Author(s): Chopin, Nicolas; Singh, Sumeetpal S.; Soto, Tomás; Vihola, Matti

Title: On resampling schemes for particle filters with weakly informative observations

Year: 2022

Version: Accepted version (Final draft)

Please cite the original version:

ON RESAMPLING SCHEMES FOR PARTICLE FILTERS WITH WEAKLY INFORMATIVE OBSERVATIONS

BY NICOLAS CHOPIN[1,a] SUMEETPAL S. SINGH[2,b] TOMÁS SOTO[3,c] MATTI VIHOLA[4,d]

[1]*ENSAE, Institut Polytechnique de Paris,* [a]*nicolas.chopin@ensae.fr*

[2]*Department of Engineering, University of Cambridge,* [b]*sss40@eng.cam.ac.uk*

[3]*School of Engineering Science, LUT University,* [c]*Tomas.Soto@lut.fi*

[4]*Department of Mathematics and Statistics, University of Jyväskylä,* [d]*matti.s.vihola@jyu.fi*

We consider particle filters with weakly informative observations (or 'potentials') relative to the latent state dynamics. The particular focus of this work is on particle filters to approximate time-discretisations of continuous-time Feynman–Kac path integral models — a scenario that naturally arises when addressing filtering and smoothing problems in continuous time — but our findings are indicative about weakly informative settings beyond this context too. We study the performance of different resampling schemes, such as systematic resampling, SSP (Srinivasan sampling process) and stratified resampling, as the time-discretisation becomes finer and also identify their continuous-time limit, which is expressed as a suitably defined 'infinitesimal generator.' By contrasting these generators, we find that (certain modifications of) systematic and SSP resampling 'dominate' stratified and independent 'killing' resampling in terms of their limiting overall resampling rate. The reduced intensity of resampling manifests itself in lower variance in our numerical experiment. This efficiency result, through an ordering of the resampling rate, is new to the literature. The second major contribution of this work concerns the analysis of the limiting behaviour of the entire population of particles of the particle filter as the time discretisation becomes finer. We provide the first proof, under general conditions, that the particle approximation of the discretised continuous-time Feynman–Kac path integral models converges to a (uniformly weighted) continuous-time particle system.

1. Introduction. Particle filters [18] have become a workhorse of non-linear stochastic filtering and statistical state space modelling. The heart of the particle filters is the 'interaction' within the particles, which is caused by the resampling (or selection) step of the algorithm. See e.g. [5] for a general introduction to particle filtering and its applications in various scientific fields.

When the observations (or more generally the potential functions) are informative, that is, the weights that go into the resampling have a high variability, empirical evidence suggests that the choice of the resampling strategy can only make a small difference. In contrast, when the observations (potential functions) are weakly informative and the weights tend to be close to uniform, the performance differences between different resampling methods can be substantial. In the weakly informative regime, it is therefore important to use an appropriate resampling scheme.

Different resampling schemes have been suggested in the literature, and some resampling schemes have been compared in terms of the conditional variance that they introduce to the weights [13]. More recently, [16] considered ordering of resampling schemes with respect to a so-called negative association, and introduced a new 'SSP' resampling scheme, which

is preferable based on their theoretical findings. These analyses, as most other theoretical analyses on resampling methods, focus on the asymptotic regime in the number of particles $N \to \infty$.

We do not consider the asymptotic $N \to \infty$, but keep N fixed instead. In contrast, we consider the asymptotic behaviour of the resampling schemes as the potentials become less and less informative. One domain of applications, where such a situation naturally arises, is time-discretisations of continuous-time particle systems [11, 12] that approximate so-called Feynman–Kac path integrals. We study the behaviour of discrete-time resampling schemes when applied with discretisations of continuous-time Feynman–Kac path integral models.

The main contributions of this work are:

- We introduce a condition for (discrete-time) resampling, namely Assumption 2, which ensures the existence of a limiting continuous-time particle system. In particular, when this assumption is satisfied and under further general conditions, Theorem 19 establishes the limiting continuous-time particle system that a particle implementation of the time-discretised Feynman–Kac path integral model converges to (as the discretisation is refined). Having established this asymptotic limit, Theorem 22 gives a result on its use for unbiased estimation of certain expectations with respect to the continuous-time Feynman–Kac path integral.
- In Section 4 we then proceed to analyse which resampling schemes satisfy Assumption 2 in a series of results, namely Propositions 4, 9, 10 and 13. When the condition holds, resampling schemes can be ordered by comparing their limiting continuous-time resampling intensities in Theorem 15 and Proposition 17. We find that certain variants of systematic resampling and SSP resampling have a common limiting overall resampling intensity, which is guaranteed to be lower than that of the stratified or so-called killing resampling. This suggests that our variants of systematic/SSP resampling can be preferable.
- Our empirical findings (Section 7) about the practical performance are in line with our theoretical results, and indicate that SSP resampling and systematic resampling, when applied after a prior partial ordering of the weights about their mean, lead to the best performing particle filters. This complements the positive findings for SSP [16], and suggests that a partial ordering of weights should always be used with systematic resampling.

Overall our results fill important gaps in the literature on particle filters, in particular concerning their continuous-time limiting behaviour. This is in contrast with e.g. [2, 10, 25], who considered directly particular (theoretical) continuous-time algorithms (based on killing resampling), and how they may be discretised.

We consider only resampling schemes that result in a uniformly weighted sample, so that the particle system remains unweighted. Furthermore, all the studied resampling schemes lead to 'single-event' continuous-time limits, meaning that at most one particle can disappear at an individual time. These exclude the popular alternative strategy, adaptive resampling [22], in which resampling is triggered at certain (random) times.

2. Hidden Markov models and particle filters. A hidden Markov model (HMM) consists of two components: a latent (unobserved) Markov chain $X_{1:T} = (X_1, \ldots, X_T)$ on state a space X with an initial probability density $f_1(x_1)$ and transition densities $f_k(x_k \mid x_{k-1})$; and conditionally independent observations $Y_{1:T}$ with conditional laws $g_k(y_k \mid x_k)$. The particle filter can be used to estimate integrals with respect to a conditional probability law, the so-called smoothing distribution:

$$p(x_{1:T} \mid y_{1:T}) \propto p(x_{1:T}, y_{1:T}) = f_1(x_1)g_1(y_1 \mid x_1) \prod_{k=2}^{T} f_k(x_k \mid x_{k-1})g_k(y_k \mid x_k).$$

The Feynman–Kac model is an abstraction and generalisation which allows for defining a family of unnormalised probability densities $\gamma(x_{1:T})$ which is equivalent to $p(x_{1:T}, y_{1:T})$ in the HMM context. It is based on 'proposal' Markov chain laws $M_1(x_1)$ and $M_k(x_k \mid x_{k-1})$ [8] and non-negative 'potential functions' $G_k(x_{1:k})$, where $x_{1:k} \in \mathsf{X}^k \to G_k(x_{1:k})$ (which can implicitly depend on $y_{1:T}$ too), so that $p(x_{1:T} \mid y_{1:T}) = \pi_T(x_{1:T})$, with π_T defined as follows:

(1) $\quad \pi_k(x_{1:k}) := \dfrac{\gamma_k(x_{1:k})}{\mathcal{Z}_k}$ where $\gamma_k(x_{1:k}) := M_1(x_1)G_1(x_1)\displaystyle\prod_{j=2}^{k} M_j(x_j \mid x_{j-1})G_j(x_{1:j})$

and $\mathcal{Z}_k := \int \gamma_k(x_{1:k})\mathrm{d}x_{1:k}$. In the HMM context, we typically set $G_1(x_1) = g_1(y_1 \mid x_1)f_1(x_1)/M_1(x_1)$ and

$$G_k(x_{1:k}) = g_k(y_k \mid x_k)f_k(x_k \mid x_{k-1})/M_k(x_k \mid x_{k-1}).$$

In the simplest case, when $M_k \equiv f_k$, we get $G_k(x_{1:k}) = g_k(y_k \mid x_k)$. Henceforth, the domain of the potential G_k will be apparent from its argument. So when $G_k : \mathsf{X} \to [0, \infty)$, an instance of which is the just mentioned simplest case, we will write $G_k(x_k)$.

The focus of this paper is in situations where G_k are 'weakly informative,' that is, when $G_k(x_k)$ is nearly constant for typical values x_k (with respect to π_k). In the HMM setting, this typically occurs when the observations $Y_k \sim g_k(\,\cdot\, \mid x_k)$ have substantial variability compared to the variability of $X_k \sim f_k(\,\cdot\, \mid x_{k-1})$, and can also occur when M_k correspond to an approximation of the smoothing distribution [cf. 27]. However, our main theoretical framework is beyond the HMM context, where $(M_{1:T}, G_{1:T})$ correspond to time-discretisations of a continuous-time path integral model (Section 3).

Hereafter, we will focus on the Feynman–Kac model in (1), assuming that π_T is well-defined, that is, the normalising constant $\mathcal{Z}_T$ is finite and strictly positive. We use the notation $a{:}b = (a, a+1, \ldots, b)$ for integers $a \leq b$, and use the same notation for indexing and double indexing of sequences. Thus for sequences $\{x_i\}_i$, $\{y^j\}_j$ and $\{z_i^j\}_{i,j}$ we write $x_{a:b} := (x_a, \ldots, x_b)$, $y^{a:b} := (y^a, \ldots, y^b)$ and $z_{a:b}^{j_{a:b}} := (z_a^{j_a}, \ldots, z_b^{j_b})$. For $N \in \mathbb{N}$, we denote $[N] = \{1, 2, \cdots, N\}$. The sequence $1{:}N$ with k omitted and ℓ duplicated is denoted as $[k \to \ell]_N := \big(1{:}(k-1), (k+1){:}\ell, \ell{:}N\big)$. The notation '$\mathrm{d}x$' implicitly stands for a σ-finite dominating measure on X, integers are equipped with the counting measure, product spaces are equipped with products of dominating measures and test functions are implicitly assumed to be measurable.

Let us then turn to the particle filter algorithm based on the Feynman–Kac model. The particle filter involves one additional ingredient: the resampling mechanism, which is determined by a probability distribution $r(\,\cdot\, \mid g^{1:N})$ on $[N]^N$, given unnormalised weights $g^{1:N} \in [0, \infty)^N$. We only consider resampling schemes r, which satisfy the following condition (which may be traced back to (4) in [7]):

ASSUMPTION 1. *Whenever* $\sum_{i=1}^{N} g^i > 0$, *the indices* $A^{1:N} \sim r(\,\cdot\, \mid g^{1:N})$ *satisfy*

(2) $$\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\left(A^i = j\right)\right] = \frac{g^j}{\sum_{i=1}^{N} g^i}$$

for all $j \in [N]$.

A resampling method that satisfies this assumption is known as being *unbiased* [1] since the expected number occurrences of outcome j in the population $A^{1:N}$ is $Ng^j/(\sum_{i=1}^{N} g^i)$. Algorithm 1 presents the particle filter in pseudo-code.

4

Algorithm 1 PARTICLEFILTER$(M_{1:T}, G_{1:T-1}, r, N)$

1: Draw $X_1^i \sim M_1(\cdot)$ and set $\boldsymbol{X}_1^i = X_1^i$ for $i \in [N]$.
2: **for** $k = 1, \ldots, T-1$ **do**
3:     Draw $A_k^{1:N} \sim r(\cdot \mid G_k(\boldsymbol{X}_k^1), \ldots, G_k(\boldsymbol{X}_k^N))$
4:     Draw $X_{k+1}^i \sim M_{k+1}(\cdot \mid X_k^{A_k^i})$ and set $\boldsymbol{X}_{k+1}^i = (\boldsymbol{X}_k^{A_k^i}, X_{k+1}^i)$ for $i \in [N]$.
5: **end for**
6: **output** $((X_1^{1:N}, \ldots, X_T^{1:N}), (A_1^{1:N}, \ldots, A_{T-1}^{1:N})))$

With the shorthand $\underline{x}_k = x_k^{1:N}$ for the whole particle system, we may write the law of the output $(\underline{X}_{1:T}, \underline{A}_{1:T-1})$ of Algorithm 1 in the following form:

$$(3) \quad \zeta^{(N)}(\underline{x}_{1:T}, \underline{a}_{1:T-1}) = \left(\prod_{i=1}^N M_1(x_1^i) \right) \left(\prod_{k=1}^{T-1} r(\underline{a}_k \mid G_k(\underline{x}_k)) \prod_{i=1}^N M_{k+1}(x_{k+1}^i \mid x_k^{a_k^i}) \right).$$

As in Algorithm 1, denote $\boldsymbol{x}_1^i = x_1^i$ and $\boldsymbol{x}_{k+1}^i = (\boldsymbol{x}_k^{a_k^i}, x_{k+1}^i)$. We have used the shorthand $G_k(\underline{x}_k)$ in the second argument of $r(\cdot \mid \cdot)$ to mean $G_k(\boldsymbol{x}_k^1), \ldots, G_k(\boldsymbol{x}_k^N)$.

