Check for updates

Adv. Appl. Probab. 1–40 (2024)
*doi:*10.1017/apr.2024.12

ANTITHETIC MULTILEVEL PARTICLE FILTERS

AJAY JASRA,* *Chinese University of Hong Kong, Shenzhen*
MOHAMED MAAMA** *** AND
HERNANDO OMBAO,** **** *King Abdullah University of Science and Technology*

Abstract

In this paper we consider the filtering of partially observed multidimensional diffusion processes that are observed regularly at discrete times. This is a challenging problem which requires the use of advanced numerical schemes based upon time-discretization of the diffusion process and then the application of particle filters. Perhaps the state-of-the-art method for moderate-dimensional problems is the multilevel particle filter of Jasra *et al.* (*SIAM J. Numer. Anal.* **55** (2017), 3068–3096). This is a method that combines multilevel Monte Carlo and particle filters. The approach in that article is based intrinsically upon an Euler discretization method. We develop a new particle filter based upon the antithetic truncated Milstein scheme of Giles and Szpruch (*Ann. Appl. Prob.* **24** (2014), 1585–1620). We show empirically for a class of diffusion problems that, for $\epsilon > 0$ given, the cost to produce a mean squared error (MSE) of $\mathcal{O}(\epsilon^2)$ in the estimation of the filter is $\mathcal{O}(\epsilon^{-2}\log(\epsilon)^2)$. In the case of multidimensional diffusions with non-constant diffusion coefficient, the method of Jasra *et al.* (2017) requires a cost of $\mathcal{O}(\epsilon^{-2.5})$ to achieve the same MSE.

Keywords: Multilevel Monte Carlo; particle filters; diffusion processes; filtering

2020 Mathematics Subject Classification: Primary 65C05
Secondary 60G05

1. Introduction

We are given a diffusion process

$$dX_t = \alpha(X_t)dt + \beta(X_t)dW_t, \tag{1.1}$$

where $X_0 = x_0 \in \mathbb{R}^d$ is given, $\alpha : \mathbb{R}^d \to \mathbb{R}^d$, $\beta : \mathbb{R}^d \to \mathbb{R}^{d\times d}$, and $\{W_t\}_{t\geq 0}$ is a standard d-dimensional Brownian motion. We consider the problem where this is a latent process and we observe it only through a sequence of data that are discrete and regular in time, and in particular where the structure of the observations is a special case of the structure in a state-space or hidden Markov model (HMM). More specifically, we suppose that data are observed

Received 5 June 2023; accepted 26 January 2024.
* Postal address: School of Data Science, Chinese University of Hong Kong, Shenzhen, China. Email address: ajayjasra@cuhk.edu.cn
** Postal address: King Abdullah University of Science and Technology, Computer, Electrical and Mathematical Sciences and Engineering, Thuwal 23955-6900, Saudi Arabia.
*** Email address: maama.mohamed@gmail.com
**** Email address: hernando.ombao@kaust.edu.sa

© The Author(s), 2024. Published by Cambridge University Press on behalf of Applied Probability Trust. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

at regular and unit times, forming a sequence of random variables $(Y_1, Y_2, \ldots)$. In particular, at time k we assume that, conditioning on all other random variables in the sequence and given the diffusion process $\{X_t\}_{t\geq 0}$, Y_k has a probability density function that depends on x_k only and is denoted by $g(x_k, y_k)$. These models have a wide class of applications from finance, econometrics, and engineering; see [3, 5, 6, 12, 22] for some particular examples in various contexts.

In this paper we consider the problems of filtering and normalizing constant (marginal likelihood) estimation—that is, the recursive-in-time (k is the time index) computation of expectations with respect to the conditional distribution of $X_k|y_1, \ldots, y_k$ (filtering) and the calculation of the associated marginal likelihood of the data $(y_1, \ldots, y_k)$. For standard HMMs, that is, where the latent process is a discrete-time Markov chain, this is a notoriously challenging problem, requiring the application of advanced numerical (Monte Carlo) methods such as the particle filter; see e.g. [5, 6] for a survey. The particle filter simulates a collection of N samples in parallel, with the samples undergoing transitions such as sampling and resampling. The first of these uses the hidden Markov chain and the resampling samples with replacement amongst the collection samples using a weight (proportional to $g(x_k, y_k)$). The scenario which we consider is even more challenging, as typically the transition density associated to (1.1), assuming it exists, is intractable. This can limit the applicability of particle filters; although there are some exceptions [8] and exact simulation methods [1, 4], these are often not general enough or too expensive to be of practical use in the filtering problem. As a result, we focus on the case where one discretizes the process (1.1) in time.

In recent years, one of the most successful methods for improving Monte-Carlo-based estimators associated to probability laws under time-discretization has been the multilevel Monte Carlo (MLMC) method [9, 10, 13]. This is an approach that considers a collapsing-sum representation of an expectation with respect to a probability law at a given level of discretization. The collapsing element is associated to differences in expectations with increasingly coarse discretization levels, with a final (single) expectation at a course level. Then, if one can sample appropriate couplings of the probability laws at consecutive levels, it is possible to reduce the cost to achieve a certain mean squared error (MSE) in several examples, diffusions being one of them. This method has been combined with the particle filter in several articles, resulting in the multilevel particle filter (MLPF); see [14, 15, 16], as well as [17] for a review and [18] for extensions.

The method of [14] is intrinsically based upon the ability to sample couplings of discretized diffusion processes. In almost all of the applications that we are aware of, this is based upon the synchronous coupling of Brownian motion for an Euler or Milstein scheme. These particular couplings inherit properties of the strong error of the associated time-discretization. The importance of the strong error rate is that it can help determine the efficiency gain of any MLMC approach, of which the MLPF is one. As is well known, in dimensions larger than two ($d > 2$), the Milstein scheme, which is of higher (strong) order than the Euler method, can be difficult to implement numerically because of the need to simulate Lévy areas. Giles and Szpruch [11] consider computing expectations associated to the law of (1.1) at a given terminal time. They show that by eliminating the Lévy area and including an antithetic-type modification of the traditional MLMC estimator, one can maintain the strong error rate of the Milstein scheme even in a multidimensional setting. Moreover, the cost of the simulation is of the same order as for Euler discretizations, which are the ones most often used in multidimensional cases.

We develop a new particle filter based upon the antithetic truncated Milstein scheme of [11]. We show empirically for a class of diffusion problems that, for $\epsilon > 0$ given, the cost to produce

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

an MSE of $\mathcal{O}(\epsilon^2)$ in the estimation of the filter is $\mathcal{O}\big(\epsilon^{-2}\log(\epsilon)^2\big)$. In the case of multidimensional diffusions with non-constant diffusion coefficient, the method of [14] requires a cost of $\mathcal{O}\big(\epsilon^{-2.5}\big)$ to achieve the same MSE. We also show how this new particle filter can be used to compute the normalizing constant recursively in time, but we do not prove anything about the efficiency gains. Our theoretical results are not optimal and show only that the cost of the new method is $\mathcal{O}\big(\epsilon^{-2.5}\big)$ for the MSE of $\mathcal{O}(\epsilon^2)$. We highlight exactly where the proof loses the rate needed.

To provide a more detailed summary of the contribution of this paper, we state the following. First, the main point is to understand how the rather clever discretization scheme of [11] can be leveraged in the context of filtering and the multilevel method. Generally, for ordinary discretization schemes in MLMC, one has to simulate coupled samples at two levels of time-discretization, one that is fine and another that is coarse. Then this can be extended to the case of filtering diffusion processes, as was done in [14] by constructing a coupling in the resampling mechanism of particle filters. In the method of [11], the authors sample a third 'antithetic' process at the finer level. We show how such a process can be incorporated into the context of particle filters with, to the best of our knowledge, a new resampling scheme. Second, there are significant technical challenges in showing that this new antithetic particle filter can provide the improvements in terms of cost and MSE that the approach of [11] does for ordinary diffusion processes. Unfortunately, we obtain a non-optimal rate; again, we highlight where the proof misses the extra rate.

The article is structured as follows. In Section 2 we give details on the model to be considered and the time-discretization associated to the process (1.1). In Section 3 we present our algorithm. Section 4 details our mathematical results and their algorithmic implications. In Section 5 we give numerical results that support our theory. Most of our mathematical proofs can be found in the appendices at the end of the article.

2. Model and discretization

2.1. State-space model

Our objective is to consider the filtering problem and normalizing constant estimation for a specific class of state-space models associated to the diffusion process (1.1). In particular, we will assume that (1.1) is subject to a certain assumption (A1) which is described in Section 4. This assumption is certainly strong enough to guarantee that (1.1) has a unique solution and in addition that the diffusion has a transition probability, which we denote, over one unit of time, by $P(x, dx')$. We are then interested in filtering associated to the state-space model

$$p(dx_{1:n}, y_{1:n}) = \prod_{k=1}^{n} P(x_{k-1}, dx_k)g(x_k, y_k),$$

where $y_{1:n} = (y_1, \ldots, y_n)^\top \in \mathsf{Y}^n$ are observations, each with conditional density $g(x_k, \cdot)$. The filter associated to this measure is, for $k \in \mathbb{N}$,

$$\pi_k(dx_k) = \frac{\int_{\mathsf{X}^{k-1}} p(dx_{1:k}, y_{1:k})}{\int_{\mathsf{X}^k} p(dx_{1:k}, y_{1:k})},$$

where $\mathsf{X} = \mathbb{R}^d$. The denominator

$$p(y_{1:k}) := \int_{\mathsf{X}^k} p(dx_{1:k}, y_{1:k})$$

Algorithm 1 Truncated Milstein scheme on [0, 1].

