\mu}(\mathrm{d}z) = \\ = 2 \int_0^\tau \left(1 - u/\tau\right) \int \mu(\mathrm{d}z) f(z) f \circ \psi_u(z) \mathrm{d}t''\end{aligned}$$

which is akin to the integration correlation time encountered in MCMC. □

This is somewhat reminiscent of what is advocated in the literature in the context of HMC or randomized HMC where the integration time t (or T when using an integrator) is randomized [29, 21].

Example 27. Consider the Gaussian scenario where $\mathsf{X} = \mathsf{V} = \mathbb{R}^d$, $\pi(x) = \mathcal{N}(x; 0, \Sigma)$ with diagonal covariance matrix such that for $i \in [\![d]\!]$, $\Sigma_{ii} = \sigma_i^2$ and $\varpi(v) = \mathcal{N}(v; 0, \mathrm{Id})$. Using the reparametrization $(x_0(i), v_0(i)) = (a_i \sigma_i \sin(\phi_i), a_i \cos \phi_i)$ the solution of Hamilton's equations is given for $i \in [\![d]\!]$ by

$$\begin{aligned}x_t(i) &= a_i \sigma \sin(t/\sigma_i + \phi_i) \\ v_t(i) &= a_i \cos(t/\sigma_i + \phi_i)\,.\end{aligned}$$

We have for each component $\mathbb{E}_\mu[X_0(i) X_t(i)] = \sigma_i^2 \cos(t/\sigma_i)$, which clearly does not vanish as t, or τ, increase: this is a particularly negative result for standard HMC where the final state of the computed integrator is used. Worse, as noted early on in [29] this implies that for heterogeneous values $\{\sigma_i^2, i \in [\![d]\!]\}$ no integration time may be suitable simultaneously for all coordinates. This motivated the introduction of random integration times [29] which leads to the average correlation

$$\mathbb{E}_\mu \left\{ \frac{1}{\tau} \int_0^\tau X_0(i) X_t(i) \mathrm{d}t \right\} = \frac{\sin(\tau/\sigma_i)}{\tau/\sigma_i}\,,$$

where it is assumed that τ is independent of the initial state. This should be contrasted with the fixed integration time scenario since as τ increases this vanishes and is even negative for some values of τ (the minimum is here reached for about $\tau_i = 4.5\sigma_i$ for a given component).

The example therefore illustrates that our approach implements this averaging feature, and therefore shares its benefits, within the context of an iterative algorithm. The example also highlights a control variate technique intepretation. More specifically in the discrete time scenarios $\{f \circ \psi^k(z), k \in [\![P]\!]\}$ can be interpreted as control variates, but can induce both positive and negative correlations.

4.3.3 Towards an optimal flow?

In this section we are looking to determine a flow for some $\tau > 0$ $[0, \tau] \ni t \mapsto \psi_t$ of an ODE of the form $\dot{z}_t = F \circ z_t$ for some field $F \colon \mathsf{Z} \to \mathsf{Z}$ which defines as above the probability model

$$\bar{\mu}(\mathrm{d}t, \mathrm{d}z) = \frac{1}{\tau} \mu^{\psi_{-t}}(\mathrm{d}z) \mathbf{1}\{0 \leq t \leq \tau\} \mathrm{d}t \,,$$

which has the property that for any $f \colon \mathsf{Z} \to \mathbb{R}$, defining $\bar{f}(t, z) := f \circ \psi_t(z)$, then $\bar{\mu}(\bar{f}) = \mu(f)$. This suggests the use of a Rao-Blackwellized estimator inspired by $\mathbb{E}_{\bar{\mu}}\left(\bar{\mu}(\bar{f} \mid \check{Z})\right)$. Assuming $\mathsf{Z} = \mathbb{R}^d \times \mathbb{R}^d$ and that μ has a density with respect to the Lebesgue measure, then for $z \in \mathsf{Z}$

$$\bar{\mu}(\bar{f} \mid z) = \frac{1}{\tau} \int_0^\tau f \circ \psi_t(z) \frac{\mu \circ \psi_t(z) \, \mathcal{J} \circ \psi_t(z)}{\int \mu \circ \psi_u(z) \, \mathcal{J} \circ \psi_u \mathrm{d}u} \mathrm{d}t \tag{38}$$

In the light of Lemma 26 we aim to find for any $z \in \mathsf{Z}$ the flow solutions $t \mapsto \psi_t(z)$ of ODEs $\dot{z}_t = F_z(z_t)$ such that the function $t \mapsto \int \mu(\mathrm{d}z) f(z) f \circ \psi_t(z) \mu \circ \psi_t(z) \, \mathcal{J} \circ \psi_t(z)$ decreases as fast as possible. This is motivated by the fact that the integral on $[0, \tau]$ of this mapping appears in the variance upper bound for (38) in Lemma 26, which we want to minimize. Note that we also expect this mapping to be smooth under general conditions not detailed in this preliminary work. For smooth enough flow and f we have, with $g := f \times \mu \times \mathcal{J}$,