Under Assumption 1, the output of Algorithm 1 satisfies the following unbiasedness condition [8, Theorem 7.4.2], which is key for particle Markov chain Monte Carlo [1]:

$$(4) \quad \mathbb{E}_{\zeta^{(N)}} \left[\left(\prod_{k=1}^T \frac{1}{N} \sum_{i=1}^N G_k(\mathbf{X}_k^i) \right) \sum_{i=1}^N W_T^i f(\mathbf{X}_T^i) \right] = \gamma_T(f) := \int f(x_{1:T}) \gamma_T(x_{1:T}) dx_{1:T},$$

where $W_T^i := G_T(\mathbf{X}_T^i) / \sum_{j=1}^N G_T(\mathbf{X}_T^j)$. In addition, under further conditions on $M_{1:T}, G_{1:T}$ and f, the following consistency result holds (see e.g. Chapter 11 of [5] and references therein, e.g. [7]):

$$(5) \quad \sum_{i=1}^N W_T^i f(\mathbf{X}_T^i) \xrightarrow{N \to \infty} \pi_T(f) := \int f(x_{1:T}) \pi_T(x_{1:T}) dx_{1:T} \qquad \text{in probability.}$$

3. Continuous-time path integral model. Continuous-time Feynman–Kac path integral models are the continuous-time analogue of hidden Markov models discussed above. The smoothing distribution is defined in terms of expectations of real-valued test functions ϕ on the path space (more precisely, the Skorohod space $D_\mathsf{X}[0, \tau]$ of càdlàg paths of a separable metric space X):

$$(6) \quad \Pi(\phi) := \frac{1}{\mathcal{Z}_\mathbb{M}} \mathbb{E}_\mathbb{M} \left[\phi(Z_{[0,\tau]}) \exp \left(- \int_0^\tau V_u(Z_u) du \right) \right]$$

$$\text{with} \quad \mathcal{Z}_\mathbb{M} := \mathbb{E}_\mathbb{M} \left[\exp \left(- \int_0^\tau V_u(Z_u) du \right) \right],$$

where $(V_u)_{0 \le u \le \tau}$ is a sufficiently regular family of non-negative potential functions on X, and $\mathbb{M}$ is the law of a Markov process $Z_{[0,\tau]} := (Z_u)_{0 \le u \le \tau}$ on X.

We focus on an approximation of the law (6) based on a time-discretisation $0 = t_1 < \cdots < t_T = \tau$ of the form

$$(7) \quad \Pi_T(\phi) := \frac{1}{\mathcal{Z}_T} \mathbb{E}_\mathbb{M} \left[\phi(\hat{Z}_{[0,\tau]}) \prod_{k=1}^{T-1} G_k(Z_{t_k}, Z_{t_{k+1}}) \right], \quad \mathcal{Z}_T := \mathbb{E}_\mathbb{M} \left[\prod_{k=1}^{T-1} G_k(Z_{t_k}, Z_{t_{k+1}}) \right],$$

where $\hat{Z}_u := \sum_{k=1}^{T-1} \mathbf{1}(u \in [t_k, t_{k+1})) Z_{t_k} + \mathbf{1}(u = t_T) Z_{t_T}$ is a càdlàg extension of the skeleton $(Z_{t_1}, \ldots, Z_{t_T})$ and where the potential functions $G_k(Z_{t_k}, Z_{t_{k+1}}) \ge 0$ are approximations

of $\exp\left(-\int_{t_k}^{t_{k+1}} V_u(Z_u)\mathrm{d}u\right)$ that can depend only on the values of Z_u at times t_k and t_{k+1}. Our theoretical focus is on a simple Euler-type form $G_k(Z_{t_k}, Z_{t_{k+1}}) := G_k(Z_{t_k}) := \exp\left(-(t_{k+1} - t_k)V_{t_k}(Z_{t_k})\right)$, but our method is also applicable to other approximation schemes. For the fixed time-discretisation $0 = t_1 < \cdots < t_T = \tau$, we may define M_1 as the initial distribution of Z_{t_1}; M_k for $2 \le k \le T$ as (an approximation of) the conditional distribution of $Z_{t_k} \mid Z_{t_{k-1}}$; $G_1 \equiv 1$; and for $1 \le k \le T-1$, $G_k(Z_{t_k})$ as just defined.

Having just declared the Markov kernels $M_{1:T}$ and the potential functions $G_{1:T-1}$, we can now employ Algorithm 1 to form a particle approximation of $\Pi_T(\cdot)$ in (7) using (5); and also an unbiased approximation of its normalising constant $\mathcal{Z}_T$ using (4). Write

$$X^{(T)} := \left((X_k^{(T)})^{1:N}\right)_{1 \le k \le T}$$

for the resulting particle system, where the superscript $^{(T)}$ refers to the discretisation $(t_k)_{1 \le k \le T} \subset [0, \tau]$. Our main focus in Sections 5 and 6 is to study the particle approximations as the discretisation is refined, that is as $T \to \infty$, but for a fixed N. We now give a flavour of the results.

The first question is whether the law of the discrete-time particle system converges in some sense to a law of the form (6) corresponding to a continuous-time Feynman–Kac model. To enable this convergence study, in Section 4 below, we introduce a stability condition (Assumption 2) that encompasses several known unbiased resampling schemes. In Sections 5 and 6, we then present the main convergence results (Theorems 19 and 22) for the particle approximations in the context of Itô diffusions for $\mathbb{M}$ (see (9)). First, Theorem 19 studies the convergence of the continuous-time extension of the population of particles $((\hat{X}_t^{(T)})^{1:N})_{0 \le t \le \tau}$, with càdlàg paths in $(\mathbb{R}^d)^N$ defined by

$$(\hat{X}_u^{(T)})^i := \sum_{k=1}^{T-1} \mathbf{1}\left(u \in [t_k, t_{k+1})\right)(X_k^{(T)})^i + \mathbf{1}\left(u = t_T\right)(X_T^{(T)})^i, \qquad i \in [N].$$

In particular, parts (i) and (ii) of Theorem 19 identify a continuous-time Markov process with càdlàg paths in $(\mathbb{R}^d)^N$ that $((\hat{X}_t^{(T)})^{1:N})_{0 \le t \le \tau}$ converges to with respect to finite-dimensional distributions as $T \to \infty$.

Then in Theorem 22 (result (12)) we show that a certain unnormalised time-marginal of this limiting continuous-time Markov process coincides with the unnormalised time-marginal of $\mathcal{Z}_\mathbb{M} \times \Pi(\cdot)$ in (6). Using part (iii) of Theorem 19 and result (12) of Theorem 22, we then conclude that the particle approximation of the (unnormalised) time-marginal Feynman–Kac path integral converges (for a fixed N) as the discretisation is refined:

$$\lim_{T \to \infty} \mathbb{E}\left\{\left(\frac{1}{N}\sum_{i=1}^N f\left((\hat{X}_\tau^{(T)})^i\right)\right)\exp\left[-\int_0^\tau \left(\frac{1}{N}\sum_{i=1}^N V_u\left((\hat{X}_u^{(T)})^i\right)\right)\mathrm{d}u\right]\right\} = \Pi(f) \times \mathcal{Z}_\mathbb{M},$$

where $f : \mathbb{R}^d \to \mathbb{R}$ is bounded and continuous, and $\Pi(f)$ and $\mathcal{Z}_\mathbb{M}$ are as in (6) with $f(Z_{[0,\tau]}) := f(Z_\tau)$. Note that the integrals in the exponential term of the left hand side are easy to evaluate as the $(\hat{X}^{(T)})^i$ are piecewise constant càdlàg paths.

4. Discretisation stable resampling schemes. Our main focus is on resampling schemes which lead to a *valid continuous-time limit* under infinitesimally refined discretisations. It turns out that the following condition naturally ensures a such continuous-time limit of a particle filter, and can provide insight about different resampling schemes.

ASSUMPTION 2. *For all $v^{1:N} \in [0, \infty)^N$ and for all $a_{1:N} \in [N]^N \setminus \{1:N\}$, the limit*

$$(8) \qquad \lim_{\Delta \to 0^+} \frac{1}{\Delta} r\left(a^{1:N} \mid \left(\exp(-\Delta v^1), \ldots, \exp(-\Delta v^N)\right)\right) =: \iota(a^{1:N}, v^{1:N}).$$

exists, and for any $v^* > 0$ the term inside the left-hand side limit is uniformly bounded for $v^{1:N} \in [0, v^*]^N$ and $\Delta \in (0, 1)$.

The limiting quantity $\iota(a^{1:N}, v^{1:N})$ can be interpreted as the *resampling intensity* corresponding to the configuration $a^{1:N} \neq 1:N$, with instantaneous potential values $v^{1:N}$. It can be thought of as the 'infinitesimal generator' stemming from the resampling r in the continuous-time limit. The sum of all resampling configurations

$$\iota^*(v^{1:N}) := \sum_{a^{1:N} \neq 1:N} \iota(a^{1:N}, v^{1:N})$$

is the *overall resampling rate*, that is, the intensity of any 'event' $a^{1:N} \neq 1:N$.

The basic and popular multinomial resampling scheme, which may be traced back to [18], does not admit a continuous-time limit: the probability of survival, that is, getting any permutation of $1:N$, does not tend to unity as $\Delta \to 0$. The same holds for the residual resampling introduced by [20, 23].

Perhaps the simplest scheme which satisfies this condition is a discrete-time version of the 'killing' resampling [9]. In discrete-time killing, the particle at index i 'survives' with probability proportional to the unnormalised weight g^i, and otherwise will be replaced with any other particle j, with probabilities proportional to g^j. We focus in particular on the following version of discrete-time killing:

DEFINITION 3 (Killing resampling).

$$r_{\text{killing}}(a^{1:N} \mid g^{1:N}) := \prod_{i=1}^N \left[\mathbf{1}\left(a^i = i\right) \frac{g^i}{g^*} + \left(1 - \frac{g^i}{g^*}\right) \sum_{j=1}^N \mathbf{1}\left(a^i = j\right) \frac{g^j}{\sum_{\ell=1}^N g^\ell} \right],$$

where $g^* = \max_{i \in [N]} g^i$.

In fact, any g^* such that $g^i/g^* \in [0, 1]$ for all $i \in [N]$ yields a valid unbiased resampling. The choice of g^* above, which was also used in the algorithmic 'rejection' variant of [24], ensures the highest survival probability, that is, $a^{1:N} = 1:N$. The following result can be verified by a direct calculation.

PROPOSITION 4. *Killing resampling satisfies Assumptions 1 and 2 and has limiting intensity*

$$\iota_{\text{killing}}(a^{1:N}, v^{1:N}) = \begin{cases} \frac{1}{N}(v^i - v_{\min}) & \text{if } a^i \neq i, \, a^{\neg i} = \neg i, \\ 0 & \text{otherwise}, \end{cases}$$

where $\neg i := (1, \ldots, i-1, i+1, \ldots, N)$ and where $v_{\min} := \min_{j \in [N]} v^j$. Consequently,

$$\iota^*_{\text{killing}}(v^{1:N}) = \frac{N-1}{N} \sum_{i=1}^N (v^i - v_{\min}) = (N-1)(\bar{v} - v_{\min}),$$

where $\bar{v} := N^{-1} \sum_{i=1}^N v^i$ is the mean of the potential values $v^{1:N}$.

4.1. *Stratified and systematic resampling.* For the rest of Section 4, we assume fixed unnormalised weights $g^{1:N} \in [0, \infty)^N \setminus \{0\}^N$ and denote the corresponding normalised weights by $w^i = g^i / \sum_{j=1}^N g^j$, and the cumulative distribution function by $F(0) \equiv 0$ and $F(i) = \sum_{j=1}^i w^j$ for $i \in [N]$. The generalised inverse $F^{-1}(u)$ is defined for $u \in (0, 1)$ as the unique index $i \in [N]$ such that $F(i-1) < u \leq F(i)$.

DEFINITION 5 (Systematic resampling). *Simulate a single $\sim U(0, 1)$, set*

$$\check{U}^i := \frac{i - 1 + U}{N}$$

and define the resampling indices as $A^i := F^{-1}(\check{U}^i)$ for $i \in [N]$.

DEFINITION 6 (Stratified resampling). *Simulate $U^{1:N} \sim U(0, 1)$, set*

$$\check{U}^i := \frac{i - 1 + U^i}{N},$$

and define the resampling indices as $A^i := F^{-1}(\check{U}^i)$ for $i \in [N]$.

We consider slightly modified versions of these resampling schemes, which rely on an auxiliary ordering of weights. This allows for simpler analysis, but our experiments also suggest potential performance gains.