1. Input level l and starting point x_0^l.
2. Generate $Z_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}_d(0, \Delta_l I_d)$, $k \in \{1, 2, \ldots, \Delta_l^{-1}\}$.
3. Generate level l: for $k \in \{0, 1, \ldots, \Delta_l^{-1} - 1\}$ with $X_0^l = x_0^l$
$$X^l_{(k+1)\Delta_l} = X^l_{k\Delta_l} + \alpha(X^l_{k\Delta_l})\Delta_l + \beta(X^l_{k\Delta_l})Z_{k+1} + H_{\Delta_l}(X^l_{k\Delta_l}, Z_{k+1}).$$
4. Output X_1^l.

is the normalizing constant or marginal likelihood. This latter object is often used in statistics for model selection.

In practice, we assume that working directly with P is not possible, because of either intractability or cost of simulation. We therefore propose to work with an alternative collection of filters based upon time-discretization, which is what we now describe.

2.2. Time-discretization

Typically one must time-discretize (1.1). We consider a time-discretization at equally spaced times, separated by $\Delta_l = 2^{-l}$. To continue with our exposition, we define the d-vector $H : \mathbb{R}^{2d} \times \mathbb{R}^+ \to \mathbb{R}^d$, $H_\Delta(x, z) = (H_{\Delta,1}(x, z), \ldots, H_{\Delta,d}(x, z))^\top$, where, for $i \in \{1, \ldots, d\}$,

$$H_{\Delta,i}(x, z) = \sum_{(j,k)\in\{1,\ldots,d\}^2} h_{ijk}(x)(z_j z_k - \Delta),$$
$$h_{ijk}(x) = \frac{1}{2} \sum_{m\in\{1,\ldots,d\}} \beta_{mk}(x) \frac{\partial \beta_{ij}(x)}{\partial x_m}.$$

We denote by $\mathcal{N}_d(\mu, \Sigma)$ the d-dimensional Gaussian distribution with mean vector μ and covariance matrix Σ; if $d = 1$ we drop the subscript d. I_d is the $d \times d$ identity matrix. A single-level version of the truncated Milstein scheme, which is the focus of this article, is presented in Algorithm 1. Ultimately, we will consider the antithetic truncated Milstein scheme in [11], which can be simulated as described in Algorithm 2. The method in Algorithm 2 is a means of approximating differences of expectations with respect to discretized laws at consecutive levels, as is used in MLMC; this latter approach will be detailed in Section 3.

To understand the use of Algorithms 1 and 2, consider computing the expectation

$$\int_{\mathsf{X}} x_1 P(x_0, dx_1),$$

which is generally approximated by considering $\int_{\mathsf{X}} x_1 P^l(x_0, dx_1)$, where we denote the transition kernel induced by Algorithm 1 by $P^l(x_0, dx_1)$. This latter integral can certainly be approximated using Monte Carlo and Algorithm 1. However, there is an alternative using both Algorithms 1 and 2, which has been established by [11] as more efficient. Clearly one can write

$$\int_{\mathsf{X}} x_1 P^l(x_0, dx_1) = \int_{\mathsf{X}} x_1 P^l(x_0, dx_1) - \int_{\mathsf{X}} x_1 P^{l-1}(x_0, dx_1) + \int_{\mathsf{X}} x_1 P^{l-1}(x_0, dx_1).$$

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

Algorithm 2 Antithetic truncated Milstein scheme on [0, 1].

1. Input level l and starting points $\left(x_0^l, x_0^{l-1}, x_0^{l,a}\right)$.
2. Generate $Z_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}_d(0, \Delta_l I_d)$, $k \in \left\{1, 2, \ldots, \Delta_l^{-1}\right\}$.
3. Generate level l: for $k \in \left\{0, 1, \ldots, \Delta_l^{-1} - 1\right\}$ with $X_0^l = x_0^l$
$$X_{(k+1)\Delta_l}^l = X_{k\Delta_l}^l + \alpha\left(X_{k\Delta_l}^l\right)\Delta_l + \beta\left(X_{k\Delta_l}^l\right)Z_{k+1} + H_{\Delta_l}\left(X_{k\Delta_l}^l, Z_{k+1}\right).$$
4. Generate level $l-1$: for $k \in \left\{0, 1, \ldots, \Delta_{l-1}^{-1} - 1\right\}$ with $X_0^{l-1} = x_0^{l-1}$
$$X_{(k+1)\Delta_{l-1}}^{l-1} = X_{k\Delta_{l-1}}^{l-1} + \alpha\left(X_{k\Delta_{l-1}}^{l-1}\right)\Delta_{l-1} + \beta\left(X_{k\Delta_{l-1}}^{l-1}\right)\left\{Z_{2(k+1)-1} + Z_{2(k+1)}\right\} + H_{\Delta_{l-1}}\left(X_{k\Delta_{l-1}}^{l-1}, Z_{2(k+1)-1} + Z_{2(k+1)}\right).$$
5. Generate antithetic level l: for $k \in \left\{0, 1, \ldots, \Delta_l^{-1} - 1\right\}$ with $X_0^{l,a} = x_0^{l,a}$
$$X_{(k+1)\Delta_l}^{l,a} = X_{k\Delta_l}^{l,a} + \alpha\left(X_{k\Delta_l}^{l,a}\right)\Delta_l + \beta\left(X_{k\Delta_l}^{l,a}\right)Z_{\rho_k} + H_{\Delta_l}\left(X_{k\Delta_l}^{l,a}, Z_{\rho_k}\right)$$
where $\rho_k = k + 2\mathbb{I}_{\{0,2,4,\ldots\}}(k)$.
6. Output $\left(X_1^l, X_1^{l-1}, X_1^{l,a}\right)$.

Now the rightmost term on the right-hand side of the above equation can be approximated by Monte Carlo and running Algorithm 1 at level $l-1$. We then consider the difference on the right-hand side, independently, which can be approximated using Algorithm 2 N times:

$$\frac{1}{N}\sum_{i=1}^{N}\left(\tfrac{1}{2}\left\{X_1^{i,l} + X_1^{i,l,a}\right\} - X_1^{i,l-1}\right),$$

where the superscript i denotes the ith sample generated. The reason such an approximation works so well is that marginally $X_1^{i,l-1}$ has the same distribution as any sample generated by running Algorithm 1 at level $l-1$; similarly, both $X_1^{i,l}$ and $X_1^{i,l,a}$ have the same distribution as any sample generated by running Algorithm 1 at level l. Moreover, the second moment of the term, which converges to zero as l increases,

$$\left\|\tfrac{1}{2}\left\{X_1^{i,l} + X_1^{i,l,a}\right\} - X_1^{i,l-1}\right\|, \tag{2.1}$$

where $\|\cdot\|$ is the usual Euclidean norm, is of the *same* order (as a function of l) as if one considered a synchronous coupling of two discretizations of the exact Milstein scheme—that is, if one could sample an exact Milstein discretization at level l $\left(X_1^{l,M}\right)$ and using the same Brownian motion as at level $l-1$ $\left(X_1^{l-1,M}\right)$, and consider the second moment of $\left\|X_1^{l,M} - X_1^{l-1,M}\right\|$. The reason why this is so important is that the rate of decrease of the aforementioned quantities is what enables MLMC (here we have described two levels) to improve over ordinary Monte Carlo. The key is that by using antithetic variates of the quantity (2.1) (in terms of square expectation) one has replicated the Milstein scheme without ever having to compute intractable Lévy areas.

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

For a given $l \in \mathbb{N}$, we will be concerned with the filter induced by the following joint measure:

$$p^l(dx_{1:n}, y_{1:n}) = \prod_{k=1}^{n} P^l(x_{k-1}, dx_k) g(x_k, y_k).$$

The filter associated to this measure is, for $k \in \mathbb{N}$,

$$\pi_k^l(dx_k) = \frac{\int_{\mathsf{X}^{k-1}} p^l(dx_{1:k}, y_{1:k})}{\int_{\mathsf{X}^k} p^l(dx_{1:k}, y_{1:k})}.$$

The associated normalizing constant is

$$p^l(y_{1:k}) = \int_{\mathsf{X}^k} p^l(dx_{1:k}, y_{1:k}).$$

Ultimately, we will seek to approximate expectations with respect to π_k^l and to estimate $p^l(y_{1:k})$ as k increases, then study the MSE of a numerical approximation relative to considering the exact filter π_k and associated normalizing constant $p(y_{1:k})$.

3. Antithetic multilevel particle filters

3.1. Multilevel particle filter

Let $\varphi \in \mathcal{B}_b(\mathsf{X})$, the latter denoting the collection of bounded and measurable real-valued functions; we write $\pi_k^l(\varphi) = \int_{\mathsf{X}} \varphi(x_k)\pi_k^l(dx_k)$. Let $(\underline{L}, \overline{L}) \in \mathbb{N}_0 \times \mathbb{N}$, with $\underline{L} < \overline{L}$ given. Our objective is the approximation of $\pi_k^{\overline{L}}(\varphi)$ sequentially in time; this is the filtering problem. This can be achieved using the MLMC identity

$$\pi_k^{\overline{L}}(\varphi) = \pi_k^{\underline{L}}(\varphi) + \sum_{l=\underline{L}+1}^{\overline{L}} \left[\pi_k^l - \pi_k^{l-1}\right](\varphi). \tag{3.1}$$

As noted in the introduction, the computational cost of approximating the right-hand side of (3.1) can be lower than that of approximating the left-hand side, when one is seeking to achieve a pre-specified MSE. The right-hand side of (3.1) can be approximated by the method in [14], but we shall consider a modification which can improve on the computational effort to achieve a given MSE. Often $\overline{L} = 1$ in the literature, but we have found in previous work that one needs to set it higher for the aforementioned improvement in computational effort.