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}_\mu \left[f(Z) \frac{g \circ \psi_t(Z)}{\bar{\mu}(Z)} \right] = \mathbb{E}_\mu \left[\frac{f(Z)}{\bar{\mu}(Z)} \langle \nabla g \circ \psi_t(Z), \dot{\psi}_t(Z) \rangle \right]$$

Pointwise, the steepest descent direction is given by

$$\dot{\psi}_t(z) = -C_t(z) \frac{f(z) \, \nabla g \circ \psi_t(z)}{|f(z) \, \nabla g \circ \psi_t(z)|}$$

for a positive function $C_t(z)$ to be determined optimally. In this scenario we therefore have

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}_\mu \left[f(Z) \frac{g \circ \psi_t(Z)}{\bar{\mu}(Z)} \right] = -\mathbb{E}_\mu \left[\frac{C_t(z)}{\bar{\mu}(Z)} |f(Z) \, \nabla g \circ \psi_t(Z)| \right]$$

and by Cauchy-Schwartz the (positive) expectation is maximized for

$$C_t(z) = C_t \frac{|f(z) \, \nabla g \circ \psi_t(z)|}{\bar{\mu}(z)} \,.$$

and the trajectories we are interested in must be such that, for some $C > 0$,

$$\dot{\psi}_t(z) = -\frac{C}{\bar{\mu}(z)} f(z) \, \nabla g \circ \psi_t(z) = -\frac{C}{\bar{\mu}(z)} f(z) F \circ \psi_t(z) \,.$$

Note that for any z the term $|f(z)/\bar{\mu}(z)|$ is only a change of speed and that the trajectory of $\mathbb{R}_+ \ni t \mapsto \psi_t(z)$ is independent of this factor, despite the remarkable fact that $\bar{\mu}(z)$ depends on this flow. The result seems fairly natural.

5 MCMC with integrator snippets

We restrict this discussion to integrator snippet based algorithms, but more general Markov snippet algorithms could be considered.

Consider again the target distribution

$$\bar{\mu}(\mathrm{d}z) = \sum_{k=0}^{T} \bar{\mu}(k, \mathrm{d}z)\,,$$

with

$$\bar{\mu}(k, \mathrm{d}z) = \frac{1}{T+1}\mu_k(\mathrm{d}z)\,,$$

where $\mu_k(\mathrm{d}z) := \mu^{\psi^{-k}}(\mathrm{d}z)$ for $k \in [\![0, T]\!]$. Assume that we are in the context of Example 1, dropping n for simplicity, the HMC algorithm using the integrator ψ^s is as follows

$$\bar{P}(z, \mathrm{d}z') = \alpha(z)\delta_{\psi^s(z)}(\mathrm{d}z) + \bar{\alpha}(z)\delta_{\sigma(z)}(\mathrm{d}z') \tag{39}$$

with here

$$\alpha(z) = \min\left\{1, \frac{\mathrm{d}\bar{\mu}^{\sigma\circ\psi^s}}{\mathrm{d}\bar{\mu}}(z)\right\} = \min\left\{1, \frac{\bar{\mu}\circ\psi^w(z)}{\bar{\mu}(z)}\right\}$$

where the last equality holds when $\upsilon \gg \bar{\mu}$, $\upsilon^{\psi^s} = \upsilon$ and we let $\bar{\mu}(z) = \mathrm{d}\bar{\mu}/\mathrm{d}\upsilon(z)$. In other works the snippet $\mathsf{z} = \big(z, \psi(z), \psi^2(z), \ldots, \psi^T(z)\big)$ is shifted along the orbit $\{\psi^k(z), k \in \mathbb{Z}\}$ by ψ^s. Naturally this needs to be combined with updates of the velocity to lead to a viable ergodic MCMC algorithm. This can be achieved with the following kernel, with $\psi^k(z) = \big(\psi_x^k(z), \psi_v^k(z)\big)$ and $A \in \mathscr{Z}$,

$$\bar{Q}(z, A) := \sum_{k,l=0}^{T} \bar{\mu}(k \mid z)\frac{1}{T+1}\int \mathbf{1}\{\psi^{-l}\big(\psi_x^k(z), v'\big) \in A\}\varpi(\mathrm{d}v')\,,$$

described algorithmically in Alg. 6. Indeed for any $(l, v') \in [\![0, T]\!] \times \mathsf{V}$, using identity (11),

$$\int \sum_{k=0}^{T} \bar{\mu}(k \mid z) \int \mathbf{1}\{\psi^{-l}\big(\psi_x^k(z), v'\big) \in A\}\bar{\mu}(\mathrm{d}z) = \int \sum_{k=0}^{T} \int \mathbf{1}\{\psi^{-l}\big(x, v'\big) \in A\}\mu(\mathrm{d}z)\,,$$