DEFINITION 7 (Mean partition). *Suppose that $u^{1:N} \in \mathbb{R}^N$. A permutation $\varpi : [N] \to [N]$ is a mean partition (order) for $u^{1:N}$, if the re-indexed vector $u_\varpi^i := u^{\varpi(i)}$ satisfies $u_\varpi^1, \ldots, u_\varpi^m \leq \bar{u}$ and $u_\varpi^{m+1}, \ldots, u_\varpi^N > \bar{u}$ for some $m \in [N]$, where $\bar{u} = N^{-1} \sum_{i=1}^N u^i$.*

A mean partition ϖ can be found in $O(N)$ time using Hoare's scheme [21].

DEFINITION 8 (Systematic/stratified resampling with order ϖ). *Let F_ϖ^{-1} denote the generalised inverse distribution function corresponding to the re-indexed weights $w_\varpi^{1:N}$. Set*

$$A^{\varpi(i)} := \varpi\big(F_\varpi^{-1}(\check{U}^i)\big),$$

where $\check{U}^{1:N}$ are defined as in systematic/stratified resampling.

In words, Definition 8 means that we process the particles in order ϖ within systematic/stratified resampling. We obtain the following convergence results, whose proofs are given in Appendix A.

PROPOSITION 9. *Let ϖ be a mean partition for $-v^{1:N}$. Stratified resampling with order ϖ (Definition 8) satisfies Assumption 2 with resampling intensity*

$$\iota_{\text{stratified}}(a^{1:N}, v^{1:N}) = \begin{cases} \sum_{j=1}^i \big(v^{\varpi(j)} - \bar{v}\big), & a^{\varpi(i)} = \varpi(i+1), a^{\varpi(\neg i)} = \varpi(\neg i) \text{ for } i \in [N-1]. \\ 0, & \text{otherwise} \end{cases}$$

The overall resampling rate is

$$\iota_{\text{stratified}}^*(v^{1:N}) = \sum_{j=1}^N j(\bar{v} - v^{\varpi(j)}).$$

PROPOSITION 10. *Let ϖ be a mean partition for $-v^{1:N}$. Systematic resampling with order ϖ (Definition 8) satisfies Assumption 2 with resampling intensity*

$$\iota_{\text{systematic}}(a^{1:N}, v^{1:N}) = \begin{cases} \left(\min\{s_\varpi^k, s_\varpi^{\ell-1}\} - \max\{s_\varpi^{k-1}, s_\varpi^\ell\} \right)_+, & a^{\varpi(1:N)} = \varpi([k \to \ell]_N), \\ 0, & \text{otherwise,} \end{cases}$$

where k, ℓ are such that $v^{\varpi(k)} \geq \bar{v}$ and $v^{\varpi(\ell)} < \bar{v}$ and $s_\varpi^0 := 0$, $s_\varpi^i := \sum_{j=1}^i (v^{\varpi(j)} - \bar{v})$. The overall resampling rate is

$$\iota^*_{\text{systematic}}(v^{1:N}) = \sum_{i=1}^N (\bar{v} - v^i)_+.$$

REMARK 11. In a discrete time implementation, ϖ can be selected as mean partition of unnormalised weights $g^{1:N}$. In practice, when the mean partition ϖ for $g^{1:N} = \exp(-\Delta v^{1:N})$ is computed by Hoare's scheme, the mean partition will converge to a mean partition for $-v^{1:N}$.

REMARK 12. We believe that 'plain' stratified and systematic resampling (without mean partition) also satisfy Assumption 2, but verification becomes more technical. In addition, our empirical findings suggest that the mean partitioned versions can be preferable.

4.2. *SSP resampling.* We consider next a variant of SSP resampling [16] based on a processing order (permutation) ϖ; see Algorithm 2. The function REPEATINDICES($r^{1:N}$) in

---

Algorithm 2 SSPRESAMPLING($w^{1:N}, \varpi$)

1: Let $r^{1:N} \leftarrow \lfloor Nw^{1:N} \rfloor$, $p^{1:N} \leftarrow Nw^{1:N} - r^{1:N}$. and $(i,j) \leftarrow (\varpi(1), \varpi(2))$.
2: **for** $k = 2, \ldots, N$ **do**
3:     Set $\delta^i \leftarrow \min\{p^j, 1 - p^i\}$ and $\delta^j \leftarrow \min\{p^i, 1 - p^j\}$
4:     With probability $\mathbf{1}\left(\delta^i > 0\right) \frac{\delta^i}{\delta^i + \delta^j}$, interchange $(i,j) \leftarrow (j,i)$.
5:     **if** $p^i + p^j < 1$ **then**
6:         Set $p^i \leftarrow p^i + \delta^i$ and $j \leftarrow \varpi(\min\{k+1, N\})$.
7:     **else**
8:         Increment $r^i \leftarrow r^i + 1$, set $p^j \leftarrow p^j - \delta^i$ and $i \leftarrow \varpi(\min\{k+1, N\})$.
9:     **end if**
10: **end for**
11: Return $A^{1:N} \leftarrow$ REPEATINDICES($r^{1:N}$)

Algorithm 2 returns the non-decreasing index vector $A^{1:N}$ such that $\#\{j \in [N] : A^j = i\} = r^i$.

PROPOSITION 13. *SSP resampling with mean partition order ϖ of $-g^{1:N}$ satisfies Assumption 2 with intensity*

$$\iota_{\text{ssp}}(a^{1:N}, v^{1:N}) = \begin{cases} \dfrac{(v^k - \bar{v})_+ (\bar{v} - v^\ell)_+}{\iota^*_{\text{ssp}}(v^{1:N})} & a^{1:N} = [k \to \ell]_N \\ 0, & \text{otherwise,} \end{cases}$$

with overall resampling intensity $\iota^*_{\text{ssp}}(v^{1:N}) = \iota^*_{\text{systematic}}(v^{1:N}) = \sum_{i=1}^N (v^i - \bar{v})_+$.

Proposition 13 follows directly from Proposition 29 and Lemma 27 in Appendix A.

REMARK 14. The overall resampling intensity of the SSP resampling coincides with systematic resampling with mean partition: $\iota^*_{\text{ssp}}(v^{1:N}) = \iota^*_{\text{systematic}}(v^{1:N})$. A closer inspection reveals that also the *marginal intensities* for elimination of a particle k, or duplication of particle ℓ, coincide. However, the elimination and duplication indices K, L, respectively, are independent in the case of SSP resampling, in contrast with systematic resampling, where they have a (somewhat complicated) dependence.

4.3. *Comparison of resampling rates and a simplified limiting scheme.* In killing, and in stratified/systematic/SSP resampling based on a mean partition ϖ, exactly one particle is eliminated and one is duplicated in the limit. Therefore, the overall resampling event rate ι^* determines the instantaneous expected number of 'deaths' in all of these schemes. This motivates comparing the overall resampling rates.

THEOREM 15. *The overall resampling intensities of killing, stratified and systematic/SSP resampling with mean partition ϖ of $-v^{1:N}$ satisfy*

$$\iota^*_{\text{killing}}(v^{1:N}) \geq \iota^*_{\text{systematic}}(v^{1:N}), \qquad and \qquad \iota^*_{\text{stratified}}(v^{1:N}) \geq \iota^*_{\text{systematic}}(v^{1:N}),$$

for all potential values $v^{1:N}$. However, ι^*_{killing} and $\iota^*_{\text{stratified}}$ do not satisfy such order in general.

Theorem 15, whose proof is given in Appendix A, shows that systematic and SSP resampling have the smallest overall resampling rate among the studied algorithms, which suggests that they may therefore be preferable over killing and stratified resampling.

Let us conclude this section with another scheme, which has the same limit as SSP resampling, but with more a transparent behaviour. Note that this scheme can only be used with fine enough discretisations.

DEFINITION 16 (Symmetrised systematic resampling). *Assume that $g^{1:N}$ are such that their corresponding normalised weights $w^{1:N}$ satisfy $p := \sum_{i=1}^{N} (Nw^i - 1)_+ \leq 1$ (cf. Assumption 24 in Appendix A).*

With probability $1 - p$, return $A^{1:N} = 1:N$; otherwise pick indices K and L on $[N]$ independently with probabilities

$$\mathbb{P}(K = k) = \frac{(1 - Nw^k)_+}{p} \qquad and \qquad \mathbb{P}(L = \ell) = \frac{(Nw^\ell - 1)_+}{p},$$

and return $A^{1:N} = [k \to \ell]_N$.

The following proposition is straightforward to check given Lemma 27 in Appendix A.

PROPOSITION 17. *Symmetrised systematic resampling (Definition 16) satisfies Assumption 2 with intensity $\iota_{\text{s.syst}}(a^{1:N}, v^{1:N}) = \iota_{\text{ssp}}(a^{1:N}, v^{1:N})$.*

REMARK 18. Symmetrised systematic resampling algorithm (Definition 16) can be used in place of another resampling scheme, such as SSP resampling, whenever the required condition is met (i.e. $p \leq 1$). Such a combination would yield slight computational benefits, as the symmetric systematic resampling only requires two uniform random variables.

5. Convergence to a continuous-time limit. Here we present a convergence result for particle filters as in Algorithm 1, targeting a time-discretised path-integral model as discussed in Section 3. The state space is $\mathsf{X} := \mathbb{R}^d$ and the transitions M_k correspond to appropriately scaled Euler-Maruyama type discretisations of the d-dimensional diffusion

$$\text{(9)} \qquad \mathrm{d}z_t = b(z_t)\mathrm{d}t + \sigma(z_t)\mathrm{d}W_t,$$

with $z_0 \sim \mu$ for some fixed $\mu \in \mathcal{P}(\mathbb{R}^d)$ and coefficient functions $b \colon \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma \colon \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$ specified below.

To this end, let $\tau > 0$ be a continuous-time horizon, $V \colon \mathbb{R}^d \to [0,\infty)$ be a bounded and continuous potential function and $(\Delta_n)_{n \in \mathbb{N}} \subset (0, \tau \wedge 1)$ be an arbitrary decreasing sequence of discretisation step sizes converging to zero.

For $n \in \mathbb{N}$, write $\widetilde{X}^{\Delta_n}$ for the $(\mathbb{R}^d)^N$-valued Markov chain given by Algorithm 1 with $T = \lfloor \tau/\Delta_n \rfloor + 1$, resampling scheme r satisfying Assumption 2 and the transitions $(M_k)_{k \in 1:T}$ and functions $(G_k)_{k \in 1:(T-1)}$ defined as follows:

- $M_1 = \mu$, and $M_k(\mathrm{d}y \mid x) = \mathbb{P}\big(x + b(x)\Delta_n + \sigma(x)B_k^{\Delta_n} \in \mathrm{d}y\big)$ for $k \in 2{:}T$, where the $B_k^{\Delta_n}$'s are independently distributed as $\mathcal{N}(0, \Delta_n I_{\mathbb{R}^d})$;
- $G_k(\underline{\boldsymbol{x}}_k) := \nu^{\Delta_n}(\underline{\boldsymbol{x}}_k) := \big(e^{-\Delta_n V(x_k^1)}, \cdots, e^{-\Delta_n V(x_k^N)}\big)$ for all $k \in 1{:}(T-1)$.

Note that each G_k in the above definition depends only on the states of the particles at time k, so that $(\widetilde{X}_k^{\Delta_n})_{k \in 1:T}$ is indeed a Markov chain.

Write then

$$X_k^{\Delta_n} := \widetilde{X}_{k+1}^{\Delta_n} \quad \text{for} \quad k \in \{0, 1, \cdots, \lfloor \tau/\Delta_n \rfloor\}.$$

This re-indexing is introduced since particles commence with 'time' index 1 in Algorithm 1. The next theorem proves convergence of the càdlàg extension of this skeleton which is defined as $X_{\lfloor t/\Delta_n \rfloor}^{\Delta_n}$ for $t \in [0, \tau]$.

Recall that by Assumption 2,

$$\lim_{n \to \infty} \frac{1}{\Delta_n} r\big(a \mid \nu^{\Delta_n}(x)\big) = \iota\big(a, (V(x^1), \cdots, V(x^N))\big) =: \iota^a(x)$$

for all $a \in [N]^N \setminus \{1{:}N\}$, with *bounded and pointwise convergence* with respect to $x := (x^1, \cdots, x^N) \in (\mathbb{R}^d)^N$ in the sense that the term inside the limit is uniformly bounded with respect to n and x.

THEOREM 19. *Let the $(\mathbb{R}^d)^N$-valued Markov chains X^{Δ_n}, $n \in \mathbb{N}$, be as above. Assume that the coefficient functions b and σ of the diffusion (9) are Lipschitz continuous and bounded, that the diffusion is uniformly non-degenerate in the sense that*

$$\text{(10)} \qquad \inf_{x \in \mathbb{R}^d} \inf_{\theta \in \mathbb{R}^d, |\theta|=1} |\sigma(x)\theta| > 0,$$

and that the functions $\iota^a \colon (\mathbb{R}^d)^N \to [0,\infty)$, $a \in [N]^N \setminus \{1{:}N\}$, are bounded and continuous.