We begin by approximating $\pi_k^{\underline{L}}(\varphi)$, which can be done using the standard particle filter as described in Algorithm 3. Algorithm 3 generates $N_{\underline{L}}$ samples in parallel, and these samples undergo sampling operations (Steps 1 and 3) and resampling operations (Step 2). The sampling step uses Algorithm 1, and the resampling step is the well-known multinomial method. We refer the reader to e.g. [6] for an introduction to particle filter methods. It is by now standard in the literature that

$$\pi_k^{N_{\underline{L}},\underline{L}}(\varphi) := \sum_{i=1}^{N_{\underline{L}}} \frac{g\big(X_k^{i,\underline{L}}, y_k\big)}{\sum_{j=1}^{N_l} g\big(X_k^{j,\underline{L}}, y_k\big)} \varphi\big(X_k^{i,\underline{L}}\big)$$

will converge almost surely to $\pi_k^{\underline{L}}(\varphi)$, where the samples $X_k^{i,\underline{L}}$ are after Step 1 or 3 of Algorithm 3. Therefore, recalling (3.1), at this stage we can approximate $\pi_k^{\underline{L}}(\varphi)$ recursively in time.

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

---

Algorithm 3 Particle filter at level $\underline{L}$.

---

1. Initialization: for $i \in \{1, \ldots, N_{\underline{L}}\}$, generate $X_1^{i,\underline{L}}$ independently using Algorithm 1 with level $\underline{L}$ and starting point x_0. Set $k = 1$.
2. Resampling: compute

$$\left(\frac{g\big(x_k^{1,\underline{L}}, y_k\big)}{\sum_{j=1}^{N_{\underline{L}}} g\big(x_k^{j,\underline{L}}, y_k\big)}, \ldots, \frac{g\big(x_k^{N_{\underline{L}},\underline{L}}, y_k\big)}{\sum_{j=1}^{N_{\underline{L}}} g\big(x_k^{j,\underline{L}}, y_k\big)}\right). \tag{3.2}$$

   For $i \in \{1, \ldots, N_{\underline{L}}\}$ generate an index a_k^i using the probability mass function in (3.2) and set $\tilde{X}_k^{i,\underline{L}} = X_k^{a_k^i,\underline{L}}$. Then set $X_k^{1:N_{\underline{L}},\underline{L}} = \tilde{X}_k^{1:N_{\underline{L}},\underline{L}}$.
3. Sampling: for $i \in \{1, \ldots, N_{\underline{L}}\}$, generate $X_{k+1}^{i,\underline{L}}|X_k^{1:N_{\underline{L}},\underline{L}}$ conditionally independently using Algorithm 1 with level $\underline{L}$ and starting point $x_k^{i,\underline{L}}$. Set $k = k + 1$.

In reference to (3.1), we will approximate the differences $\big[\pi_k^l - \pi_k^{l-1}\big](\varphi)$ sequentially in time and independently of using Algorithm 3 and for each l. As stated in the introduction, our objective is to leverage the work of [11] as described in Algorithm 2 and be able to insert this within a particle filter as given in Algorithm 3. The particle filter has two main operations, which are sampling and resampling. The sampling mechanism is to be achieved by using Algorithm 2. The problem is now the resampling approach. The main idea is to ensure that we resample the level-l coordinate just as if we had done so using the particle filter at level l; similarly, we want the same for the antithetic-level coordinate and finally we resample the level-$(l-1)$ coordinate just as if we had used the particle filter at level $l-1$. Critically, the resampling for these three coordinates should be correlated, because if this is not the case, the effect of using Algorithm 2 is lost; see [19] for a theoretical justification of this point. The reason we want to maintain the resampling as if the particle filter had been used is that we know this will produce consistent estimates of the filter. We seek to correlate the resampling because this can help ensure that we reduce the cost of achieving a pre-specified MSE, relative to using the particle filter.

Algorithm 4 presents a new resampling method which we shall employ. This resampling method achieves the objectives discussed above and gives rise to the new coupled particle filter, which is described in Algorithm 5. Algorithm 5 consists of propagating N_l samples of the three coordinates sequentially in time, using Algorithm 2 for the sampling and Algorithm 4 for the resampling. In this context one can define, for $(k, l, N_l, \varphi) \in \mathbb{N}^3 \times \mathcal{B}_b(\mathsf{X})$,

$$\pi_k^{N_l,l}(\varphi) := \sum_{i=1}^{N_l} W_k^{i,l}\varphi\big(X_k^{i,l}\big), \tag{3.3}$$

$$\bar{\pi}_k^{N_l,l-1}(\varphi) := \sum_{i=1}^{N_l} \bar{W}_k^{i,l-1}\varphi\big(X_k^{i,l-1}\big),$$

$$\pi_k^{N_l,l,a}(\varphi) := \sum_{i=1}^{N_l} W_k^{i,l,a}\varphi\big(X_k^{i,l,a}\big),$$

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

Algorithm 4 Maximal coupling-type resampling.

1. Input $\left(U_1^{1:N}, U_2^{1:N}, U_3^{1:N}\right)$ and probabilities $\left(W_1^{1:N}, W_2^{1:N}, W_3^{1:N}\right)$.
2. For $i \in \{1, \ldots, N\}$ generate $U \sim \mathcal{U}_{[0,1]}$ (uniform distribution on [0, 1]):
   - If $U < \sum_{i=1}^N \min\left\{W_1^i, W_2^i, W_3^i\right\}$ generate $a^i \in \{1, \ldots, N\}$ using the probability mass function
   $$\mathbb{P}(i) = \frac{\min\left\{W_1^i, W_2^i, W_3^i\right\}}{\sum_{j=1}^N \min\left\{W_1^j, W_2^j, W_3^j\right\}}$$
   and set $\tilde{U}_j^i = U_j^{a^i}, j \in \{1, 2, 3\}$.
   - Otherwise generate $\left(a_1^i, a_2^i, a_3^i\right) \in \{1, \ldots, N\}^3$ using any coupling of the probability mass functions
   $$\mathbb{P}_j(i) = \frac{W_j^i - \min\left\{W_1^i, W_2^i, W_3^i\right\}}{\sum_{k=1}^N \left[W_j^k - \min\left\{W_1^k, W_2^k, W_3^k\right\}\right]},$$
   and set $\tilde{U}_j^i = U_j^{a_j^i}, j \in \{1, 2, 3\}$. In all of our implementations, $\left(a_1^i, a_2^i, a_3^i\right)$ are generated independently with the given marginals.
3. Set $U_j^i = \tilde{U}_j^i, (i, j) \in \{1, \ldots, N\} \times \{1, 2, 3\}$.
4. Output $\left(U_1^{1:N}, U_2^{1:N}, U_3^{1:N}\right)$.

and estimate $\left[\pi_k^l - \pi_k^{l-1}\right](\varphi)$ as

$$\left[\pi_k^l - \pi_k^{l-1}\right]^{N_l}(\varphi) := \frac{1}{2}\left\{\pi_k^{N_l,l}(\varphi) + \pi_k^{N_l,l,a}(\varphi)\right\} - \bar{\pi}_k^{N_l,l-1}(\varphi).$$

The samples in $\left(\pi_k^{N_l,l}(\varphi), \pi_k^{N_l,l,a}(\varphi), \bar{\pi}_k^{N_l,l-1}(\varphi)\right)$ are obtained after Step 1 or Step 3 in Algorithm 5.

3.2. Final algorithm

To conclude, the antithetic multilevel particle filter (AMLPF) is as follows:

1. At level $\underline{L}$ run Algorithm 3 sequentially in time.
2. At level $l \in \{\underline{L}+1, \underline{L}+2, \ldots, \overline{L}\}$, independently of all other l and of Step 1, run Algorithm 3 sequentially in time.

To approximate the right-hand side of (3.1) at any time k, one can use the estimator

$$\pi_k^{\overline{L},ML}(\varphi) := \pi_k^{N_{\underline{L}},\underline{L}}(\varphi) + \sum_{l=\underline{L}+1}^{\overline{L}} \left[\pi_k^l - \pi_k^{l-1}\right]^{N_l}(\varphi),$$

where $\pi_k^{N_{\underline{L}},\underline{L}}(\varphi)$ has the definition (3.3) when $l = \underline{L}$ and has been obtained using Algorithm 3.

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

---

Algorithm 5 A new coupled particle filter for $l \in \mathbb{N}$ given.

---

1. Initialization: for $i \in \{1, \ldots, N_l\}$, generate $U_1^{i,l} = \left(X_1^{i,l}, \bar{X}_1^{i,l-1}, X_1^{i,l,a}\right)$ independently using Algorithm 2 with level l and starting points $\left(x_0^l, \bar{x}_0^{l-1}, x_0^{l,a}\right)$. Set $k = 1$.
2. Coupled resampling: compute

$$W_k^{1:N_l,l} = \left(\frac{g\left(x_k^{1,l}, y_k\right)}{\sum_{j=1}^{N_l} g\left(x_k^{j,l}, y_k\right)}, \ldots, \frac{g\left(x_k^{N_l,l}, y_k\right)}{\sum_{j=1}^{N_l} g\left(x_k^{j,l}, y_k\right)} \right),$$

$$\bar{W}_k^{1:N_l,l-1} = \left(\frac{g\left(\bar{x}_k^{1,l-1}, y_k\right)}{\sum_{j=1}^{N_l} g\left(\bar{x}_k^{j,l-1}, y_k\right)}, \ldots, \frac{g\left(\bar{x}_k^{N_l,l-1}, y_k\right)}{\sum_{j=1}^{N_l} g\left(\bar{x}_k^{j,l-1}, y_k\right)} \right),$$

$$W_k^{1:N_l,l,a} = \left(\frac{g\left(x_k^{1,l,a}, y_k\right)}{\sum_{j=1}^{N_l} g\left(x_k^{j,l,a}, y_k\right)}, \ldots, \frac{g\left(x_k^{N_l,l,a}, y_k\right)}{\sum_{j=1}^{N_l} g\left(x_k^{j,l,a}, y_k\right)} \right).$$

   Then, using $\left(X_k^{1:N_l,l}, \bar{X}_k^{1:N_l,l-1}, X_k^{1:N_l,l,a}\right)$ and $\left(W_k^{1:N_l,l}, \bar{W}_k^{1:N_l,l-1}, W_k^{1:N_l,l,a}\right)$, call Algorithm 4.
3. Coupled sampling: for $i \in \{1, \ldots, N_l\}$, generate $U_{k+1}^{i,l} | U_k^{1:N_l,l}$ conditionally independently using Algorithm 2 with level l and starting point $\left(x_k^{i,l}, \bar{x}_k^{i,l-1}, x_k^{i,l,a}\right)$. Set $k = k + 1$.