and hence

$$\begin{aligned}
\bar{\mu}\bar{Q}(A) &= \frac{1}{T+1}\sum_{l=0}^{T}\int \mathbf{1}\{\psi^{-l}\big(x, v'\big) \in A\}\varpi(\mathrm{d}v')\mu(\mathrm{d}z) \\
&= \frac{1}{T+1}\sum_{l=0}^{T}\int \mathbf{1}\{\psi^{-l}\big(x, v\big) \in A\}\mu(\mathrm{d}z) \\
&= \int \mathbf{1}\{z \in A\}\frac{1}{T+1}\sum_{l=0}^{T}\mu^{\psi^{-l}}(\mathrm{d}z) \\
&= \bar{\mu}(A)\,.
\end{aligned}$$

Again samples from this MCMC targetting $\bar{\mu}$ can be used to estimate expectations with respect to μ using the identity (11). This is closely related to the "windows of states" approach of [28, 29, 31], where a window of states is

what we call a Hamiltonian snippet in the present manuscript. Indeed the windows of states approach corresponds to the Markov update

$$P(z, A) = \sum_{k,l=0}^{T} \frac{1}{T} \int \bar{P}\big(\psi^{-k}(z), \mathrm{d}z'\big) \bar{\mu}(l \mid z') \mathbf{1}\{\psi^l(z') \in A\}\,, \tag{40}$$

which, we show below, leaves μ invariant. Indeed, note that for $k, l \in [\![0, T]\!]$ and $A \in \mathscr{Z}$,

$$\int \mu(\mathrm{d}z) \int \bar{P}\big(\psi^{-k}(z), \mathrm{d}z'\big) \bar{\mu}(l \mid z') \mathbf{1}\{\psi^l(z') \in A\} = \int_A \mu^{\psi^{-k}}(\mathrm{d}z) \int \bar{P}\big(z, \mathrm{d}z'\big) \bar{\mu}(l \mid z') \mathbf{1}\{\psi^l(z') \in A\}\,,$$

and therefore

$$\begin{aligned}
\int \mu(\mathrm{d}z) \int P(z, \mathrm{d}z') \mathbf{1}\{z' \in A\} &= \int \bar{\mu}(\mathrm{d}z) \bar{P}(z, \mathrm{d}z') \sum_{l=0}^{T} \bar{\mu}(l \mid z') \mathbf{1}\{\psi^l(z') \in A\} \\
&= \int \bar{\mu}(\mathrm{d}z') \sum_{l=0}^{T} \bar{\mu}(l \mid z') \mathbf{1}\{\psi^l(z') \in A\} \\
&= \int \mu(\mathrm{d}z) \mathbf{1}\{z \in A\}\,,
\end{aligned}$$

where we obtain the last line from (11). P is not μ-reversible in general, making theoretical comparisons challenging.

1 Given z
2 Sample $k \sim \bar{\mu}(\cdot \mid z)$ and $l \sim \mathcal{U}([\![0, T]\!])$.
3 Compute $\psi^k(z) = \big(\psi_x^k(z), \psi_v^k(z)\big)$.
4 Sample $v' \sim \varpi(\cdot)$
5 Return $\psi^{-l}(\psi_x^k(z), v')$

Algorithm 6: Kernel $\bar{Q}$

6 Discussion

We have shown how mappings used in various Monte Carlo schemes relying on numerical integrators of ODE can be implemented to fully exploit all computations to design robust and efficient sampling algorithms. Numerous questions remain open, including the tradeoff between N and T. A precise analysis of this question is made particularly difficult by the fact that integration along snippets is straightforwardly parallelizable, while resampling does not lend itself to straightforward parallelisation.

Another point is concerned with the particular choice of mutation Markov kernel $\bar{M}_n$, or $\bar{M}$, in (12) or (23). Indeed such a kernel starts with a transition from samples approximating the snippet distribution $\bar{\mu}_{n-1}$ to μ_{n-1}, which is then followed by a reweighting of samples leading to a representation of $\bar{\mu}_n$. Instead, for illustration, one could suggest using an SMC sampler with (39) as mutation kernel.

In relation to the discussion in Remark 20, a natural question is how our scheme would compare with a "Rao-Blackwellized" SMC where weights of the type (35), derived from (6) are used.

We leave all these questions for future investigations.

Acknowledgements

The authors would like to thank Carl Dettman for very useful discussions on Boltzman's conjecture. Research of CA and MCE supported by EPSRC grant 'CoSInES (COmputational Statistical INference for Engineering and Security)' (EP/R034710/1), and EPSRC grant Bayes4Health, 'New Approaches to Bayesian Data Science: Tackling Challenges from the Health Sciences' (EP/R018561/1). Research of CZ was supported by a CSC Scholarship.

A Notation and definitions

We will write $\mathbb{N} = \{0, 1, 2, \dots\}$ for the set of natural numbers and $\mathbb{R}_+ = (0, \infty)$ for positive real numbers. Throughout this section $(\mathsf{E}, \mathscr{E})$ is a generic measurable space.