(i) *There exists a continuous-time process $(Z_t)_{t \geq 0}$ with càdlàg paths in $(\mathbb{R}^d)^N$ such that*

$$\lim_{n \to \infty} (X_{\lfloor t_1/\Delta_n \rfloor}^{\Delta_n}, \cdots, X_{\lfloor t_T/\Delta_n \rfloor}^{\Delta_n}) = (Z_{t_1}, \cdots, Z_{t_T})$$

in distribution for all finite $\{t_1, \cdots, t_T\} \subset [0, \tau]$.

(ii) *The limit process Z in part (i) has infinitesimal generator*

$$\mathcal{L}f(x) := \sum_{j \in [N]} L^{(j)} f(x) + \sum_{a \in [N]^N \setminus \{1:N\}} \iota^a(x)\big(f(x^{a(1:N)}) - f(x)\big)$$

for infinitely differentiable and compactly supported $f \colon \mathbb{R}^{dN} \to \mathbb{R}$ *and* $x \in \mathbb{R}^{dN}$, *where* L *is the generator corresponding to the* d-*dimensional diffusion* (9), $L^{(j)}f(x)$ *stands for* $L[y \mapsto f(x^1, \cdots, x^{j-1}, y, x^{j+1}, \ldots, x^N)](x^j)$ *and* $x^{a(1:N)} = (x^{a(1)}, \cdots, x^{a(N)}) \in \mathbb{R}^{dN}$ *for* $x := (x^1, \cdots, x^N) \in \mathbb{R}^{dN}$ *and* $a \in [N]^N$.

(iii) *Let* $\mathcal{V} \colon [0, \infty) \times \mathbb{R}^{dN} \to [0, \infty)$ *be a bounded and continuous function. Then*

$$\lim_{n \to \infty} \mathbb{E}\Big[f(X_{\lfloor \tau/\Delta_n \rfloor}^{\Delta_n}) \prod_{k=0}^{\lfloor \tau/\Delta_n \rfloor - 1} e^{-\Delta_n \mathcal{V}(k\Delta_n, X_k^{\Delta_n})}\Big] = \mathbb{E}\Big[f(Z_\tau) \exp\big(-\int_0^\tau \mathcal{V}(u, Z_u)\mathrm{d}u\big)\Big]$$

for all bounded and continuous $f \colon \mathbb{R}^{dN} \to \mathbb{R}$.

The proof of Theorem 19 is given in Appendix B and the Supplementary Material [6].

REMARK 20. Regarding Theorem 19:

(i) The assumption about the boundedness and continuity of the functions ι^a above often follows automatically from the corresponding properties of the potential function V (cf. Proposition 4).

(ii) Theorem 5 in the Supplementary Material [6] is a more general variant of part (iii) of Theorem 19.

(iii) The result is formulated for time-homogeneous coefficient and potential functions b, σ and V for simpler exposition. In fact, by considering a time-augmented state space (i.e. $[0, \infty) \times \mathbb{R}^{dN}$ instead of $\mathbb{R}^{dN}$), an analysis similar to the one in the Supplementary Material [6] can be carried out for time-dependent coefficient and potential functions, resulting in a variant of the Theorem with a time-inhomogeneous limit process Z. We omit the details; see e.g. [14, Chapter 4, Section 7] for basic results corresponding to such generalisations.

(iv) In the special case where killing resampling is used, we recover as the limit the continuous-time particle system described in e.g. Section 1.5.2 of [8] or [10]. See also e.g. [25, Example 3.1.3 and Proposition 3.4] for a continuos-time particle model with overall resampling rate that interestingly coincides with that of the systematic and SSP resampling schemes (see Propositions 10 and 13).

6. Unbiased estimation of Feynman–Kac measures. Continuing the theme of Section 5, we explain how the unbiasedness condition of the resampling r (Assumption 1) leads to an unbiasedness property for the jumping intensities $\iota(a, \cdot)$ (Definition 21 below). This property will be applied for time-marginal Feynman–Kac measures of the particle filter on the continuous-time limit, namely Theorem 22 below, which is a continuous-time variant of the well-known property (4).

DEFINITION 21. *We say that a resampling scheme* r *satisfying Assumption 2 is* asymptotically unbiased *if*

$$\sum_{a \in [N]^N \setminus \{1:N\}} \iota(a, v^{1:N})\big(\#\{j \in [N] : a^j = i\} - 1\big) = \frac{1}{N}\sum_{j=1}^N v^j - v^i \tag{11}$$

for all $v^{1:N} \in [0, \infty)^N$ *and* $i \in [N]$.

In order to state the main result of this section, let us introduce the following notation: for functions $f \colon \mathbb{R}^d \to \mathbb{R}$, write $\overline{f} \colon (\mathbb{R}^d)^N \to \mathbb{R}$ for the function $x \mapsto \frac{1}{N}\sum_{i=1}^N f(x^i)$.

THEOREM 22. *Let*

(i) *r be an asymptotically unbiased resampling scheme;*
(ii) $Z := (Z_t)_{t \geq 0}$ *be the continuous-time particle filter in Theorem 19;*
(iii) $z := (z_t)_{t \geq 0}$ *be the solution to the d-dimensional stochastic differential equation (9) with initial distribution μ (the solution is unique in law under the assumptions of Theorem 19);*
(iv) $V \colon \mathbb{R}^d \to [0, \infty)$ *be the bounded and continuous potential function in (the construction preceding) Theorem 19.*

Then

$$(12) \qquad \mathbb{E}\Big[\overline{f}(Z_t) \exp\Big(-\int_0^t \overline{V}(Z_u)\mathrm{d}u\Big)\Big] = \mathbb{E}\Big[f(z_t) \exp\Big(-\int_0^t V(z_u)\mathrm{d}u\Big)\Big]$$

for all $t > 0$ and bounded and measurable $f \colon \mathbb{R}^d \to \mathbb{R}$.

Regarding assumption (i) in Theorem 22, we note that it follows from our standard Assumptions 1 and 2:

PROPOSITION 23. *Suppose that the resampling r satisfies Assumptions 1 and 2. Then it is asymptotically unbiased.*

In particular, the assumption holds for all resampling schemes examined in Section 4, but it may also hold for a richer class of resampling schemes, including ones that are not necessarily unbiased in the sense of Assumption 1.

The proofs of Theorem 22 and Proposition 23 are presented in Appendix C.

7. Experiments. We compare empirically the behaviour of a number of resampling algorithms in two experiments. Our first experiment is on an Ornstein-Uhlenbeck latent process and a 'box-shape' potential ('OU'). The second experiment is about the inference of a Cox process, that is, an inhomogeneous Poisson process with latent intensity, where we use the particle filter within a PMMH (particle marginal Metropolis-Hastings) sampler.

7.1. *Ornstein-Uhlenbeck process and box potential.* In this experiment, $\mathbb{M}$ corresponds to the law of a stationary Ornstein-Uhlenbeck process $(Z_t)_{t \in [0, \tau]}$ with initial distribution $Z_0 \sim N(0, \sigma_\infty^2)$ which solves the following stochastic differential equation

$$\mathrm{d}Z_t = -\theta Z_t \mathrm{d}t + \sigma \mathrm{d}W_t,$$

where $(W_t)_{t \geq 0}$ is the standard Brownian motion. The parameters are $\theta = 0.1$ and $\sigma = 1$ and the stationary variance is $\sigma_\infty = \sigma/\sqrt{2\theta} \approx 2.236$.

The potential function is a (relatively narrow) 'box-shaped' potential function of the following form:

$$V(x) := 6 \times \mathbf{1}\left(|x - 0.5| > 0.1\right).$$

We study the performance of the particle filter with different resampling schemes in gradually finer discretisation $\log_2 \Delta \in \{-12, -10, \dots, 0\}$, with different number of particles $N \in \{64, 128, 256, 512\}$. We repeat the particle filter 10,000 times with each configuration to obtain the root mean squared errors (RMSE) in the figures. We consider the resampling schemes discussed in Section 4. The resampling schemes that include mean partition order, are named with 'Partition'.
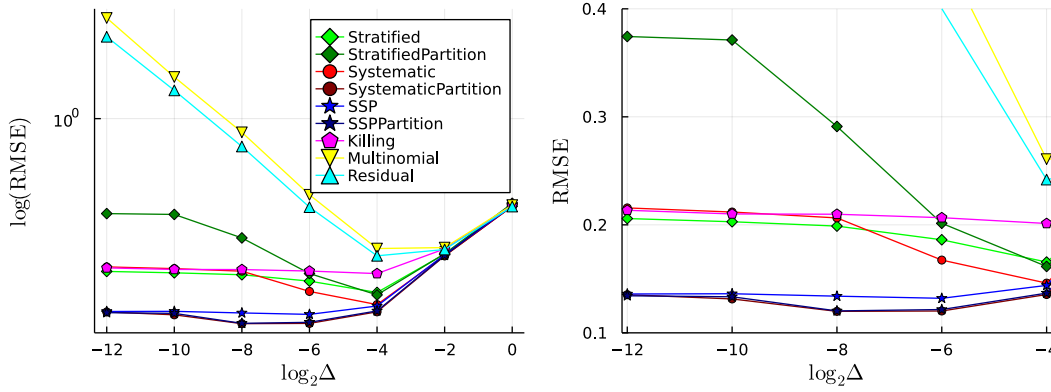
FIG 1. *Normalising constant estimates with $N = 512$ in the OU example. While the left panel uses a log scale and compares all the resampling schemes, the right panel uses a linear scale to compare more finely those resampling schemes that do not diverge as $\Delta \to 0$.*

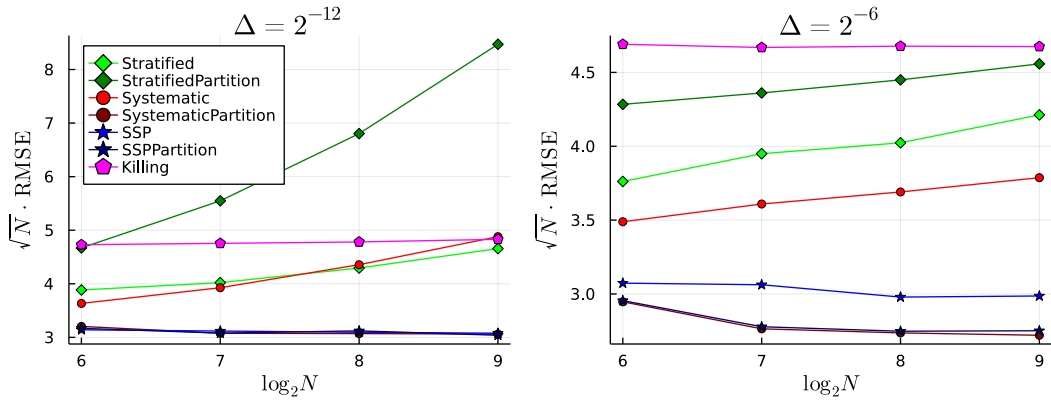FIG 2. *Normalising constant estimates with different N in the OU example. Each panel corresponds to a different value of Δ.*

For each Δ, we calculate the 'ground truth' of the normalising constant, defined as $\mathcal{Z}_T$ in (7), over all scenarios (all resampling schemes, all N). Taking this normalising constant as the truth, we construct unbiased estimators of the relative normalising constant (true value 1) and the filtering (the last state $\hat{Z}_\tau$ in (7)) and smoothing expectations (the first state $\hat{Z}_0$). For the latter, we use the 'filter smoother', that is, use traced back paths in estimation.

Figure 1 shows the normalising constant estimate mean squared errors (MSEs) in the case $N = 512$. When $\Delta = 2^0$, the performance is almost identical across resampling schemes, as we are not in the weakly informative setting, but this is no longer the case as $\Delta \to 0$. As expected, the performance of multinomial and residual resampling decay as $\Delta \to 0$, while other resampling schemes remain stable. The increase in relative RMSE for the multinomial scheme supports related results in [4] that show the variance of the normalising constant estimate increases exponentially with the number of resampling instances. Similar to multinomial, residual resampling rate does not stabilise (see comment after Assumption 2) and hence the observed variance increase as Δ decreases. The zoomed figure on the right suggests that the best performance is obtained with SystematicPartition and SSPPartition. SSP resampling is also close, and seems indistinguishable in the $\Delta \to 0$ limit.