The normalizing constant $p^{\overline{L}}(y_{1:k})$ can also be approximated using the AMLPF. For any $(k, l, N_l, \varphi) \in \mathbb{N}^3 \times \mathcal{B}_b(\mathsf{X})$, we define

$$\eta_k^{N_l,l}(\varphi) := \sum_{i=1}^{N_l} \varphi\left(X_k^{i,l}\right), \tag{3.4}$$

$$\bar{\eta}_k^{N_l,l-1}(\varphi) := \sum_{i=1}^{N_l} \varphi\left(X_k^{i,l-1}\right),$$

$$\eta_k^{N_l,l,a}(\varphi) := \sum_{i=1}^{N_l} \varphi\left(X_k^{i,l,a}\right).$$

Then we have the following approximation:

$$p^{\overline{L},ML}(y_{1:k}) := \prod_{j=1}^{k} \eta_j^{N_{\underline{L}},\underline{L}}(g_j) + \sum_{l=\underline{L}+1}^{\overline{L}} \left\{ \frac{1}{2} \prod_{j=1}^{k} \eta_j^{N_l,l,}(g_j) + \frac{1}{2} \prod_{j=1}^{k} \eta_j^{N_l,l,a}(g_j) - \prod_{j=1}^{k} \bar{\eta}_j^{N_{l-1},l-1,}(g_j) \right\}$$

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

where $g_j(x) = g(x, y_j)$ is used as shorthand, g_j is assumed bounded and measurable for each $j \in \mathbb{N}$, and $\eta_k^{N_{\underline{L}},\underline{L}}(\varphi)$ has the definition (3.4) when $l = \underline{L}$ and has been obtained using Algorithm 3. This new estimator approximates the quantity

$$p^{\underline{L}}(y_{1:k}) + \sum_{l=\underline{L}+1}^{\overline{L}} \left[p^l(y_{1:k}) - p^{l-1}(y_{1:k})\right]$$

as was done in [15] and essentially uses the ordinary particle filter normalizing constant estimator to approximate $p^{\underline{L}}(y_{1:k})$ (see e.g. [6]). The approximation of the increment $\left[p^l(y_{1:k}) - p^{l-1}(y_{1:k})\right]$ follows the approach used for the filter based upon the output from the AMLPF. This estimator is unbiased in terms of $p^{\overline{L}}(y_{1:k})$ (see e.g. [6]) and is consistent as the number of samples used at each level grows. We note, however, that there is nothing in the estimator that ensures that it is almost surely positive, whereas it does approximate a positive quantity.

4. Mathematical results

We now present our main mathematical result, which relates to the $\mathbb{L}_2$-convergence of our estimator of the filter. We remark that we will consider the estimation of the normalizing constant in a sequel paper. The result is proved under two assumptions (A1) and (A2), which are stated below. We write $\mathsf{X}_2 = \mathbb{R}^{d \times d}$. By $\mathcal{C}_b^2(\mathsf{X}, \mathbb{R})$ we denote the collection of twice continuously differentiable functions from X to $\mathbb{R}$ (resp. X_2) with bounded derivatives of orders 1 and 2. Similarly, $\mathcal{C}^2(\mathsf{X}, \mathsf{X})$ and $\mathcal{C}^2(\mathsf{X}, \mathsf{X}_2)$ are respectively the collections of twice continuously differentiable functions from X to X and from X to X_2. For $\varphi \in \mathcal{B}_b(\mathsf{X})$ we use the notation $\pi_k^{\overline{L}}(\varphi) = \int_{\mathsf{X}} \varphi(x)\pi_k^{\overline{L}}(dx)$. By $\mathbb{E}$ we denote the expectation with respect to the law used to generate the AMLPF. We drop the dependence upon the data in $g(x_k, y_k)$ and simply write $g_k(x_k)$ in this section.

Our assumptions are as follows:

(A1)
- For each $(i, j) \in \{1, \ldots, d\}$, $\alpha_i \in \mathcal{B}_b(\mathsf{X})$, $\beta_{ij} \in \mathcal{B}_b(\mathsf{X})$.
- We have $\alpha \in \mathcal{C}^2(\mathsf{X}, \mathsf{X})$, $\beta \in \mathcal{C}^2(\mathsf{X}, \mathsf{X}_2)$.
- We have that $\beta(x)\beta(x)^\top$ is uniformly positive definite.
- There exists a $C < +\infty$ such that for any $(x, i, j, k, m) \in \mathsf{X} \times \{1, \ldots, d\}^4$,

$$\max\left\{\left|\frac{\partial \alpha_i}{\partial x_m}(x)\right|, \left|\frac{\partial \beta_{ij}}{\partial x_m}(x)\right|, \left|\frac{\partial h_{ijk}}{\partial x_m}(x)\right|, \left|\frac{\partial^2 \alpha_i}{\partial x_k \partial x_m}(x)\right|, \left|\frac{\partial^2 \beta_{ij}}{\partial x_k \partial x_m}(x)\right|\right\} \le C.$$

(A2)
- For each $k \in \mathbb{N}$, $g_k \in \mathcal{B}_b(\mathsf{X}) \cap \mathcal{C}^2(\mathsf{X}, \mathbb{R})$.
- For each $k \in \mathbb{N}$ there exists a $0 < C < +\infty$ such that for any $x \in \mathsf{X}$ $g_k(x) \ge C$.
- For each $k \in \mathbb{N}$ there exists a $0 < C < +\infty$ such that for any $(x, j, m) \in \mathsf{X} \times \{1, \ldots, d\}^2$,

$$\max\left\{\left|\frac{\partial g_k}{\partial x_j}(x)\right|, \left|\frac{\partial^2 g_k}{\partial x_j \partial x_m}(x)\right|\right\} \le C.$$

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

Theorem 1. *Assume* (A1)–(A2). *Then, for any* $(k, \varphi) \in \mathbb{N} \times \mathcal{B}_b(\mathsf{X}) \cap \mathcal{C}_b^2(\mathsf{X}, \mathbb{R})$*, there exists a* $C < +\infty$ *such that for any* $\left(\underline{L}, \overline{L}, l, N_{\underline{L}}, \ldots, N_{\overline{L}}\right) \in \mathbb{N}_0^2 \times \{\underline{L}, \ldots, \overline{L}\} \times \mathbb{N}^{\overline{L}-\underline{L}+1}$ *with* $\underline{L} < \overline{L}$*,*

$$\mathbb{E}\left[\left(\pi_k^{\overline{L},ML}(\varphi) - \pi_k^{\overline{L}}(\varphi)\right)^2\right] \leq C\left(\frac{1}{N_{\underline{L}}} + \sum_{l=\underline{L}+1}^{\overline{L}} \frac{\Delta_l^{\frac{1}{2}}}{N_l} + \sum_{(l,q)\in \mathsf{D}_{\underline{L},\overline{L}}} \frac{\Delta_l^{\frac{1}{2}}}{N_l N_q}\right),$$

where $\mathsf{D}_{\underline{L},\overline{L}} = \{(l, q) \in \{\underline{L}, \ldots, \overline{L}\} : l \neq q\}$.

Proof. This follows by using the C_2-inequality to separate the terms $\pi_k^{N_{\underline{L}},\underline{L}}(\varphi) - \pi_k^{\underline{L}}(\varphi)$ and $\sum_{l=\underline{L}+1}^{\overline{L}} \left[\pi_k^l - \pi_k^{l-1}\right]^{N_l}(\varphi) - \sum_{l=\underline{L}+1}^{\overline{L}} \left[\pi_k^l - \pi_k^{l-1}\right](\varphi)$. Then one uses standard results for particle filters (see e.g. [2, Lemma A.3]) for the $\pi_k^{N_{\underline{L}},\underline{L}}(\varphi) - \pi_k^{\underline{L}}(\varphi)$, and one multiplies out the brackets and uses Theorems 2 and 3 for the other term. □

The implication of this result is as follows. We know that the bias $\left|\pi_k^{\underline{L}}(\varphi) - \pi_k(\varphi)\right|$ is $\mathcal{O}(\Delta_{\overline{L}})$. This can be proved in a similar manner to [14, Lemma D.2], using the fact that the truncated Milstein scheme is a first-order (weak) method. Therefore, for $\epsilon > 0$ given, one can choose $\overline{L}$ such that $\Delta_{\overline{L}} = \mathcal{O}(\epsilon)$ (so that the squared bias is $\mathcal{O}(\epsilon^2)$). Then, to choose $N_{\underline{L}}, \ldots, N_{\overline{L}}$, one can set $N_{\underline{L}} = \mathcal{O}(\epsilon^{-2})$ and, just as in e.g. [9], $N_l = \mathcal{O}\left(\epsilon^{-2}\Delta_l^{3/4}\Delta_{\overline{L}}^{-1/4}\right)$, $l \in \{\underline{L}+1, \ldots, \overline{L}\}$. If one does this, one can verify that the upper bound in Theorem 1 is $\mathcal{O}(\epsilon^2)$, so the MSE is also $\mathcal{O}(\epsilon^2)$. The cost to run the algorithm per time step is $\mathcal{O}\left(\sum_{l=\underline{L}}^{\overline{L}} N_l \Delta_l^{-1}\right) = \mathcal{O}\left(\epsilon^{-2.5}\right)$. However, as we see in our simulations, the cost should be lower, at $\mathcal{O}\left(\epsilon^{-2}\log(\epsilon)^2\right)$; this indicates that our proof is not optimal. In the case that (1.1) has a non-constant diffusion coefficient with $d > 2$, the approach in [14], based upon using Euler discretizations, would obtain an MSE of $\mathcal{O}(\epsilon^2)$ at a cost of $\mathcal{O}\left(\epsilon^{-2.5}\right)$.