- For $A \subset \mathsf{E}$ we let $A^{\complement}$ be its complement.
- $\mathcal{M}(\mathsf{E}, \mathscr{E})$ (resp. $\mathcal{P}(\mathsf{E}, \mathscr{E})$ is the set of measures (resp. probability distributions) on $(\mathsf{E}, \mathscr{E})$
- For a set $A \in \mathscr{E}$, its complement in E is denoted by $A^{\complement}$. We denote the corresponding indicator function by $\mathbf{1}_A : \mathsf{E} \to \{0, 1\}$ and may use the notation $\mathbf{1}\{z \in A\} := \mathbf{1}_A(z)$.
- For μ a probability measure on $(\mathsf{E}, \mathscr{E})$ and a measurable function $f \colon \mathsf{E} \to \mathbb{R}$ and , we let $\mu(f) := \int f(x)\mu(\mathrm{d}x)$.
- For two probability measures μ and ν on $(\mathsf{E}, \mathscr{E})$ we let $\mu \otimes \nu$ be a measure on $(\mathsf{E} \times \mathsf{E}, \mathscr{E} \otimes \mathscr{E})$ such that $\mu \otimes \nu(A \times B) = \mu(A)\nu(B)$ for $A, B \in \mathscr{E}$.
- For a Markov kernel $P(x, \mathrm{d}y)$ on $\mathsf{E} \times \mathscr{E}$, we write
  - $\mu \otimes P$ for the probability measure on $(\mathsf{E} \times \mathsf{E}, \mathscr{E} \otimes \mathscr{E})$ such that for $\bar{A} \in \mathscr{E} \otimes \mathscr{E}$, the minimal product σ-algebra, $\mu \otimes P(\bar{A}) = \int_{\bar{A}} \mu(\mathrm{d}x)P(x, \mathrm{d}y)$.
  - $\mu \hat{\otimes} P$ for the probability measure on $(\mathsf{E} \times \mathsf{E}, \mathscr{E} \otimes \mathscr{E})$ such that for $A, B \in \mathscr{E}$ $\mu \hat{\otimes} P(A \times B) = \mu \otimes P(B \times A)$.
- For μ, ν probability distributions on $(\mathsf{E}, \mathscr{E})$ and kernels $M, L \colon \mathsf{E} \times \mathscr{E} \to [0, 1]$ such that $\mu \otimes M \gg \nu \hat{\otimes} L$ then we denote
$$\frac{\mathrm{d}\nu \hat{\otimes} L}{\mathrm{d}\mu \otimes M}(z, z')$$
the corresponding Radon-Nikodym derivative such that for $f \colon \mathsf{E} \times \mathsf{E} \to \mathbb{R}$,
$$\int f(z, z') \frac{\mathrm{d}\nu \hat{\otimes} L}{\mathrm{d}\mu \otimes M}(z, z')\mu \otimes M\big(\mathrm{d}(z, z')\big) = \int f(z, z')\nu \hat{\otimes} L\big(\mathrm{d}(z', z)\big)\,.$$
- A point mass distribution at x will be denoted by $\delta_x(\mathrm{d}y)$; it is such that for $f \colon \mathsf{E} \to \mathbb{R}$
$$\int f(x)\delta_x(\mathrm{d}y) = f(x)$$
- In order to alleviate notation, for $M \in \mathbb{N}$, $(z^{(i)}, w_i) \in \mathsf{E} \times [0, \infty)$, $i \in [\![M]\!]$, we refer to $\big\{(z^{(i)}, w_i), i \in [\![M]\!]\big\}$ as weighted samples to mean $\big\{(z^{(i)}, \tilde{w}_i), i \in [\![M]\!]\big\}$ where $\tilde{w}_i \propto w_i$ but $\sum_{i=1}^M \tilde{w}_i = 1$.

- We say that a set of weighted samples, or particles, $\{(z_i, w_i) \in \mathsf{Z} \times \mathbb{R}_+ : i \in [\![N]\!]\}$ for $N \geq 1$ represents a distribution μ whenever for f μ-integrable

$$\sum_{i=1}^{N} \frac{w_i}{\sum_{j=1}^{N} w_j} f(z_i) \approx \mu(f)\,,$$

  in either in the L^p sense for some $p \geq 1$.

- For $M \in \mathbb{N}$, $w_i \in [0, \infty)$, $i \in [\![M]\!]$, we let $K \sim \mathrm{Cat}\,(w_1, w_2, \ldots, w_M)$ mean that $\mathbb{P}(K = k) \propto w_k$.
- For $M, N \in \mathbb{N}$, $w_{ij} \in [0, \infty)$, $i \in [\![M]\!] \times [\![N]\!]$, we let $K \sim \mathrm{Cat}\,(w_{ij}, i \in [\![M]\!] \times [\![N]\!])$ mean that $\mathbb{P}(K = (k, l)) \propto w_{kl}$.
- for $f : \mathbb{R}^m \to \mathbb{R}^n$ we let
  - $\nabla \otimes f$ be the transpose of the Jacobian
  - for $n = 1$ we let $\nabla f = (\nabla \otimes f)^\top$ be the gradient,
  - $\nabla \cdot f$ be the divergence.