Figure 2 shows a similar picture for varying N, with two choices of Δ. The values of the y-axis are RMSEs multiplied by $\sqrt{N}$, which is expected to stabilise if a central limit theorem
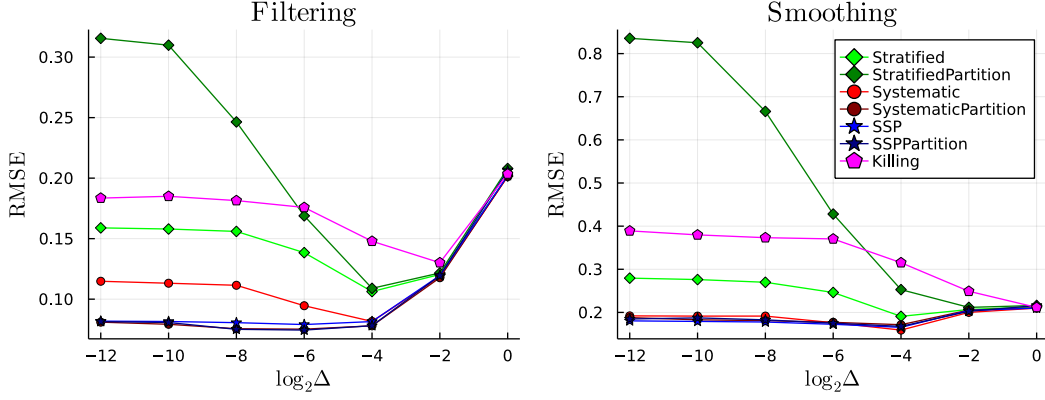
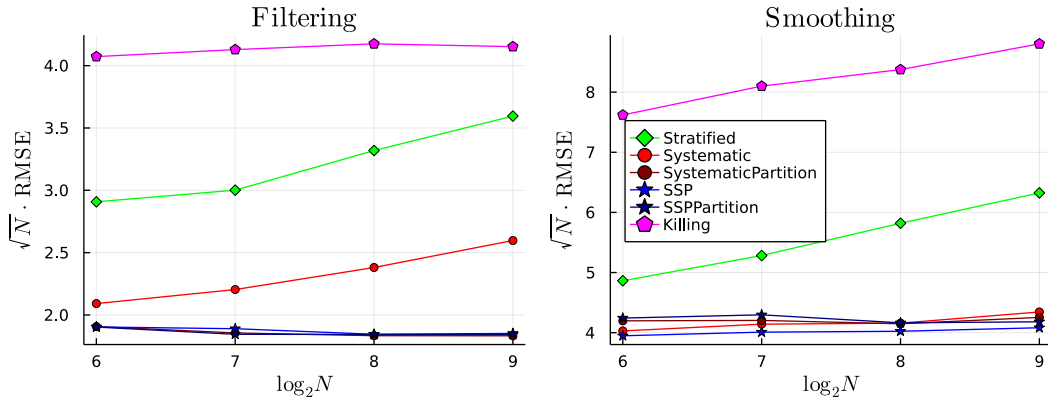FIG 3. *Filtering and smoothing estimates with $N = 512$ in the OU example.*

FIG 4. *Filtering and smoothing estimates with $\Delta = 2^{-12}$ and varying N in the OU example.*

type result holds. The results suggest that Killing, SSP, SSPPartition and SystematicPartition indeed stabilise, with the latter two being again the best.

Figure 3 shows filtering and smoothing estimate MSEs for stable resampling schemes similar to Figure 1. The conclusions are similar, except for systematic resampling, which seems to be competitive with the best schemes in smoothing, but not in filtering.

Figure 4 shows filtering and smoothing estimate MSEs scaled with $\sqrt{N}$ similar to Figure 2. Again, Killing, SSP, SSPPartition and SystematicPartition seem stable in the case of filtering, but smoothing with Killing has not stabilised yet.

7.2. *Comparison with adaptive resampling.* Adaptive resampling [22] is a commonly used method with particle filters, where resampling is performed only when so-called effective sample size of the weights falls below a predefined threshold (fraction of particles). Adaptive resampling is out of the scope of our theoretical framework, but can be useful in practice also in the weak potentials setting, so we compare empirically how adaptive resampling performs in the OU example (Section 7.1).

Figure 5 shows the performance with adaptive resampling with $N = 512$ particles and threshold $t_{\text{res}} = 0.5$, in the filtering and smoothing, similar to Figure 3. Adaptive resampling seems stable with all resamplings, the differences between resamplings are small, and the performance is competitive with the best non-adaptive resamplings. The behaviour with smaller number of particles is qualitatively similar to $N = 512$ (results not shown).
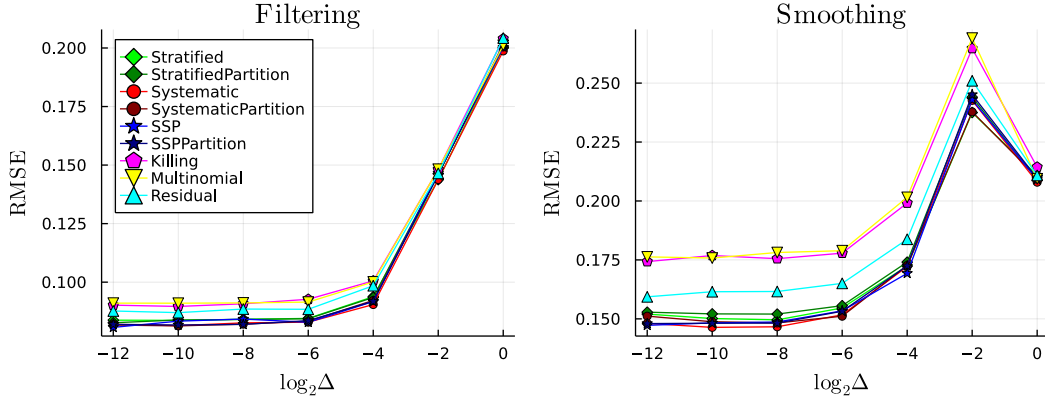
FIG 5. *Adaptive resampling with threshold* 0.5 *and* $N = 512$ *in the OU example.*
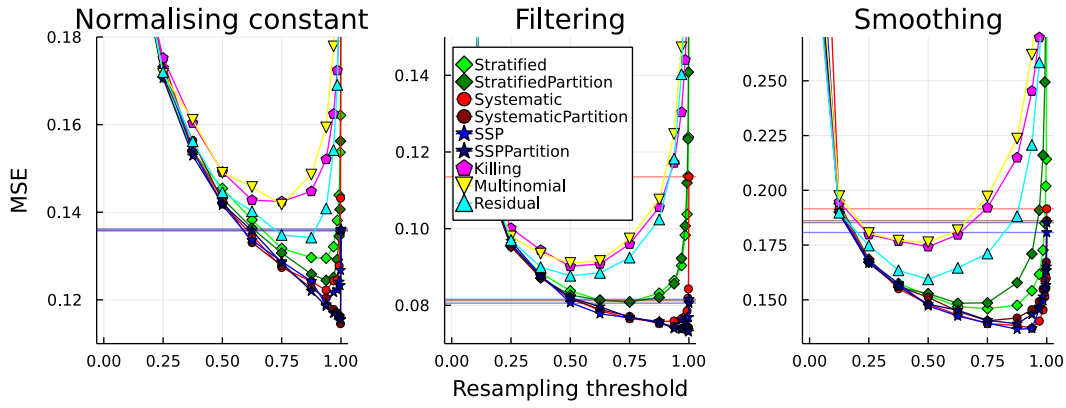
FIG 6. *Adaptive resampling with* $\log_2 \Delta = -12$ *and varying threshold in the OU example. The horizontal lines indicate the performance with non-adaptive resmplings* ($t_{\mathrm{res}} = 1.0$).

The threshold value, which controls how often resampling is triggered, is a tuning parameter of the method. We repeated the experiment with a range of thresholds $t_{\mathrm{res}} \in \{k/8 : k = 0, \ldots, 8\} \cup \{1 - 2^{-k} : k = 4, \ldots, 9\}$. Figure 6 shows a comparison of the results with finest discretisation $\Delta = 2^{-12}$. The differences between resamplings are small with low threshold values, but more noticeable with higher thresholds. For normalising constant estimation and filtering, adaptive Multinomial resampling does not reach the efficiency of the best non-adaptive schemes. In contrast, adaptive resampling can improve on the smoothing performance, and for instance with $t_{\mathrm{res}} = 0.5$ all adaptive resamplings outperform the best non-adaptive resampling. Interestingly, the optimal threshold value appears to depend on the resampling. For Multinomial, Residual and Killing, the optimal value is close to 0.5, but for SSPPartition and SystematicPartition, the optimal threshold is closer to one.

7.3. *Cox process with Particle marginal Metropolis-Hastings.* In our second example, we consider a Cox process model, that is, an inhomogeneous Poisson process with random intensity. We infer the latent intensity based on event times $0 < \tau_1 < \cdots < \tau_m < \tau$, leading to the following model:

$$(13) \qquad \Pi(\phi) := \frac{1}{\mathcal{Z}_{\mathbb{M}}} \mathbb{E}_{\mathbb{M}} \left[\phi \left(Z_{[0,\tau]} \right) \exp \left(- \int_0^\tau V_u(Z_u) \mathrm{d}u \right) \prod_{i=1}^m V_{\tau_i}(Z_{\tau_i}) \right],$$
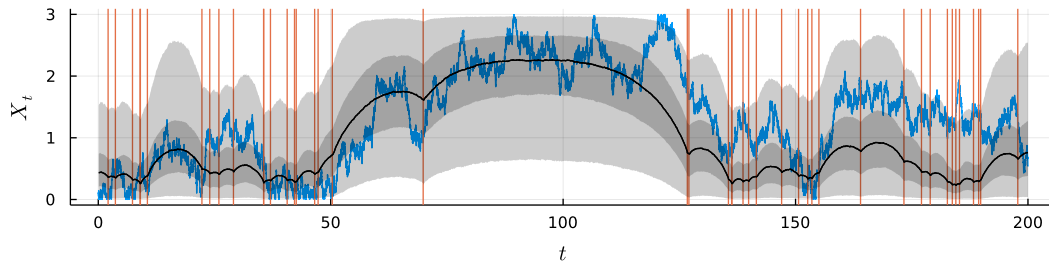
FIG 7. *Generated latent state (blue) and observed times y (red) in the Cox process experiment, as well as the posterior smoothing 50% and 95% credible intervals with PMMH using SystematicPartition and 32 particles.*

where $\mathbb{M}$ stands for the law of a reflected Brownian motion on $[a, b]$, and the potential $V_u(z) = \beta e^{-\alpha z}$.

We approximate the reflected Brownian motion with the discrete-time dynamics $M_1 = N(0, 1)$ and $M_k(\cdot \mid X_{k-1})$ for $k \geq 1$:

$$\hat{X}_k \sim N(X_{k-1}, \Delta_k \sigma^2); \qquad X_k = \text{reflect}(\hat{X}_k; a, b),$$

where $\Delta_k = t_k - t_{k-1}$ is the time difference between X_{k-1}, and X_k and reflect implements a folding back to $[a, b]$.

We consider synthetic data $\tau = (\tau_1, \ldots, \tau_n)$ generated from the model, where Z_t is the càdlàg extension of the skeleton $Z_{t_k} = X_k$ on $[0, T]$. We use the constant step size $\Delta_k = \Delta = 0.01$ and the parameter values $\sigma = 0.3$, $\alpha = 1.0$, $\beta = 0.5$ and $T = 200$ in the simulation.

We then use the particle marginal Metropolis-Hastings (PMMH) [1] to do posterior inference with independent $N(0, 2.5)$ prior for all log-transformed parameters $\theta = (\log \sigma, \log \alpha, \log \beta)$. The discretisation mesh is a uniform grid as in the data generation, augmented with the data points. The potentials are defined as follows:

$$\log G_k(x) = -\Delta_{k+1} \beta \exp(-\alpha x) + 1(t_k \in \tau)(\log(\beta) - \alpha x)$$

That is, the latter part is only included in case of data point is observed at t_k. The initial value of the PMMH is set to $\theta_0 = (0, 0, 0)$. We use the continuous covariance adaptation scheme of [19] within PMMH [cf. 26] during the entire simulation of 500,000 iterations, with 50,000 taken as as burn-in. We repeat the experiment with $N \in \{16, 32, 64, 128, 256\}$ particles and the same range of resampling algorithms as in the previous experiment.

Figure 7 shows the data in the experiment, and illustrates the inference outcome for the latent state. It is intuitive that there is substantial uncertainty in longer intervals with no observations. In these intervals, the potentials are weak, and so the resampling strategy is expected to have an impact in the efficiency.

Figure 8 (left) shows the PMMH acceptance rate in the different scenarios. The same group of SSP, SSPPartition and SystematicPartition attains the highest rates, and with multinomial and residual resampling, the acceptance remains notably lower. To attain a 10% acceptance rate, residual resampling needs 128 particles in contrast with 32 particles for the best resampling schemes.