5. Numerical results

5.1. Models

We consider three different models for our numerical experiments.

5.1.1. *Model 1: geometric Brownian motion (GBM) process.* The first model we use is

$$dX_t = \mu X_t dt + \sigma X_t dW_t, \quad X_0 = x_0.$$

We set $Y_n | X_n = x \sim \mathcal{N}\left(\log(x), \tau^2\right)$, where $\tau^2 = 0.02$ and $\mathcal{N}(x, \tau^2)$ is the Gaussian distribution with mean x and variance τ^2. We choose $x_0 = 1$, $\mu = 0.02$, and $\sigma = 0.2$.

5.1.2. *Model 2: Clark–Cameron SDE.* The second model we consider is the Clark–Cameron stochastic differential equation (SDE) model (e.g. [11]) with initial conditions $x_0 = (0, 0)^\top$:

$$dX_{1,t} = dW_{1,t},$$
$$dX_{2,t} = X_{1,t} dW_{2,t},$$

where $X_{j,t}$ denotes the jth dimension of X_t, $j \in \{1, \ldots, d\}$. In addition, $Y_n | X_n = x \sim \mathcal{N}\left(\frac{X_1 + X_2}{2}, \tau^2\right)$, where $\tau^2 = 0.1$.

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

5.1.3. *Model 3: multidimensional SDEs with a nonlinear diffusion term (NLMs).* For our last model we use the following multidimensional SDEs, with $x_0 = (0, 0)^\top$:

$$dX_{1,t} = \theta_1(\mu_1 - X_{1,t})dt + \frac{\sigma_1}{\sqrt{1 + X_{1,t}^2}} dW_{1,t},$$

$$dX_{2,t} = \theta_2(\mu_2 - X_{1,t})dt + \frac{\sigma_2}{\sqrt{1 + X_{1,t}^2}} dW_{2,t}.$$

We set $Y_n|X_n = x \sim \mathcal{L}\left(\frac{X_1+X_2}{2}, s\right)$, where $\mathcal{L}(m, s)$ is the Laplace distribution with location m and scale s. The values of the parameters that we choose are $(\theta_1, \theta_2) = (1, 1)$, $(\mu_1, \mu_2) = (0, 0)$, $(\sigma_1, \sigma_2) = (1, 1)$, and $s = \sqrt{0.1}$.

5.2. Simulation settings

We will consider a comparison of the particle filter, MLPF (as in [14]), and AMLPF. For our numerical experiments, we consider multilevel estimators at levels $l \in \{3, 4, 5, 6, 7\}$. Resampling is performed adaptively. For the particle filters, resampling is done when the effective sample size (ESS) is less than $1/2$ of the particle numbers. Suppose that we have a collection of weights which sum to one, $(W^1, \ldots, W^N)$, for instance $\left(W_k^{1,l}, \ldots, W_k^{N_l,l}\right)$ in Algorithm 5. Then the ESS can be calculated using the formula

$$\frac{1}{\sum_{i=1}^N (W^i)^2}.$$

This is a number between 1 and N and is a measure of the variability of the weights. Typically one would like this measure to be close to N. For the coupled filters, we use the ESS of the coarse filter as the measure of discrepancy, and resampling occurs at the same threshold as the particle filter. Each simulation is repeated 100 times.

5.3. Simulation results

We now present our numerical simulations to show the benefits of the AMLPF for the three models under consideration. In Figures 1 and 2 we present the rate of the cost and the MSE for the estimator for the filter and normalizing constant. The cost is computed using the CPU time, although we found similar results using the formula of Section 4. The figures show that as we increase the level from $l = 3$ to $l = 7$, the difference between the costs of the two methods increases. In particular, the figures show the advantage and accuracy of using AMLPF. Table 1 presents the estimated rates of the MSE with respect to cost for the normalizing constant and filter. The results agree with the theory [14] and the results of this article, which predicts a rate of -1.5 for the particle filter, -1.25 for the MLPF, and -1 for the AMLPF.

Appendix A. Introduction and some notation

The purpose of this appendix is to provide the necessary technical results to prove Theorem 1. For a full understanding of the arguments, we advise the reader to read the proofs in order. It is possible to skip Section B and simply refer to the proofs, but the reader may miss some notation or nuances.

The structure of this appendix is as follows. We first consider some properties of the antithetic truncated Milstein scheme of [11]. We present several new results which are needed

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

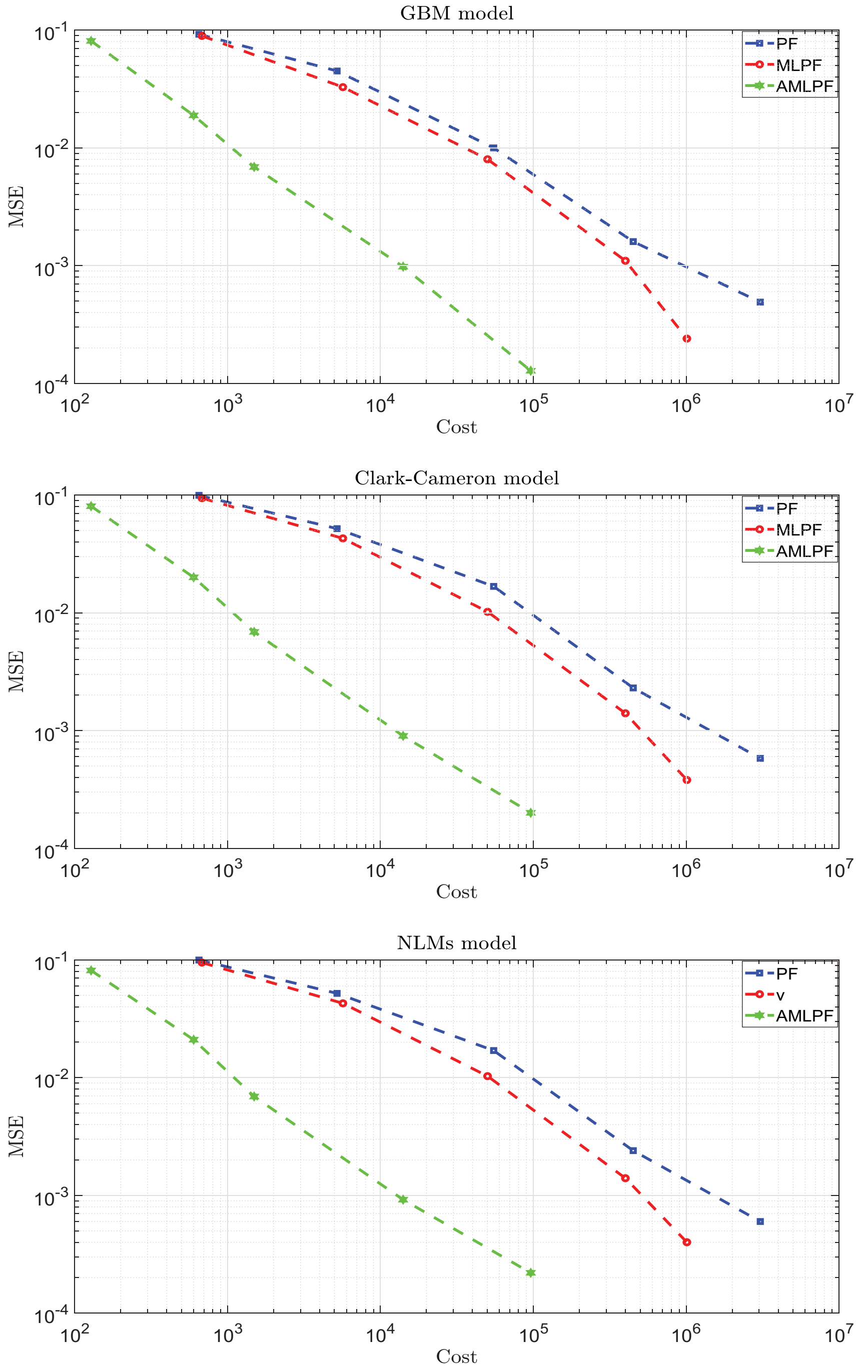

FIGURE 1. Cost rates as a function of the MSE. The results are for the filter.