B Radon-Nikodym derivative

The general formalism required and used throughout the paper relies on a unique measure theoretic tool, the Radon-Nikodym derivative. We gather here definitions and intermediate results used throughout, pointing out the simplicity of the tools involved and and the benefits they bring.

Definition 28 (Pushforward)**.** Let μ be a measure on $(\mathsf{Z}, \mathscr{Z})$ and $\psi : (\mathsf{Z}, \mathscr{Z}) \to (\mathsf{Z}', \mathscr{F}')$ a measurable function. The pushforward of μ by ψ is defined by

$$\mu^{\psi}(A) := \mu(\psi^{-1}(A)), \qquad A \in \mathscr{Z}',$$

where $\psi^{-1}(A) = \{z \in \mathsf{Z} : \psi(z) \in A\}$ is the preimage of A under ψ.

If μ is a probability distribution then μ^{ψ} is the probability measure associated with $\psi(Z)$ when $Z \sim \mu$.

Definition 29 (Dominating and equivalent measures)**.** For two measures μ and ν on the same measurable space $(\mathsf{Z}, \mathscr{Z})$,

1. ν is said to dominate μ if for all measurable $A \in \mathscr{Z}$, $\mu(A) > 0 \Rightarrow \nu(A) > 0$ – this is denoted $\nu \gg \mu$.
2. μ and ν are equivalent, written $\mu \equiv \nu$, if $\mu \gg \nu$ and $\nu \gg \mu$.

We will need the notion of Radon-Nikodym derivative [7, Theorems 32.2 & 16.11]:

Theorem 30 (Radon–Nikodym)**.** *Let μ and ν be σ-finite measures on $(\mathsf{Z}, \mathscr{Z})$. Then $\nu \ll \mu$ if and only if there exists an essentially unique, measurable, non-negative function $f : \mathsf{Z} \to [0, \infty)$ such that*

$$\int_A f(z)\mu(\mathrm{d}z) = \nu(A), \qquad A \in \mathscr{E}.$$

Therefore we can view $\mathrm{d}\nu/\mathrm{d}\mu := f$ as the density of ν w.r.t μ and in particular if g is integrable w.r.t. ν then

$$\int g(z)\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(z)\mu(\mathrm{d}z) = \int g(z)\nu(\mathrm{d}z)\,.$$

If μ is a measure and f a non-negative, measurable function then $\mu \cdot f$ is the measure $(\mu \cdot f)(A) = \int \mathbf{1}_A(z) f(z) \mu(\mathrm{d}z)$, i.e. the measure $\nu = \mu \cdot f$ such that f is the Radon–Nikodym derivative $\mathrm{d}\nu/\mathrm{d}\mu = f$.

The following establishes the expression of an expectation with respect to the pushforward μ^ψ in terms of expectations with respect to μ [7, Theorem 16.13].

Theorem 31 (Change of variables)**.** *A function $f : \mathsf{Z}' \to \mathbb{R}$ is integrable w.r.t. μ^ψ if and only if $f \circ \psi$ is integrable w.r.t. μ, in which case*

$$\int_{\mathsf{Z}'} f(z) \mu^\psi(\mathrm{d}z) = \int_{\mathsf{Z}} f \circ \psi(z) \mu(\mathrm{d}z)\,. \tag{41}$$

We now establish results useful throughout the manuscript. The central identity used throughout the manuscript is a direct application of Theorem 31 for $\psi \colon \mathsf{Z} \to \mathsf{Z}$ invertible

$$\int_{\mathsf{Z}'} f \circ \psi(z) \mu^{\psi^{-1}}(\mathrm{d}z) = \int_{\mathsf{Z}} f(z) \mu(\mathrm{d}z)$$

which seems tautological since it can be summarized as follows: for $Z \sim \mu$, then $\psi^{-1}(Z) \sim \mu^{\psi^{-1}}$ and $\psi \circ \psi^{-1}(Z) \sim \mu$! However the interest of the approach stems from the following properties.