Figure 8 (right) illustrates the mean inverse relative efficiencies (IREs) [17], that is, mean asymptotic variances of the standardised log-transformed parameters, multiplied by number of particles. The asymptotic variances are calculated by batch means [15], and standardisation is based on mean and variance estimates calculated from all outputs. The results are in line with earlier findings, but suggest that a low number of particles (even as low as 8) might be optimal in some cases. However, this might well be anomaly due to underestimation of asymptotic variance, which is supported by inspection of autocorrelation plots of the first parameter ($\log \alpha$) shown in Figure 9. Note that the lags are chosen inversely proportional to N to account for varying cost per iteration.
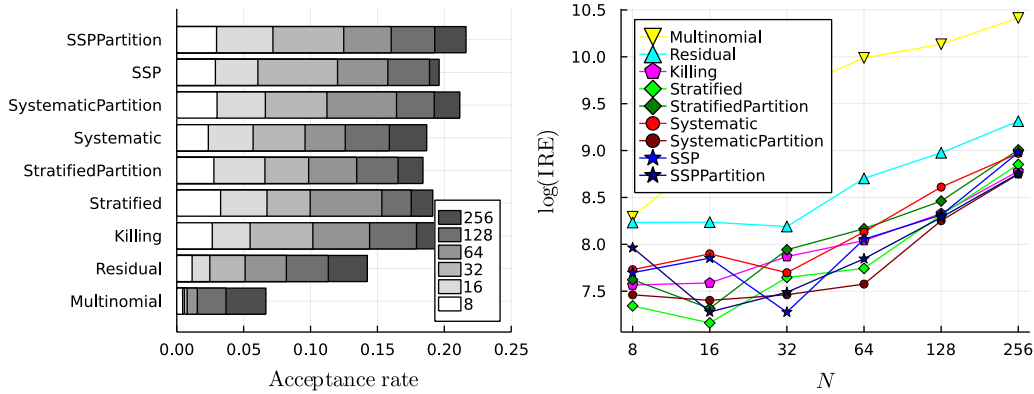
FIG 8. *Acceptance rate (left) and mean inverse relative efficiency (right) of PMMH in the Cox process experiment.*
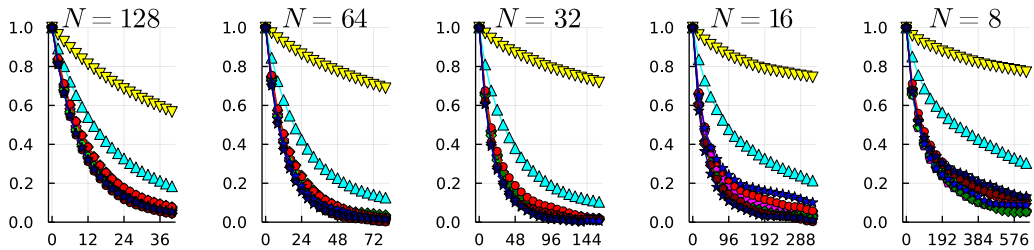

FIG 9. *Autocorrelation functions of $\log \alpha$ in the Cox process experiment.*

8. Discussion. We investigated the effect of resampling methods in a particle filter targeting a HMM with uninformative observations, by considering discretisations of continuous-time path integral models. We introduced a general condition for discrete-time resampling schemes which guarantees convergence to a non-degenerate particle system the continuous-time limit. We are unaware of earlier results establishing continuous-time limits of particle filters with different resampling strategies.

Resampling methods which satisfy our condition are 'safe' to use with weakly informative observations/potentials. We introduced modified versions of stratified/systematic/SSP resampling, which are shown to satisfy the condition. The modified strategies add a simple (and computationally cheap) algorithmic step to the resampling schemes, which orders the weights about their mean value. The modified algorithms lend themselves to a theoretical analysis, which reveals that systematic and SSP resampling schemes yield the smallest overall resampling rate, and may therefore be preferable.

Our empirical results complement our theoretical findings: systematic and SSP resampling with mean ordering had the best performance in all experiments. Because of the appealing theoretical properties of SSP resampling [cf. 16], it can be recommended also in the weakly informative regime. However, the systematic resampling may remain preferable in some settings, because of its slightly lower computational cost. Interestingly, the mean partition order, which was necessary for theoretical analysis, appears to improve the performance of systematic resampling as well. Based on our findings, we recommend that systematic resampling is always used together with the mean partition ordering of the weights. SSP resampling appears to perform well also without such pre-ordering.

Adaptive resampling [22] and further refinements, such as partial interaction schemes [28], can also be useful in the weakly informative setting, but are out of the scope of our theoretical

framework. Our empirical comparison suggests that adaptive resampling can further improve performance of the studied resampling algorithms. However, optimal choice of threshold is non-trivial, as it seems to depend on the resampling scheme.

APPENDIX A: PROOFS FOR SECTION 4

We first establish results for weights that are mean partitioned and nearly constant.

ASSUMPTION 24. *Let $w^{1:N}$ be normalised weights and write $w^i = \frac{1+\epsilon^i}{N}$ where $\epsilon^i = Nw^i - 1$. Suppose that $\sum_i |\epsilon^i| < 2$ and that there exists $m \in [N]$ such that $\epsilon^1, \ldots, \epsilon^m \leq 0$ and $\epsilon^{m+1}, \ldots, \epsilon^N > 0$.*

In what follows, under Assumption 24, we denote $c^0 = 0$ and $c^i := -\sum_{j=1}^i \epsilon^j$ for $i \in [N]$. Then we may write the distribution function corresponding to $w^{1:N}$ as follows:

$$F(i) = \sum_{j=1}^i w^j = \frac{1}{N}\left(i - c^i\right) \qquad \text{for} \qquad i = 0, \ldots, N.$$

LEMMA 25. *Under Assumption 24:*

(i) $c^1 \leq \cdots \leq c^m$, $c^m > \cdots > c^N$ *and* $c^i \in [0,1)$ *for* $i \in [m]$.
(ii) *For $u \in (0,1)$ and $\check{u}^i := (i - 1 + u)/N$, the following hold:*

$$F(i-1) < \check{u}^i \leq F(i) \iff u \leq 1 - c^i$$

$$F(i) < \check{u}^i \leq F(i+1) \iff u > 1 - c^i.$$

PROOF. Because $\sum_i |\epsilon^i| = \sum_i (\epsilon^i)_+ + \sum_i (-\epsilon^i)_+ < 2$, and $\sum_i \epsilon^i = 0$ so $\sum_i (\epsilon^i)_+ = \sum_i (-\epsilon^i)_+ < 1$, from which (i) follows, and (ii) is a direct consequence of $c^i \in [0,1)$. □

LEMMA 26. *Let $A^{1:N}$ be indices from stratified resampling (Definition 6). If Assumption 24 holds, then $A^i \in \{i, i+1\}$ for all $i \in [N]$ and for any $K \subset [N-1]$ and $S = [N] \setminus K$,*

$$\mathbb{P}(A^i = i,\, A^j = j+1 \text{ for all } i \in S \text{ and } j \in K) = \left(\prod_{j \in S}(1-c^j)\right)\left(\prod_{i \in K} c^i\right).$$

PROOF. Because the $\check{U}_i$'s are independent, we may write the probability of interest as

$$\left(\prod_{j \in S} \mathbb{P}\big(F(j-1) < \check{U}^j \leq F(j)\big)\right)\left(\prod_{i \in K} \mathbb{P}\big(F(i) < \check{U}^i \leq F(i+1)\big)\right),$$

from which the result follows by Lemma 25. □

LEMMA 27. *Let $v^{1:N} \geq 0$. The normalised weights $w_\Delta^{1:N}$ corresponding to unnormalised weights $g_\Delta^i = \exp(-\Delta v^i)$ may be written as*

$$w_\Delta^i = \frac{1 + \epsilon_\Delta^i}{N}, \qquad \text{where} \qquad \epsilon_\Delta^i = \Delta(\bar{v} - v^i) + r_\Delta^i$$

where $\bar{v} = N^{-1}\sum_{i=1}^N v^i$ stands for the mean potential and the error terms $r_\Delta^i = o(\Delta)$ and satisfy $|r_\Delta^i| \leq c\Delta$ for all $\Delta \in (0,1)$, where the constant c depends only on N and $\max_i v^i$. Consequently,

$$c_\Delta^i := -\sum_{j=1}^i \epsilon_\Delta^j = \Delta \sum_{j=1}^i (v^j - \bar{v}) + \tilde{r}_\Delta^i, \qquad \tilde{r}_\Delta^i = o(\Delta), \qquad |\tilde{r}_\Delta^i| \leq \tilde{c}\Delta.$$

PROOF. Direct calculation for $\epsilon_\Delta^i = N g_\Delta^i / \sum_{j=1}^N g_\Delta^j - 1$ yields that

$$\lim_{\Delta \to 0} \frac{d}{d\Delta} \epsilon_\Delta^i = \frac{1}{N} \sum_{j=1}^N v^j - v^i = \bar{v} - v^i,$$

and properties of the error term can be verified, for instance, by using a Taylor expansion for the exponential function. $\square$

PROOF OF PROPOSITION 9. Suppose first that $-v^{1:N}$ are mean ordered, and that Δ is sufficiently small so that also $w_\Delta^{1:N} \propto \exp(-\Delta v^{1:N})$ satisfy Assumption 24. Lemma 27 together with Lemma 26 give

$$\lim_{\Delta \to 0+} \frac{1}{\Delta} r\big(a^{1:N} \mid \exp(-\Delta v^{1:N})\big) = \begin{cases} \sum_{j=1}^i (v^j - \bar{v}), & a^i = i+1 \text{ and } a^{\neg i} = \neg i. \\ 0, & \text{otherwise.} \end{cases}$$

The corresponding overall resampling rate is therefore

$$\sum_{i=1}^N \sum_{j=1}^i (v^j - \bar{v}) = \sum_{j=1}^N (N + 1 - j)(v^j - \bar{v}) = \sum_{j=1}^N j(\bar{v} - v^j).$$

The claim follows from this result applied to re-indexed v_ϖ and a_ϖ. $\square$

LEMMA 28. *Let $A^{1:N}$ be indices from systematic resampling (Definition 5). If Assumption 24 holds, then for any $k \in [m]$ and $\ell \in [N] \setminus [m]$,*

$$\mathbb{P}\big(A^i = i \text{ for } i < k \text{ and } i \geq \ell, \, A^j = j + 1 \text{ for } k \leq j < \ell\big)$$
$$= \big(\min\{c^k, c^{\ell-1}\} - \max\{c^{k-1}, c^\ell\}\big)_+,$$

and these events are the only possible in addition to the 'no resampling' event, for which

$$\mathbb{P}(A^{1:N} = 1{:}N) = 1 - c^m.$$

PROOF. By Lemma 25, the event is equivalent to

$$\mathbb{P}\big(U \leq 1 - c^i \text{ for } i < k \text{ and } i \geq \ell, \, U > 1 - c^j \text{ for } k \leq j < \ell\big)$$
$$= \mathbb{P}\big(U \in (1 - c^k, 1 - c^{k-1}], U \in (1 - c^{\ell-1}, 1 - c^\ell]\big),$$

thanks to the monotonicity properties of $c^{1:m}$ and $c^{m+1:N}$. The latter follows similarly because $c^m = \max_{i \in [N]} c^i$. $\square$

PROOF OF PROPOSITION 10. Suppose first that $-v^{1:N}$ are mean ordered. Lemma 27 with Lemma 28 yield

$$\lim_{\Delta \to 0+} \frac{1}{\Delta} r\big(a^{1:N} \mid \exp(-\Delta v^{1:N})\big) = \begin{cases} \big(\min\{s^k, s^{\ell-1}\} - \max\{s^{k-1}, s^\ell\}\big)_+ & a^{1:N} = [k \to \ell]_N, \\ 0, & \text{otherwise,} \end{cases}$$

where $s^0 = 0$ and $s^i = \sum_{j=1}^i (v^j - \bar{v})$. The overall resampling rate is

$$\sum_{k=1}^m \sum_{\ell=m+1}^N \big(\min\{s^k, s^{\ell-1}\} - \max\{s^{k-1}, s^\ell\}\big)_+ = \sum_{k=1}^m (s^k - s^{k-1}) = s^m,$$

because $s^{0:m}$ is increasing and $s^{m:N}$ is decreasing. The result follows by re-indexing wrt. ϖ. $\square$

PROPOSITION 29. *Suppose that the normalised weights $w^{1:N}$ satisfy Assumption 24 and $\sum_{i=1}^{N} |\epsilon_i| < 1$. Then, for the SSP resampling with identity order ϖ, the only events with non-zero probability in addition to $1{:}N$ are of the form $A^{1:N} = [k \to \ell]_N$, with probabilities:*

$$\mathbb{P}(A^{1:N} = [k \to \ell]_N) = \frac{(-\epsilon^k)_+ (\epsilon^\ell)_+}{\sum_{i=1}^{N} (\epsilon^i)_+}, \qquad for\ k, \ell \in [N].$$

PROOF. Thanks to Assumption 24, the initial values of $p^{1:N}$ satisfy $p^i = 1 + \epsilon^i$ for $i = 1, \ldots, m$ and $p^i = \epsilon^j$ for $j = m+1, \ldots, N$, and now $c^i = -\sum_{j=1}^{i} \epsilon^j \in [0, 1/2)$.