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

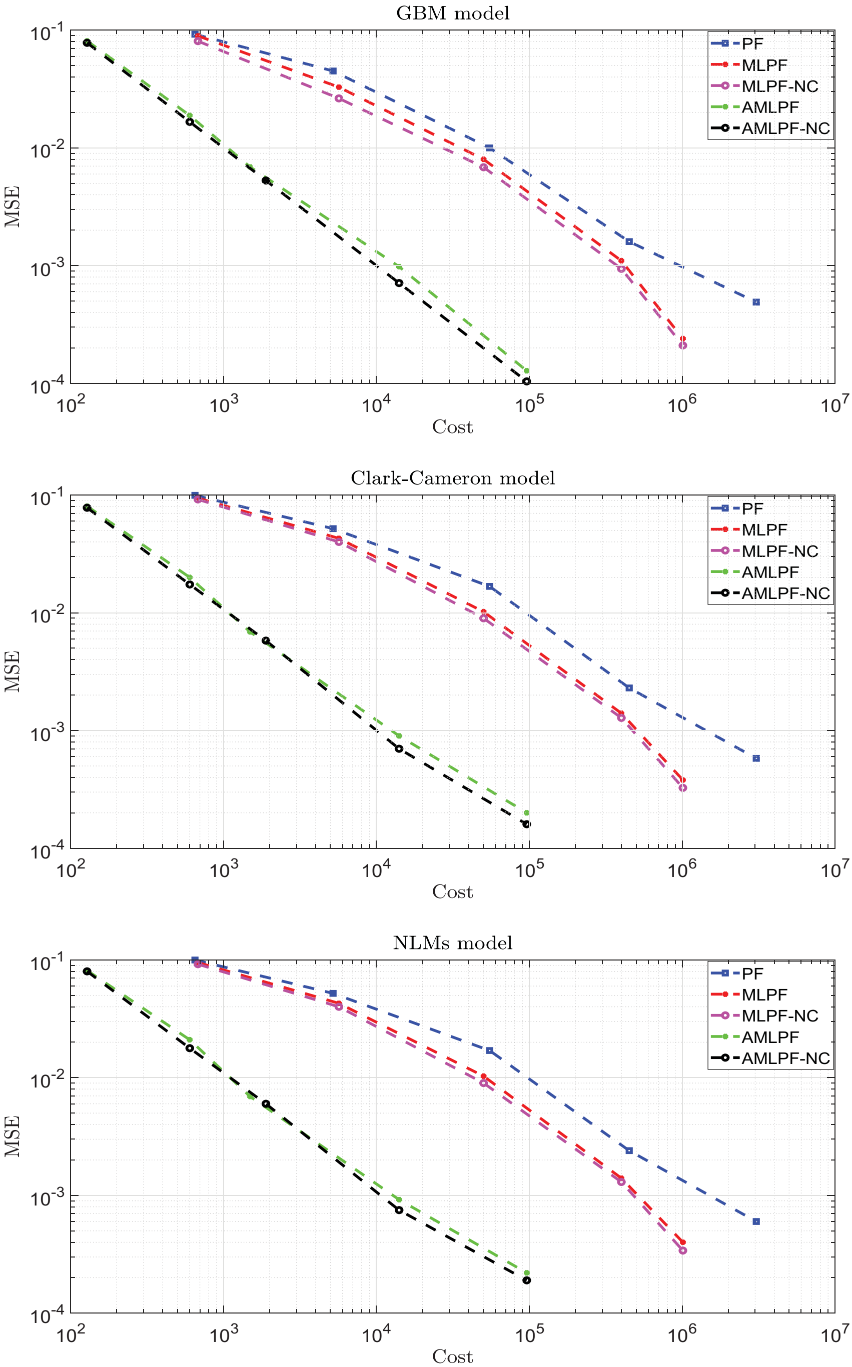

FIGURE 2. Cost rates as a function of the MSE in our algorithms, with results for the normalizing constants.

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

TABLE 1. Estimated rates of MSE with respect to cost. ‘NC’ stands for normalizing constant.

Model	PF (filter)	MLPF (NC)	MLPF (filter)	AMLPF (NC)	AMLPF (filter)
GBM	–1.53	–1.21	–1.23	–1.04	–1.02
Clark–Cameron	–1.55	–1.28	–1.26	–1.07	–1.05
NLMs	–1.56	–1.29	–1.27	–1.08	–1.06

directly for the subsequent proofs. In Section C we consider the coupled particle filter as in Algorithm 5. This section is split into three subsections. The first of these, Section C.1, considers the application of the results in Section B to a coupled particle filter. These results feed into the final two subsections, which consider the convergence of $\left[\pi_k^l - \pi_k^{l-1}\right]^{N_l}(\varphi)$ in $\mathbb{L}_p$ (Section C.2) and the associated bias (Section C.3). This is the culmination of the work in this appendix and is summarized in Theorems 2 and 3. To prove our results we use two major assumptions, (A1) and (A2), which can be found in Section 4 of the main text.

A.1. Some notation

Let $(\mathsf{V}, \mathcal{V})$ be a measurable space. For $\varphi : \mathsf{V} \to \mathbb{R}$ we write $\mathcal{B}_b(\mathsf{V})$ for the collection of bounded measurable functions. Let $\varphi : \mathbb{R}^d \to \mathbb{R}$, and let $\text{Lip}(\mathbb{R}^d)$ denote the collection of real-valued functions that are Lipschitz with respect to $\|\cdot\|$ (where $\|\cdot\|$ denotes the $\mathbb{L}_2$-norm of a vector $x \in \mathbb{R}^d$). That is, $\varphi \in \text{Lip}(\mathbb{R}^d)$ if there exists a $C < +\infty$ such that for any $(x, y) \in \mathbb{R}^{2d}$,

$$|\varphi(x) - \varphi(y)| \le C\|x - y\|.$$

For $\varphi \in \mathcal{B}_b(\mathsf{V})$, we write the supremum norm as $\|\varphi\|_\infty = \sup_{x\in\mathsf{V}} |\varphi(x)|$. For a measure μ on $(\mathsf{V}, \mathcal{V})$ and a function $\varphi \in \mathcal{B}_b(\mathsf{V})$, the notation $\mu(\varphi) = \int_{\mathsf{V}} \varphi(x)\mu(dx)$ is used. For $A \in \mathcal{V}$, the indicator function is written as $\mathbb{I}_A(x)$. If $K : \mathsf{V} \times \mathcal{V} \to [0, \infty)$ is a non-negative operator and μ is a measure, we use the notation $\mu K(dy) = \int_{\mathsf{V}} \mu(dx)K(x, dy)$, and for $\varphi \in \mathcal{B}_b(\mathsf{V})$, $K(\varphi)(x) = \int_{\mathsf{V}} \varphi(y)K(x, dy)$. Throughout, we use C to denote a generic finite constant whose value may change at each appearance and whose dependencies (on model and simulation parameters) are clear from the context.

Appendix B. Proofs for antithetic truncated Milstein scheme

The proofs of this section focus on the antithetic truncated Milstein discretization over a unit of time (i.e. as in Algorithm 2). The case we consider is almost identical to that in [11], except that we impose that our initial points $\left(x_0^l, x_0^{l-1}, x_0^{l,a}\right)$ need not be equal. This constraint is important in our subsequent proofs for the coupled (and multilevel) particle filter. Our objective is to prove a result similar to [11, Theorem 4.10], and to that end we make the assumption (A1), which is stronger than [11, Assumption 4.1]. The stronger assumption is related to the boundedness of the drift and diffusion coefficients of (1.1). The reason we require this is that it greatly simplifies our (subsequent) proofs if the constants in the results below do not depend on the initial points $\left(x_0^l, x_0^{l-1}, x_0^{l,a}\right)$; this would be the case otherwise. In this section, $\mathbb{E}$ denotes the expectation with respect to the law associated to Algorithm 2.

The final result in this section, Proposition B.1, is our adaptation of [11, Theorem 4.10] and is proved using simple modifications of Lemmata 4.6, 4.7, and 4.9 and Corollary 4.8 in [11].

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

The proofs of Lemma 4.7 and Corollary 4.8 in [11] need not be modified, so we proceed to prove analogues of Lemmata 4.6 and 4.9.

Lemma B.1. *Assume* (A1). *Then for any* $p \in [1, \infty)$ *there exists a* $C < +\infty$ *such that for any* $l \in \mathbb{N}$,

$$\mathbb{E}\left[\max_{k\in\left\{0,2,\ldots,\Delta_l^{-1}\right\}} \|X_{k\Delta_l}^l - X_{k\Delta_l}^{l,a}\|^p\right] \leq C\left(\Delta_l^{\frac{p}{2}} + \|x_0^l - x_0^{l-1}\|^p + \|x_0^{l,a} - x_0^{l-1}\|^p\right).$$

Proof. Using the C_p-inequality, we have that

$$\mathbb{E}\left[\max_{k\in\left\{0,2,\ldots,\Delta_l^{-1}\right\}} \|X_{k\Delta_l}^l - X_{k\Delta_l}^{l,a}\|^p\right]$$
$$\leq C\left(\mathbb{E}\left[\max_{k\in\left\{0,2,\ldots,\Delta_l^{-1}\right\}} \|X_{k\Delta_l}^l - X_{k\Delta_l}^{l-1}\|^p\right] + \mathbb{E}\left[\max_{k\in\left\{0,2,\ldots,\Delta_l^{-1}\right\}} \|X_{k\Delta_l}^{l,a} - X_{k\Delta_l}^{l-1}\|^p\right]\right).$$