Lemma 32. *Let $\psi \colon \mathsf{Z} \to \mathsf{Z}$ be measurable and integrable, μ and υ be σ-finite measures on $(\mathsf{Z}, \mathscr{Z})$ such that $\upsilon \gg \mu$ and $\upsilon \gg \upsilon^{\psi^{-1}}$. Then*

1. *$\upsilon^{\psi^{-1}} \gg \mu^{\psi^{-1}}$ and therefore $\upsilon \gg \mu^{\psi^{-1}}$,*
2. *for υ-almost all $z \in \mathsf{Z}$,*
$$\frac{\mathrm{d}\mu^{\psi^{-1}}}{\mathrm{d}\upsilon}(z) = \frac{\mathrm{d}\mu}{\mathrm{d}\upsilon} \circ \psi(z) \frac{\mathrm{d}\upsilon^{\psi^{-1}}}{\mathrm{d}\upsilon}$$
3. *we have*
$$\mu \gg \mu^{\psi^{-1}} \iff \upsilon\big(\big\{z \in \mathsf{Z} \colon \mathrm{d}\mu^{\psi^{-1}}/\mathrm{d}\upsilon(z) > 0, \mathrm{d}\mu/\mathrm{d}\upsilon(z) = 0\big\}\big) = 0$$
in which case for υ-almost all $z \in \mathsf{Z}$
$$\frac{\mathrm{d}\mu^{\psi^{-1}}}{\mathrm{d}\mu}(z) = \begin{cases} \frac{\mu \circ \psi}{\mu}(z) & \mu(z) > 0 \\ 0 & \textit{otherwise} \end{cases}$$
and therefore
$$\int_{\mathsf{Z}'} f \circ \psi(z) \mu^{\psi^{-1}}(\mathrm{d}z) = \int_{\mathsf{Z}'} f \circ \psi(z) \frac{\mu \circ \psi(z)}{\mu(z)} \frac{\mathrm{d}\upsilon^\psi}{\mathrm{d}\upsilon} \mu(\mathrm{d}z).$$

Proof. For the first part of the first statement, let $A \in \mathscr{Z}$ such that $\upsilon^{\psi^{-1}}(A) = \upsilon\big(\psi(A)\big) > 0$, then since $\psi(A) \in \mathscr{Z}$ and $\upsilon \gg \mu$ we deduce $\mu\big(\psi(A)\big) = \mu^{\psi^{-1}}(A) > 0$ and we conclude; the second parts follows from $\upsilon \gg \upsilon^{\psi^{-1}}$. For the

second statement for $f\colon \mathsf{Z} \to \mathbb{R}$ bounded and measurable,

$$\begin{aligned}\int f(z)\frac{\mathrm{d}\mu^{\psi^{-1}}}{\mathrm{d}\upsilon}(z)\upsilon(\mathrm{d}z) &= \int f(z)\mu^{\psi^{-1}}(\mathrm{d}z)\\ &= \int f\circ\psi^{-1}(z)\mu(\mathrm{d}z)\\ &= \int f\circ\psi^{-1}(z)\frac{\mathrm{d}\mu}{\mathrm{d}\upsilon}(z)\upsilon(\mathrm{d}z)\\ &= \int f(z)\frac{\mathrm{d}\mu}{\mathrm{d}\upsilon}\circ\psi(z)\upsilon^{\psi^{-1}}(\mathrm{d}z)\\ &= \int f(z)\frac{\mathrm{d}\mu}{\mathrm{d}\upsilon}\circ\psi(z)\frac{\mathrm{d}\upsilon^{\psi^{-1}}}{\mathrm{d}\upsilon}(z)\upsilon(\mathrm{d}z)\end{aligned}$$

The third statement is given as [7, Problem 32.6.], which we solve in Lemma 34. □

Corollary 33. *In the scenario when $\psi\colon \mathsf{Z}\to\mathsf{Z}$ and υ are such that $\upsilon^{\psi}=\upsilon$ then $\mathrm{d}\upsilon^{\psi}/\mathrm{d}\upsilon \equiv 1$.*

Lemma 34 (Billingsley, Problem 32.6.)**.** *Assume μ,ν and υ are σ- finite and that $\upsilon \gg \nu,\mu$. Then $\mu \gg \nu$ if and only if $\upsilon\big(\{z\in\mathsf{Z}\colon \mathrm{d}\nu/\mathrm{d}\upsilon(z)>0, \mathrm{d}\mu/\mathrm{d}\upsilon(z)=0\}\big)=0$, in which case*

$$\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(z) = \mathbf{1}\{z\in\mathsf{Z}: \mathrm{d}\mu/\mathrm{d}\upsilon(z)>0\}\frac{\mathrm{d}\nu}{\mathrm{d}\upsilon}/\frac{\mathrm{d}\mu}{\mathrm{d}\upsilon}(z)\,.$$

Proof. Let $S := \big\{z\in\mathsf{Z}\colon \mathrm{d}\nu/\mathrm{d}\upsilon(z)>0, \mathrm{d}\mu/\mathrm{d}\upsilon(z)=0\big\}$. For $f\colon \mathsf{Z}\to\mathbb{R}$ integrable we always have

$$\int f(z)\nu(\mathrm{d}z) = \int \mathbf{1}\{z\in S\}f(z)\frac{\mathrm{d}\nu}{\mathrm{d}\upsilon}(z)\upsilon(\mathrm{d}z) + \int \mathbf{1}\big\{z\in S^{\complement}\big\}f(z)\frac{\mathrm{d}\nu}{\mathrm{d}\upsilon}/\frac{\mathrm{d}\mu}{\mathrm{d}\upsilon}(z)\mu(\mathrm{d}z)\,.$$