Note that the state of Algorithm 2 after lines 3–4 is independent of the order of the indices (i, j) before, so without loss of generality, we may assume that $i < j = k$ always before line 3. We may deduce inductively that after line 3 with $k \in \{2{:}m\}$:

- $p^i = 1 - c^{k-1}$ and $p^j = 1 + \epsilon^k$, and
- $p^i > 1/2$ and $p^j > 1/2$ and therefore $\delta^i = 1 - p^i = c^{k-1}$ and $\delta^j = 1 - p^j = -\epsilon^k$.

With probability $\delta^j/(\delta^i + \delta^j) = -\epsilon^k/c^k$, the indices at next iteration will be $i = k$ and $j = k+1$, and $r^{1:k-1}$ have all been incremented by one. The probability to end up with indices $i = k$ and $j = m+1$ after iteration m is $(-\epsilon^k/c^k)\prod_{i=k+1}^{m}(c^{i-1}/c^i) = -\epsilon^k/c^m$, in which case all $r^{1:m}$ have been incremented by one, except for r^k.

Given the above scenario happens, then in the steps $k \in \{(m+1){:}(N-1)\}$ of the algorithm, it is again easy to see inductively that $p^i + p^j = 1 - c^k < 1$ so $\delta^{i,j} = p^{j,i}$ and that $p^i = 1 - c^{j-1}$. The probability to end up with $i = \ell$ and $j = N$ after iteration $N-1$ is therefore $\epsilon^\ell/(1-c^\ell)\prod_{j=\ell+1}^{N-1}(1-c^{j-1})/(1-c^j) = \epsilon^\ell/(1-c^{N-1})$, in which case in the beginning of the last step, $p^i + p^j = 1 - c^N = 1$. Now, r^ℓ will be incremented by one with probability $(1 - c^{N-1})$. We conclude the overall probability of outcome $[k \to \ell]_N$, which is equivalent to incrementing r^ℓ and $r^{1:m}$ by one except for r^k. $\qquad\square$

PROOF OF THEOREM 15. Note that for any i such that $(\bar{v} - v^i)_+ > 0$, that is, $v^i < \bar{v}$, we have

$$(\bar{v} - v^i)_+ = \bar{v} - v^i \le \bar{v} - v_{\min},$$

and there are of course at most $N-1$ such i, so

$$\iota^*_{\text{systematic}}(v^{1:N}) \le \#\{i : v^i < \bar{v}\}(\bar{v} - v_{\min}) \le (N-1)(\bar{v} - v_{\min}) = \iota^*_{\text{killing}}(v^{1:N}).$$

Assuming mean ordered $-v^{1:N}$ we may write

$$\iota^*_{\text{stratified}}(v^{1:N}) - \iota^*_{\text{systematic}}(v^{1:N}) = \sum_{j=1}^{m} j(\bar{v} - v^j) + \sum_{j=m+1}^{N} (j-1)(\bar{v} - v^j)$$

$$\ge m \sum_{j=1}^{m} (\bar{v} - v^j) + m \sum_{j=m+1}^{N} (\bar{v} - v^j) = 0.$$

To see that there cannot be such an order between ι^*_{killing} and $\iota^*_{\text{stratified}}$, consider $N = 3$ and strictly decreasing $v^{1:3}$. Now, $v_{\min} = v^3$ and

$$\iota^*_{\text{killing}}(v^{1:N}) - \iota^*_{\text{stratified}}(v^{1:N}) = 2(\bar{v} - v^3) - (\bar{v} - v^1) - 2(\bar{v} - v^2) - 3(\bar{v} - v^3) = v^2 - \bar{v}$$

$$= \frac{2}{3}\left(v^2 - \frac{v^1 + v^3}{2}\right),$$

which can be positive or negative depending on $v^2 \in (v^3, v^1)$. A similar example can be constructed for any $N > 3$ (we omit the details). $\qquad\square$

APPENDIX B: PROOF OF THEOREM 19

This section is dedicated to an outline of the proof of Theorem 19. The details of the proof are given in the Supplementary Material [6].

B.1. A new construction of X_k^Δ in Theorem 19 . For notational convenience, we present a new (equivalent) construction for the $(\mathbb{R}^d)^N$-valued Markov chains $\{X_k^\Delta\}_{k\in\mathbb{N}_0}$ for $\Delta\in\{\Delta_n:n\in\mathbb{N}\}$. This new construction is self-contained in the sense that it does not make reference to Algorithm 1 (which we did when introducing this chain in Section 5).

Denote by $\mu:=M_1\in\mathcal{P}(\mathbb{R}^d)$ the initial distribution for the particles, and by $W:=(W^1,\cdots,W^N)$ a fixed dN-dimensional Brownian motion (so that the W^j's are independent d-dimensional Brownian motions) with respect to a filtration $(\mathcal{F}_t)_{t\geq0}$. We then redefine the Markov chain $X^\Delta:=(X_k^\Delta)_{k\in\mathbb{N}_0}$ by $(X_0^\Delta)^j\sim\mu$ independently for $j\in[N]$ and

$$(X_{k+1}^\Delta)^j = (X_k^\Delta)^{A_k(j)} + b\big((X_k^\Delta)^{A_k(j)}\big)\Delta + \sigma\big((X_k^\Delta)^{A_k(j)}\big)\big(W_{(k+1)\Delta}^j - W_{k\Delta}^j\big)$$

for $k\in\mathbb{N}_0$ and $j\in[N]$. Here $A_k\in[N]^N$ stands for the multi-index resulting from the resampling at time k, i.e. $A_k\sim r(\cdot\,|\,\nu^\Delta(X_k^\Delta))$, and $A_k(j)\in[N]$ stands for the j'th index of A_k.

More precisely, we may write

$$(14)\qquad A_k = \sum_{\ell=1}^{N^N}\mathbf{1}\Big(\sum_{i=1}^{\ell-1}r(a_i\,|\,\nu^\Delta(X_k^\Delta)) < U_k^\Delta \leq \sum_{i=1}^{\ell}r(a_i\,|\,\nu^\Delta(X_k^\Delta))\Big)a_\ell,$$

where $\{a_1,\cdots,a_{N^N}\}$ is some fixed enumeration of $[N]^N$, and the U_k^Δ's are uniform random variables on $(0,1)$ independent of each other and the W^j's. We can take each U_k^Δ to be $\mathcal{F}_{k\Delta}$-measurable. The Markov chain X^Δ is then $(\mathcal{F}_{k\Delta})_{k\in\mathbb{N}_0}$-adapted for all $\Delta\in\{\Delta_n:n\in\mathbb{N}\}$.

Now define the continuous-time scaling Z^Δ of this Markov chain by

$$(15)\qquad Z_t^\Delta := X_{\lfloor t/\Delta\rfloor}^\Delta,\quad t\in[0,\infty),$$

so that Z^Δ is for all Δ an $(\mathcal{F}_t)_{t\geq0}$-adapted process with paths in $D_{\mathbb{R}^{dN}}[0,\infty)$, the Skorohod space of càdlàg paths in $\mathbb{R}^{dN}$.

B.2. Outline of the proof of Theorem 19 . The main steps in the proof of parts (i) and (ii) of the theorem are summarised below, where we add a prefix 'S' to the references to the Supplementary Material [6], so for instance Theorem S5 means Theorem 5 of [6]:

- In order to use a convergence result from [14], our first goal is to show that the family of processes $(Z^{\Delta_n})_{n\in\mathbb{N}}$ is relatively compact with respect to convergence in distribution, i.e. that $\{\mathbb{P}(Z^{\Delta_n}\in\cdot):n\in\mathbb{N}\}$ is a relatively compact set in the weak topology of $\mathcal{P}(D_{\mathbb{R}^{dN}}[0,\infty))$. This is done in Proposition S1.
- The second step is to declare the continuous-time process $(Z_t)_{t\geq0}$ that the càdlàg extensions $(Z_t^{\Delta_n})_{t\geq0}$ will converge to. The process $(Z_t)_{t\geq0}$ is the canonical càdlàg process $(\mathbb{X}_t)_{t\geq0}$ introduced immediately before Proposition S2, equipped with a law possessing the generator $\mathcal{L}$. Proposition S2 proves (by establishing the well-posedness of the corresponding martingale problem) that this law is uniquely determined and Markovian.
- We then show in Proposition S3 that the appropriately-scaled discrete-time derivative of the transition kernel of $(X_k^{\Delta_n})_{k\in\mathbb{N}_0}$, the skeleton of $(Z_t^{\Delta_n})_{t\geq0}$, converges in a suitable sense to the generator $\mathcal{L}$ as $n\to\infty$.
- Proposition S4 and Theorem S5 then complete the proof of parts (i) and (ii).
- The proof of part (iii) is presented in the final part of the Supplementary Material [6].

APPENDIX C: PROOF OF THEOREM 22

We start with the following auxiliary result, which states the intuitively simple fact that although the sample paths of the particle filter Z are discontinuous with probability 1, the probability of discontinuities (i.e. resampling-induced jumps) at any given time is negligible.

PROPOSITION 30. *Let Z be the càdlàg process in Theorem 19. Then*

$$\mathbb{P}\big(Z_t = Z_{t-}\big) = 1$$

for all $t > 0$.

PROOF. We only give a brief outline of a proof. It suffices to show that

(16) $$\mathbb{E}\big[|f(Z_t) - f(Z_{t-})|\big] = 0$$

for any bounded and Lipschitz continuous $f \colon \mathbb{R}^{dN} \to \mathbb{R}$. The càdlàg property implies that

$$f(z_t) - f(z_{t-}) = \lim_{\delta \to 0^+} \frac{1}{\delta} \Big(\int_t^{t+\delta} f(z_u)\mathrm{d}u - \int_{t-\delta}^t f(z_u)\mathrm{d}u \Big)$$

for all $z \in D_{\mathbb{R}^{dN}}[0,\infty)$ and $t > 0$ with bounded and pointwise convergence, and the expression inside the limit is for each $\delta > 0$ a continuous function of $z \in D_{\mathbb{R}^{dN}}[0,\infty)$. We can then proceed as in the proof of Theorem 5 of the Supplementary Material [6], noting that resampling-induced jumps of the processes Z^{Δ_n} happen with arbitrarily small probability on arbitrarily small time intervals. $\qquad\square$

PROOF OF THEOREM 22. It suffices to establish (12) for $f \in C_c^\infty(\mathbb{R}^d)$, the space of infinitely differentiable and compactly supported functions on $\mathbb{R}^{dN}$, since any bounded and measurable function on $\mathbb{R}^d$ can be approximated pointwise by an uniformly bounded sequence of functions in $C_c^\infty(\mathbb{R}^d)$.

Write

$$\mathcal{Q}_t(F) := \mathbb{E}\Big[F(Z_t)\exp\Big(-\int_0^t \overline{V}(Z_u)\mathrm{d}u\Big)\Big]$$

for $F \in C_c^\infty(\mathbb{R}^{dN})$ and

$$Q_t(f) := \mathbb{E}\Big[f(z_t)\exp\Big(-\int_0^t V(z_u)\mathrm{d}u\Big)\Big]$$

for $f \in C_c^\infty(\mathbb{R}^d)$.