Denoting by X_t^x the solution of (1.1) with initial point x at time t, we have

$$\mathbb{E}\left[\max_{k\in\left\{0,2,\ldots,\Delta_l^{-1}\right\}} \|X_{k\Delta_l}^l - X_{k\Delta_l}^{l,a}\|^p\right]$$
$$\leq C\Bigg(\mathbb{E}\left[\max_{k\in\left\{0,2,\ldots,\Delta_l^{-1}\right\}} \left\|X_{k\Delta_l}^l - X_{k\Delta_l}^{x_0^l}\right\|^p\right] + \mathbb{E}\left[\max_{k\in\left\{0,2,\ldots,\Delta_l^{-1}\right\}} \left\|X_{k\Delta_l}^{x_0^l} - X_{k\Delta_l}^{x_0^{l-1}}\right\|^p\right]$$
$$+\mathbb{E}\left[\max_{k\in\left\{0,2,\ldots,\Delta_l^{-1}\right\}} \left\|X_{k\Delta_l}^{l-1} - X_{k\Delta_l}^{x_0^{l-1}}\right\|^p\right] + \mathbb{E}\left[\max_{k\in\left\{0,2,\ldots,\Delta_l^{-1}\right\}} \left\|X_{k\Delta_l}^{l,a} - X_{k\Delta_l}^{x_0^{l,a}}\right\|^p\right]$$
$$+\mathbb{E}\left[\max_{k\in\left\{0,2,\ldots,\Delta_l^{-1}\right\}} \left\|X_{k\Delta_l}^{x_0^{l,a}} - X_{k\Delta_l}^{x_0^{l-1}}\right\|^p\right] + \mathbb{E}\left[\max_{k\in\left\{0,2,\ldots,\Delta_l^{-1}\right\}} \left\|X_{k\Delta_l}^{l-1} - X_{k\Delta_l}^{x_0^{l-1}}\right\|^p\right]\Bigg).$$

The proof is then easily concluded via the strong convergence result [11, Lemma 4.2] and standard results for diffusion processes (e.g. via Gronwall and [20, Corollary V.11.7]). □

Remark B.1. Note that the proof also establishes the following: for any $p \in [1, \infty)$ there exists a $C < +\infty$ such that for any $l \in \mathbb{N}$,

$$\mathbb{E}\left[\max_{k\in\left\{0,2,\ldots,\Delta_l^{-1}\right\}} \|X_{k\Delta_l}^l - X_{k\Delta_l}^{l-1}\|^p\right] \leq C\left(\Delta_l^{\frac{p}{2}} + \|x_0^l - x_0^{l-1}\|^p\right)$$

and

$$\mathbb{E}\left[\max_{k\in\left\{0,2,\ldots,\Delta_l^{-1}\right\}} \|X_{k\Delta_l}^{l,a} - X_{k\Delta_l}^{l-1}\|^p\right] \leq C\left(\Delta_l^{\frac{p}{2}} + \|x_0^{l,a} - x_0^{l-1}\|^p\right).$$

To state our next result, which mirrors [11, Lemma 4.9], we need to introduce a significant amount of notation directly from [11]. Below, we set $\overline{X}_{k\Delta_l}^l = \frac{1}{2}(X_{k\Delta_l}^l + X_{k\Delta_l}^{l,a})$, $k \in \left\{0, \ldots, \Delta_l^{-1}\right\}$. For $(i, k) \in \{1, \ldots, d\} \times \left\{2, 4, \ldots, \Delta_l^{-1}\right\}$ we define

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

$$
\begin{aligned}
R_{i,k}^{(1)} &= \left\{ \frac{1}{2}\left(\alpha_i(X_{k\Delta_l}^{l}) + \alpha_i(X_{k\Delta_l}^{l,a}) \right) - \alpha_i(\overline{X}_{k\Delta_l}^{l}) \right\} \Delta_{l-1}, \\
M_{i,k}^{(1)} &= \sum_{j=1}^{d} \left\{ \frac{1}{2}\left(\beta_{ij}(X_{k\Delta_l}^{l}) + \beta_{ij}(X_{k\Delta_l}^{l,a}) \right) - \beta_{ij}(\overline{X}_{k\Delta_l}^{l}) \right\} [W_{j,(k+2)\Delta_l} - W_{j,k\Delta_l}], \\
M_{i,k}^{(2)} &= \sum_{(j,m)\in\{1,\ldots,d\}^2} \left\{ \frac{1}{2}\left(h_{ijm}(X_{k\Delta_l}^{l}) + h_{ijm}(X_{k\Delta_l}^{l,a}) \right) - h_{ijm}(\overline{X}_{k\Delta_l}^{l}) \right\} \Big([W_{j,(k+2)\Delta_l} - W_{j,k\Delta_l}] \\
&\quad \times [W_{m,(k+2)\Delta_l} - W_{m,k\Delta_l}] - \Delta_{l-1} \Big), \\
M_{i,k}^{(3)} &= \sum_{(j,m)\in\{1,\ldots,d\}^2} \frac{1}{2}\left(h_{ijm}(X_{k\Delta_l}^{l}) - h_{ijm}(X_{k\Delta_l}^{l,a}) \right) \Big([W_{j,(k+1)\Delta_l} - W_{j,k\Delta_l}][W_{m,(k+2)\Delta_l} \\
&\quad - W_{m,(k+1)\Delta_l}] - [W_{m,(k+1)\Delta_l} - W_{m,k\Delta_l}][W_{j,(k+2)\Delta_l} - W_{j,(k+1)\Delta_l}] \Big), \\
R_{i,k}^{l} &= \sum_{j=1}^{d} \frac{\partial \alpha_i}{\partial x_j}(X_{k\Delta_l}^{l}) \Bigg(\alpha_j(X_{k\Delta_l}^{l})\Delta_l + \sum_{(m,n)\in\{1,\ldots,d\}^2} h_{imn}(X_{k\Delta_l}^{l}) \Big([W_{m,(k+1)\Delta_l} \\
&\quad - W_{m,k\Delta_l}][W_{n,(k+1)\Delta_l} - W_{n,k\Delta_l}] - \Delta_l \Big) \Bigg) \Delta_l + \frac{1}{2} \sum_{(j,m)\in\{1,\ldots,d\}^2} \frac{\partial^2 \alpha_i}{\partial x_j \partial x_m}(\xi_1^l)[X_{j,(k+1)\Delta_l}^{l} \\
&\quad - X_{j,k\Delta_l}^{l}][X_{m,(k+1)\Delta_l}^{l} - X_{m,k\Delta_l}^{l}]\Delta_l, \\
M_{i,k}^{(1,l)} &= \sum_{(j,m)\in\{1,\ldots,d\}^2} \frac{\partial \alpha_i}{\partial x_j}(X_{k\Delta_l}^{l}) \beta_{jm}(X_{k\Delta_l}^{l}) [W_{m,(k+1)\Delta_l} - W_{m,k\Delta_l}]\Delta_l, \\
M_{i,k}^{(2,l)} &= \sum_{(j,m)\in\{1,\ldots,d\}^2} \frac{\partial \beta_{ij}}{\partial x_m}(X_{k\Delta_l}^{l}) \Bigg(\alpha_m(X_{k\Delta_l}^{l})\Delta_l + \sum_{(n,p)\in\{1,\ldots,d\}^2} h_{mnp}(X_{k\Delta_l}^{l}) [[W_{n,(k+1)\Delta_l} - W_{n,k\Delta_l}] \\
&\quad \times [W_{p,(k+1)\Delta_l} - W_{p,k\Delta_l}] - \Delta_l] \Bigg) [W_{j,(k+2)\Delta_l} - W_{j,(k+1)\Delta_l}], \\
M_{i,k}^{(3,l)} &= \sum_{(j,m)\in\{1,\ldots,d\}^2} \left\{ h_{ijm}(X_{(k+1)\Delta_l}^{l}) - h_{ijm}(X_{k\Delta_l}^{l}) \right\} \Big([W_{j,(k+2)\Delta_l} - W_{j,(k+1)\Delta_l}][W_{m,(k+2)\Delta_l} \\
&\quad - W_{m,(k+1)\Delta_l}] - \Delta_l \Big),
\end{aligned}
$$

where, in $R_{i,k}^{l}$, ξ_1^l is some point which lies on the line between $X_{k\Delta_l}^{l}$ and $X_{(k+1)\Delta_l}^{l}$. In the case of $R_{i,k}^{l}$ and $M_{i,k}^{(j,l)}$, $j \in \{1, 2, 3\}$, one can substitute $X_{k\Delta_l}^{l}$, ξ_1^l, $X_{(k+1)\Delta_l}^{l}$ for $X_{k\Delta_l}^{l,a}$, $\xi_1^{l,a}$, $X_{(k+1)\Delta_l}^{l,a}$ (where $\xi_1^{l,a}$ is some point which lies on the line between $X_{k\Delta_l}^{l,a}$ and $X_{(k+1)\Delta_l}^{l,a}$); when we do so, we use the notation $R_{i,k}^{l,a}$ and $M_{i,k}^{(j,l,a)}$, $j \in \{1, 2, 3\}$. Finally we set

$$
\begin{aligned}
M_k &= \sum_{j=1}^{3} M_k^{(j)} + \frac{1}{2} \sum_{j=1}^{3} \left\{ M_k^{(j,l)} + M_k^{(j,l,a)} \right\}, \\
R_k &= R_k^{(1)} + \frac{1}{2} \left\{ R_{i,k}^{l} + R_{i,k}^{l,a} \right\}.
\end{aligned}
$$

We are now in a position to give our analogue of [11, Lemma 4.9].

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

Lemma B.2. *Assume* (A1). *Then the following hold:*

- *For any* $p \in [1, \infty)$ *there exists a* $C < +\infty$ *such that for any* $l \in \mathbb{N}$,

$$\max_{k\in\{2,4,\ldots,\Delta_l\}} \mathbb{E}[|R_k|^p] \leq C\Delta_l^p \left(\Delta_l^p + \|x_0^l - x_0^{l-1}\|^{2p} + \|x_0^{l,a} - x_0^{l-1}\|^{2p}\right).$$

- *For any* $p \in [1, \infty)$ *there exists a* $C < +\infty$ *such that for any* $l \in \mathbb{N}$,

$$\max_{k\in\{2,4,\ldots,\Delta_l\}} \mathbb{E}[|M_k|^p]$$

$$\leq C\left(\Delta_l^{\frac{3p}{2}} + \Delta_l^{\frac{p}{2}}\left\{\|x_0^l - x_0^{l-1}\|^{2p} + \|x_0^{l,a} - x_0^{l-1}\|^{2p}\right\} + \Delta_l^p\left\{\|x_0^l - x_0^{l-1}\|^p + \|x_0^{l,a} - x_0^{l-1}\|^p\right\}\right).$$

Proof. The proof of this result essentially follows from [11] and controlling the terms (in $\mathbb{L}_p$). The expressions $\frac{1}{2}\sum_{j=1}^3 \left\{M_k^{(j,l)} + M_k^{(j,l,a)}\right\}$ and $\frac{1}{2}\{R_{i,k}^l + R_{i,k}^{l,a}\}$ can be dealt with exactly as in [11, Lemma 4.7], so we need only consider the terms $\sum_{j=1}^3 M_k^{(j)}$ and $R_k^{(1)}$. It will suffice to control in $\mathbb{L}_p$ any of the d coordinates of the aforementioned vectors, which is what we do below.