Assume $\mu\gg\nu$ then from above for any $f\colon \mathsf{Z}\to\mathbb{R}$ integrable

$$\begin{aligned}\int \mathbf{1}\{z\in S\}f(z)\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(z)\mu(\mathrm{d}z) &= \int \mathbf{1}\{z\in S\}f(z)\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(z)\frac{\mathrm{d}\mu}{\mathrm{d}\upsilon}(z)\upsilon(\mathrm{d}z)\\ &= 0\,,\end{aligned}$$

and therefore $\upsilon\big(S\big)=0$ and we conclude from the first identity above. Now assume that $\upsilon\big(S\big)=0$, then

$$\begin{aligned}\int f(z)\nu(\mathrm{d}z) &= \int \mathbf{1}\{z\in S\}f(z)\frac{\mathrm{d}\nu}{\mathrm{d}\upsilon}(z)\upsilon(\mathrm{d}z) + \int \mathbf{1}\big\{z\in S^{\complement}\big\}f(z)\frac{\mathrm{d}\nu}{\mathrm{d}\upsilon}/\frac{\mathrm{d}\mu}{\mathrm{d}\upsilon}(z)\mu(\mathrm{d}z)\\ &= \int \mathbf{1}\big\{z\in S^{\complement}\big\}f(z)\frac{\mathrm{d}\nu}{\mathrm{d}\upsilon}/\frac{\mathrm{d}\mu}{\mathrm{d}\upsilon}(z)\mu(\mathrm{d}z)\,.\end{aligned}$$

The equivalence is therefore established and when either conditions is satisfied we have

$$\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(z) = \mathbf{1}\big\{z\in S^{\complement}\big\}\frac{\mathrm{d}\nu}{\mathrm{d}\upsilon}/\frac{\mathrm{d}\mu}{\mathrm{d}\upsilon}(z)$$

and we conclude. □

References

[1] Christophe Andrieu, Anthony Lee, and Sam Livingstone. A general perspective on the Metropolis-Hastings kernel. *arXiv preprint arXiv:2012.14881*, 2020.

[2] Christophe Andrieu, James Ridgway, and Nick Whiteley. Sampling normalizing constants in high dimensions using inhomogeneous diffusions, 2018.

[3] Mark A Beaumont. Approximate bayesian computation in evolution and ecology. *Annual review of ecology, evolution, and systematics*, 41:379–406, 2010.

[4] Alexandros Beskos, Natesh Pillai, Gareth Roberts, Jesus-Maria Sanz-Serna, and Andrew Stuart. Optimal tuning of the hybrid Monte Carlo algorithm. *Bernoulli*, 19(5A):1501–1534, 2013.

[5] Michael Betancourt. Nested sampling with constrained Hamiltonian Monte Carlo. In *AIP Conference Proceedings*, volume 1305, pages 165–172. American Institute of Physics, 2011.

[6] Joris Bierkens and Gareth Roberts. A piecewise deterministic scaling limit of lifted Metropolis–Hastings in the Curie–Weiss model. *The Annals of Applied Probability*, 27(2):846–882, 2017.

[7] P Billingsley. Probability and measure. 3rd wiley. *New York*, 1995.

[8] Alexandre Bouchard-Côté, Sebastian J Vollmer, and Arnaud Doucet. The bouncy particle sampler: A nonreversible rejection-free Markov chain monte carlo method. *Journal of the American Statistical Association*, 113(522):855–867, 2018.

[9] Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A course in metric geometry*, volume 33. American Mathematical Society, 2022.

[10] Mari Paz Calvo, Daniel Sanz-Alonso, and Jesús María Sanz-Serna. HMC: reducing the number of rejections by not using leapfrog and some results on the acceptance rate. *Journal of Computational Physics*, 437:110333, 2021.

[11] Cédric M Campos and Jesús María Sanz-Serna. Extra chance generalized hybrid Monte Carlo. *Journal of Computational Physics*, 281:365–374, 2015.

[12] Joseph T Chang and David Pollard. Conditioning as disintegration. *Statistica Neerlandica*, 51(3):287–317, 1997.

[13] Nicolas Chopin and James Ridgway. Leave pima indians alone: binary regression as a benchmark for Bayesian computation. *Statistical Science*, 32(1):64–87, 2017.

[14] Hai-Dang Dau and Nicolas Chopin. Waste-free sequential Monte Carlo. *arXiv preprint arXiv:2011.02328*, 2020.

[15] Pierre Del Moral and Pierre Del Moral. *Feynman-kac formulae*. Springer, 2004.

[16] Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436, 2006.