For fixed F, the measure flow $\mathcal{Q}_t(F)$ is differentiable with respect to $t > 0$. In order to show this and compute $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{Q}_t(F)$, write $E_t := \exp(-\int_0^t \overline{V}(Z_u)\mathrm{d}u)$ for $t > 0$. Then for $\delta > 0$ with $\delta \ll 1$,

$$F(Z_{t+\delta})E_{t+\delta} - F(Z_t)E_t = E_t\big(F(Z_{t+\delta}) - F(Z_t)\big) + F(Z_{t+\delta})\big(E_{t+\delta} - E_t\big).$$

By the martingale problem (see the Supplementary Material [6]), $F(Z_{t+\delta}) - F(Z_t)$ can be written as $M_{t+\delta} - M_t + \int_t^{t+\delta} \mathcal{L}F(Z_u)\mathrm{d}u$ for some martingale M (with respect to the filtration $\mathcal{F}^Z$ generated by Z). Using this in combination with the tower property of conditional expectations (with respect to $\mathcal{F}_t^Z$), we can calculate

$$\mathcal{Q}_{t+\delta}(F) - \mathcal{Q}_t(F) = \mathbb{E}\Big[E_t \int_t^{t+\delta} \mathcal{L}F(Z_u)\mathrm{d}u\Big] + \mathbb{E}\big[F(Z_{t+\delta})\big(E_{t+\delta} - E_t\big)\big].$$

The càdlàg property in conjunction with the dominated convergence theorem then implies

$$\lim_{\delta \to 0^+} \frac{\mathcal{Q}_{t+\delta}(F) - \mathcal{Q}_t(F)}{\delta} = \mathbb{E}[\mathcal{L}F(Z_t)E_t] - \mathbb{E}[F(Z_t)\overline{V}(Z_t)E_t] = \mathcal{Q}_t\big((\mathcal{L} - \overline{V})F\big).$$

For negative δ, we may compute $\mathcal{Q}_{t+\delta}(F) - \mathcal{Q}_t(F)$ in a similar manner by using the tower property with respect to the filtration $\mathcal{F}^Z_{t+\delta}$ instead of $\mathcal{F}^Z_t$. Then Proposition 30 above together with the dominated convergence theorem imply

$$(17) \qquad \lim_{\delta \to 0^-} \frac{\mathcal{Q}_{t+\delta}(F) - \mathcal{Q}_t(F)}{\delta} = \mathcal{Q}_t\big((\mathcal{L} - \overline{V})F\big).$$

In a similar (in fact easier since the sample paths of z are automatically continuous) way we may calculate

$$(18) \qquad \frac{\mathrm{d}}{\mathrm{d}t} Q_t(f) = Q_t\big((L - V)f\big)$$

for any $f \in C^\infty_c(\mathbb{R}^d)$, where L is the infinitesimal generator corresponding to the diffusion (9).

Now the left-hand side of the statement of the Theorem is $\mathcal{Q}_t(\overline{f}) =: \hat{Q}_t(f)$ and the right-hand side is $Q_t(f)$. We will show that the evolution equation for $\hat{Q}$ is of the same form as (18). To this end, recall that $\mathcal{L} = \mathcal{L}^{\mathrm{mut}} + \mathcal{L}^{\mathrm{jump}}$ as in (11) of the Supplementary Material [6]. For $f \in C^\infty_c(\mathbb{R}^d)$, we simply get $\mathcal{L}^{\mathrm{mut}}(\overline{f}) = \overline{Lf}$, and further

$$\mathcal{L}^{\mathrm{jump}}(\overline{f})(x) = \sum_{a \neq 1:N} \iota^a(x)\Big(\frac{1}{N}\sum_{i=1}^N f(x^{a(i)}) - \frac{1}{N}\sum_{i=1}^N f(x^i)\Big)$$

$$= \frac{1}{N}\sum_{i=1}^N \Big(\sum_{a \neq 1:N} \iota^a(x)(\#_{a,i} - 1)\Big) f(x^i),$$

where $\#_{a,i} := \#\{j : a(j) = i\}$. Thus, comparing the right-hand sides of (17) and (18) (with $\overline{f}$ in place of F), we see that

$$\frac{\mathrm{d}}{\mathrm{d}t} \hat{Q}_t(f) = \hat{Q}_t\big((L - V)f\big)$$

is equivalent to

$$(19) \qquad \mathcal{Q}_t\Big(\frac{1}{N}\sum_{i=1}^N \Big(\sum_{a \neq 1:N} \iota^a(x)(\#_{a,i} - 1)\Big) f(x^i)\Big) = \mathcal{Q}_t(\overline{V} \cdot \overline{f} - \overline{Vf}).$$

Writing out the expression inside the right-hand side parentheses and comparing the coefficients of the $f(x^i)$'s, one sees that (19) will follow from the identity

$$(20) \qquad \sum_{a \neq 1:N} \iota^a(x)(\#_{a,i} - 1) = \overline{V}(x) - V(x^i)$$

for all i and x, which is simply assumption (i) of the Theorem.

Thus the measure flows Q and $\hat{Q}$ satisfy the same evolution equation, and by the assumptions we have $Q_0 = \hat{Q}_0$. Our next task is to verify that this evolution equation is well-posed in a suitable sense.

In order to work in a space of probability measures, let us denote by $\dot{\mathbb{R}}^d$ the standard one-point compactification of $\mathbb{R}^d$ with infinity point o. Define the probability measures $Q^o_t \in \mathcal{P}(\dot{\mathbb{R}}^d)$ and $\hat{Q}^o_t \in \mathcal{P}(\dot{\mathbb{R}}^d)$, $t \geq 0$, by

$$Q^o_t(f) := Q_t(f_{|\mathbb{R}^d}) + \big(1 - Q_t(1_{|\mathbb{R}^d})\big)f(o) = Q_t\big((f - f(o))_{|\mathbb{R}^d}\big) + f(o)$$

for (bounded and) continuous $f \colon \dot{\mathbb{R}}^d \to \mathbb{R}$ and similarly for $\hat{Q}$ in place of Q. Define M as the collection of continuous functions f on $\dot{\mathbb{R}}^d$ such that

$$\big(f - f(o)\big)_{|\mathbb{R}^d} \in C_c^\infty(\mathbb{R}^d),$$

and define the linear operator A on M by

$$Af(x) = L\big(f - f(o)\big)_{|\mathbb{R}^d}(x) + V(x)\big(f(o) - f(x)\big), \quad x \in \dot{\mathbb{R}}^d,$$

with the understanding that $Af(o) = 0$.

Now the flows Q^o and $\hat{Q}^o$ both satisfy the forward equation

$$(21) \qquad \mu_t(f) = \mu_0(f) + \int_0^t \mu_u(Af)\mathrm{d}u, \quad t > 0,$$

for (the natural extensions of) $f \in C_c^\infty(\mathbb{R}^d)$, and it is easy to see that this extends to $f \in M$. We are thus in a place to apply a uniqueness result from [14]: it is routinely verified that A satisfies the positive maximum principle, M is an algebra of functions that is dense in the space of continuous functions on $\dot{\mathbb{R}}^d$ (with respect to the sup-norm), and the $D_{\dot{\mathbb{R}}^d}[0,\infty)$-martingale problem for (A, M) is well-posed (see the Supplementary Material [6]). Thus [14, Chapter 4, Proposition 9.19] yields well-posedness for the forward equation (21).

In particular, $Q_t^o(f) = \hat{Q}_t^o(f)$ for all $t > 0$ and (extensions of) $f \in C_c^\infty(\mathbb{R}^d)$, which translates to $Q_t(f) = \hat{Q}_t(f)$. $\qquad\square$

Finally, let us prove Proposition 23:

PROOF OF PROPOSITION 23. Let $i \in [N]$, $v^{1:N} \in [0,\infty)^N$ and $\Delta > 0$. By Assumption 1,

$$\frac{Ne^{-\Delta v^i}}{\sum_{j=1}^N e^{-\Delta v^j}} = \sum_{j=1}^N r\big(a(j) = i \mid \exp(-\Delta v^{1:N})\big) = \sum_{j=1}^N \sum_{a \in [N]^N} r\big(a \mid \exp(-\Delta v^{1:N})\big)\mathbf{1}\,(a(j) = i)$$

$$= \sum_{a \in [N]^N} r\big(a \mid \exp(-\Delta v^{1:N})\big)\#\{j : a(j) = i\},$$

and subtracting $\sum_{a \in [N]^N} r(a \mid \exp(-\Delta v^{1:N})) \equiv 1$ from this yields

$$\sum_{a \neq 1:N} r\big(a \mid \exp(-\Delta v^{1:N})\big)\big(\#\{j : a(j) = i\} - 1\big) = \frac{Ne^{-\Delta v^i} - \sum_{j=1}^N e^{-\Delta v^j}}{\sum_{j=1}^N e^{-\Delta v^j}}.$$

Dividing this by Δ and taking $\Delta \to 0^+$ leads to (11), as in the proof of Lemma 27. $\qquad\square$

SUPPLEMENTARY MATERIAL

Supplement to: "On resampling schemes for particle filters with weakly informative observations"
Detailed proof of Theorem 19.

Source codes
Julia [3] source codes of the experiments in Section 7.

REFERENCES

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

[2] ARNAUDON, M. and DEL MORAL, P. (2020). A duality formula and a particle Gibbs sampler for continuous time Feynman-Kac measures on path spaces. *Electron. J. Probab.* **25** Paper No. 157, 54. https://doi.org/10.1214/20-ejp546

[3] BEZANSON, J., EDELMAN, A., KARPINSKI, S. and SHAH, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM review* **59** 65–98.

[4] CÉROU, F., MORAL, P. D. and GUYADER, A. (2011). A nonasymptotic theorem for unnormalized Feynman–Kac particle models. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* **47** 629 – 649.

[5] CHOPIN, N. and PAPASPILIOPOULOS, O. (2020). *An introduction to sequential Monte Carlo*. *Springer Series in Statistics*. Springer, Cham.

[6] CHOPIN, N., SINGH, S. S., SOTO, T. and VIHOLA, M. (2022). Supplement to: "On resampling schemes for particle filters with weakly informative observations".

[7] CRISAN, D., DEL MORAL, P. and LYONS, T. (1999). Discrete filtering using branching and interacting particle systems. *Markov Process. Related Fields* **5** 293–318.

[8] DEL MORAL, P. (2004). *Feynman-Kac Formulae*. Springer.

[9] DEL MORAL, P. (2013). *Mean field simulation for Monte Carlo integration*. Chapman and Hall/CRC.

[10] DEL MORAL, P., JACOB, P. E., LEE, A., MURRAY, L. and PETERS, G. W. (2013). Feynman-Kac particle integration with geometric interacting jumps. *Stoch. Anal. Appl.* **31** 830–871. https://doi.org/10.1080/07362994.2013.817247

[11] DEL MORAL, P. and MICLO, L. (2000). A Moran particle system approximation of Feynman–Kac formulae. *Stoch. Proc. Appl.* **86** 193–216.

[12] DEL MORAL, P. and MICLO, L. (2000). Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering. In *Séminaire de probabilités XXXIV* 1–145. Springer.

[13] DOUC, R. and CAPPÉ, O. (2005). Comparison of resampling schemes for particle filtering. In *Proc. ISPA 2005* 64–69.

[14] ETHIER, S. and KURTZ, T. (1986). *Markov processes. Characterization and convergence*. John Wiley & Sons, Inc., New York.

[15] FLEGAL, J. M. and JONES, G. L. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *Ann. Statist.* **38** 1034–1070.

[16] GERBER, M., CHOPIN, N. and WHITELEY, N. (2019). Negative association, ordering and convergence of resampling methods. *Ann. Statist.* **47** 2236–2260.

[17] GLYNN, P. W. and WHITT, W. (1992). The asymptotic efficiency of simulation estimators. *Oper. Res.* **40** 505–520.

[18] GORDON, N. J., SALMOND, D. J. and SMITH, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F* **140** 107–113.

[19] HAARIO, H., SAKSMAN, E. and TAMMINEN, J. (2001). An Adaptive Metropolis Algorithm. *Bernoulli* **7** 223–242.

[20] HIGUCHI, T. (1997). Monte Carlo filter using the genetic algorithm operators. *J. Statist. Comput. Simulation* **59** 1–23.

[21] HOARE, C. A. (1962). Quicksort. *The Computer Journal* **5** 10–16.

[22] LIU, J. S. and CHEN, R. (1995). Blind deconvolution via sequential imputations. *J. Amer. Statist. Assoc.* **90** 567–576.

[23] LIU, J. S. and CHEN, R. (1998). Sequential Monte Carlo methods for dynamic systems. *J. Amer. Statist. Assoc.* **93** 1032–1044.

[24] MURRAY, L. M., LEE, A. and JACOB, P. E. (2016). Parallel resampling in the particle filter. *J. Comput. Graph. Statist.* **25** 789–805.

[25] ROUSSET, M. (2006). On the control of an interacting particle estimation of Schrödinger ground states. *SIAM J. Math. Anal.* **38** 824–844. https://doi.org/10.1137/050640667

[26] VIHOLA, M. (2020). Ergonomic and reliable Bayesian inference with adaptive Markov chain Monte Carlo. In *Wiley StatsRef: Statistics Reference Online* (M. Davidian, R. S. Kenett, N. T. Longford, G. Molenberghs, W. W. Piegorsch and F. Ruggeri, eds.) **stat08286** Wiley.

[27] VIHOLA, M., HELSKE, J. and FRANKS, J. (2020). Importance sampling type estimators based on approximate marginal MCMC. *Scand. J. Stat.* **47** 1339–1376.

[28] WHITELEY, N., LEE, A. and HEINE, K. (2016). On the role of interaction in sequential Monte Carlo algorithms. *Bernoulli* **22** 494–529.