[17] Simon Duane, Anthony D Kennedy, Brian J Pendleton, and Duncan Roweth. Hybrid Monte Carlo. *Physics letters B*, 195(2):216–222, 1987.

[18] Mauro Camara Escudero. *Approximate Manifold Sampling: Robust Bayesian Inference for Machine Learning*. PhD thesis, School of Mathematics, January 2024.

[19] Youhan Fang, J. M. Sanz-Serna, and Robert D. Skeel. Compressible generalized hybrid Monte Carlo. *The Journal of Chemical Physics*, 140(17):174108, 05 2014.

[20] Ernst Hairer, SP Nörsett, and G. Wanner. *Solving ordinary differential equations I, Nonstiff problems*. Springer-Verlag, 1993.

[21] Matthew Hoffman, Alexey Radul, and Pavel Sountsov. An adaptive-MCMC scheme for setting trajectory lengths in Hamiltonian Monte Carlo. In *International Conference on Artificial Intelligence and Statistics*, pages 3907–3915. PMLR, 2021.

[22] Matthew D Hoffman and Pavel Sountsov. Tuning-Free Generalized Hamiltonian Monte Carlo. In *International Conference on Artificial Intelligence and Statistics*, pages 7799–7813. PMLR, 2022.

[23] Paul B Mackenze. An improved hybrid Monte Carlo method. *Physics Letters B*, 226(3-4):369–371, 1989.

[24] Florian Maire and Pierre Vandekerkhove. On markov chain monte carlo for sparse and filamentary distributions. *ArXiv e-prints*, 2018.

[25] Florian Maire and Pierre Vandekerkhove. Markov kernels local aggregation for noise vanishing distribution sampling. *SIAM Journal on Mathematics of Data Science*, 4(4):1293–1319, 2022.

[26] Youssef Marzouk, Tarek Moselhy, Matthew Parno, and Alessio Spantini. *Sampling via Measure Transport: An Introduction*, pages 1–41. Springer International Publishing, Cham, 2016.

[27] Xiao-Li Meng and Stephen Schilling. Warp bridge sampling. *Journal of Computational and Graphical Statistics*, 11(3):552–586, 2002.

[28] Radford M Neal. An improved acceptance procedure for the hybrid Monte Carlo algorithm. *Journal of Computational Physics*, 111(1):194–203, 1994.

[29] Radford M. Neal. *MCMC Using Hamiltonian Dynamics*, chapter 5. CRC Press, 2011.

[30] Ari Pakman and Liam Paninski. Exact Hamiltonian Monte Carlo for truncated multivariate Gaussians. *Journal of Computational and Graphical Statistics*, 23(2):518–542, 2014.

[31] Zhaohui S. Qin and Jun S. Liu. Multipoint Metropolis method with application to Hybrid Monte Carlo. *Journal of Computational Physics*, 172(2):827–840, 2001.

[32] James Ridgway. Computation of gaussian orthant probabilities in high dimension. *Statistics and computing*, 26(4):899–916, 2016.

[33] Grant M Rotskoff and Eric Vanden-Eijnden. Dynamical computation of the density of states and Bayes factors using nonequilibrium importance sampling. *Physical review letters*, 122(15):150602, 2019.

[34] Gabriel Stoltz, Mathias Rousset, et al. *Free energy computations: A mathematical perspective*. World Scientific, 2010.

[35] D Szasz. *Boltzmann's ergodic hypothesis, a conjecture for centuries?*, chapter Hard ball systems and the Lorentz gas, pages 421–446. Springer, 2000.

[36] Esteban G Tabak and Eric Vanden-Eijnden. Density estimation by dual ascent of the log-likelihood. *Communications in Mathematical Sciences*, 8(1):217–233, 2010.

[37] Erik H Thiede, Brian Van Koten, Jonathan Weare, and Aaron R Dinner. Eigenvector method for umbrella sampling enables error analysis. *The Journal of chemical physics*, 145(8), 2016.

[38] Achille Thin, Yazid Janati El Idrissi, Sylvain Le Corff, Charles Ollion, Eric Moulines, Arnaud Doucet, Alain Durmus, and Christian X Robert. Neo: Non equilibrium sampling on the orbits of a deterministic transform. *Advances in Neural Information Processing Systems*, 34:17060–17071, 2021.

[39] G.M. Torrie and J.P. Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187–199, 1977.

[40] Paul F Tupper. Ergodicity and the numerical simulation of Hamiltonian systems. *SIAM Journal on Applied Dynamical Systems*, 4(3):563–587, 2005.

[41] J. von Neumann. Proof of the quasi-ergodic hypothesis. *Proc. Natl. Acad. Sci. USA*, 18:70–82, 1932.

[42] Chang Zhang. *On the Improvements and Innovations of Monte Carlo Methods*. PhD thesis, School of Mathematics, https://research-information.bris.ac.uk/en/studentTheses/on-the-improvements-and-innovations-of-monte-carlo-methods, June 2022.