Beginning with $R_{i,k}^{(1)}$, using the second-order Taylor expansion in [11], one has

$$R_{i,k}^{(1)} = \frac{1}{8}\sum_{(j,m)\in\{1,\ldots,d\}^2}\left(\frac{\partial^2\alpha_i}{\partial x_j\partial x_m}(\xi_1^l) + \frac{\partial^2\alpha_i}{\partial x_j\partial x_m}(\xi_2^l)\right)\left(X_{j,k\Delta_l}^l - X_{j,k\Delta_l}^{l,a}\right)\left(X_{m,k\Delta_l}^l - X_{m,k\Delta_l}^{l,a}\right)\Delta_l,$$

where ξ_1^l is some point between $\overline{X}_{k\Delta_l}^l$ and $X_{k\Delta_l}^l$ and ξ_2^l is some point between $\overline{X}_{k\Delta_l}^l$ and $X_{k\Delta_l}^{l,a}$. Then it follows easily that

$$\mathbb{E}\big[|R_{i,k}^{(1)}|^p\big] \leq C\Delta_l^p\mathbb{E}\Big[\Big\|X_{k\Delta_l}^l - X_{k\Delta_l}^{l,a}\Big\|^{2p}\Big].$$

Application of Lemma B.1 yields that

$$\mathbb{E}\big[|R_{i,k}^{(1)}|^p\big] \leq C\Delta_l^p\left(\Delta_l^p + \|x_0^l - x_0^{l-1}\|^{2p} + \|x_0^{l,a} - x_0^{l-1}\|^{2p}\right),$$

which is the desired result.

For $M_{i,k}^{(1)}$, again one has

$$M_{i,k}^{(1)} = \frac{1}{16}\sum_{(j,m,n)\in\{1,\ldots,d\}^3}\left\{\frac{\partial^2\beta_{ij}}{\partial x_m\partial x_n}(\xi_3^l) + \frac{\partial^2\beta_{ij}}{\partial x_m\partial x_n}(\xi_4^l)\right\}\left(X_{m,k\Delta_l}^l - X_{m,k\Delta_l}^{l,a}\right)\left(X_{n,k\Delta_l}^l - X_{n,k\Delta_l}^{l,a}\right)$$
$$\times\big[W_{j,(k+2)\Delta_l} - W_{j,k\Delta_l}\big],$$

where ξ_3^l is some point between $\overline{X}_{k\Delta_l}^l$ and $X_{k\Delta_l}^l$ and ξ_4^l is some point between $\overline{X}_{k\Delta_l}^l$ and $X_{k\Delta_l}^{l,a}$. Then, using the independence of the Brownian increment (with the random variables $\left(X_{m,k\Delta_l}^l - X_{m,k\Delta_l}^{l,a}\right)\left(X_{n,k\Delta_l}^l - X_{n,k\Delta_l}^{l,a}\right)\Big)$ and the same approach as above, one has

$$\mathbb{E}\big[|M_{i,k}^{(1)}|^p\big] \leq C\Delta_l^{p/2}\mathbb{E}\big[\|X_{k\Delta_l}^l - X_{k\Delta_l}^{l,a}\|^{2p}\big],$$

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

and applying Lemma B.1 yields

$$\mathbb{E}\big[|M_{i,k}^{(1)}|^p\big] \le C\Delta_l^{p/2}\left(\Delta_l^p + \|x_0^l - x_0^{l-1}\|^{2p} + \|x_0^{l,a} - x_0^{l-1}\|^{2p}\right).$$

For $M_{i,k}^{(2)}$ one has

$$M_{i,k}^{(2)} = \frac{1}{4}\sum_{(j,m,n)\in\{1,\dots,d\}^3}\left\{\frac{\partial h_{ijm}}{\partial x_n}(\xi_5^l) + \frac{\partial h_{ijm}}{\partial x_n}(\xi_6^l)\right\}\left(X_{n,k\Delta_l}^l - X_{n,k\Delta_l}^{l,a}\right)$$
$$\times\Big(\big[W_{j,(k+2)\Delta_l} - W_{j,k\Delta_l}\big]\big[W_{m,(k+2)\Delta_l} - W_{m,k\Delta_l}\big] - \Delta_{l-1}\Big),$$

where ξ_5^l is some point between $\overline{X}_{k\Delta_l}^l$ and $X_{k\Delta_l}^l$ and ξ_6^l is some point between $\overline{X}_{k\Delta_l}^l$ and $X_{k\Delta_l}^{l,a}$. As the Brownian increments are independent of $\left(X_{n,k\Delta_l}^l - X_{n,k\Delta_l}^{l,a}\right)$ and the dimensions are also independent of each other, one obtains

$$\mathbb{E}\big[|M_{i,k}^{(2)}|^p\big] \le C\Delta_l^p\mathbb{E}\big[\|X_{k\Delta_l}^l - X_{k\Delta_l}^{l,a}\|^p\big].$$

Lemma B.1 gives

$$\mathbb{E}\big[|M_{i,k}^{(2)}|^p\big] \le C\Delta_l^p\left(\Delta_l^{p/2} + \|x_0^l - x_0^{l-1}\|^p + \|x_0^{l,a} - x_0^{l-1}\|^p\right).$$

For $M_{i,k}^{(3)}$, one has

$$M_{i,k}^{(3)} = \frac{1}{4}\sum_{(j,m,n)\in\{1,\dots,d\}^3}\left\{\frac{\partial h_{ijm}}{\partial x_n}(\xi_7^l) + \frac{\partial h_{ijm}}{\partial x_n}(\xi_8^l)\right\}(X_{k\Delta_l}^l - X_{k\Delta_l}^{l,a})\Big(\big[W_{j,(k+1)\Delta_l} - W_{j,k\Delta_l}\big]$$
$$\times\big[W_{m,(k+2)\Delta_l} - W_{m,(k+1)\Delta_l}\big] - \big[W_{m,(k+1)\Delta_l} - W_{m,k\Delta_l}\big]\big[W_{j,(k+2)\Delta_l} - W_{j,(k+1)\Delta_l}\big]\Big),$$

where ξ_7^l is some point between $\overline{X}_{k\Delta_l}^l$ and $X_{k\Delta_l}^l$ and ξ_8^l is some point between $\overline{X}_{k\Delta_l}^l$ and $X_{k\Delta_l}^{l,a}$. Using essentially the same approach as for $M_{i,k}^{(2)}$, one obtains

$$\mathbb{E}\big[|M_{i,k}^{(3)}|^p\big] \le C\Delta_l^p\left(\Delta_l^{p/2} + \|x_0^l - x_0^{l-1}\|^p + \|x_0^{l,a} - x_0^{l-1}\|^p\right),$$

which concludes the proof. □

Proposition B.1. *Assume* (A1). *Then for any* $p \in [1, \infty)$ *there exists a* $C < +\infty$ *such that for any* $l \in \mathbb{N}$,

$$\mathbb{E}\left[\max_{k\in\{0,2,\dots,\Delta_l^{-1}\}}\|\overline{X}_{k\Delta_l}^l - X_{k\Delta_l}^{l-1}\|^p\right]$$
$$\le C\left(\Delta_l^p + \left\{\|x_0^l - x_0^{l-1}\|^{2p} + \|x_0^{l,a} - x_0^{l-1}\|^{2p}\right\} + \Delta_l^{\frac{p}{2}}\left\{\|x_0^l - x_0^{l-1}\|^p + \|x_0^{l,a} - x_0^{l-1}\|^p\right\}\right).$$

Proof. We omit the proof, because it is identical to that of [11, Theorem 4.10], except that one uses Lemma B.2 above instead of [11, Lemma 4.9]. □

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

Appendix C. Proofs for coupled particle filter

In this section, $\mathbb{E}$ denotes the expectation with respect to the law that generates the AMLPF.

C.1. Rate proofs

Our analysis will apply to any coupling used in the second bullet point of Algorithm 4, Step 2. At any time point k of Algorithm 5, we will denote the resampled index of particle $i \in \{1, \ldots, N_l\}$ by $I_k^{i,l}$ (level l), $I_k^{i,l-1}$ (level $l-1$), and $I_k^{i,l,a}$ (level l antithetic). Now let $I_k^l(i) = I_k^{i,l}$, $I_k^{l-1}(i) = I_k^{i,l-1}$, $I_k^{l,a}(i) = I_k^{i,l,a}$, and define S_k^l as the collection of indices that choose the same ancestor at each resampling step, i.e.

$$\mathsf{S}_k^l = \big\{ i \in \{1, \ldots, N_l\} : I_k^l(i) = I_k^{l-1}(i) = I_k^{l,a}(i), I_{k-1}^l \circ I_k^l(i) = I_{k-1}^{l-1} \circ I_k^{l-1}(i) = I_{k-1}^{l,a} \circ I_k^{l,a}(i), \ldots, \\ I_1^l \circ I_2^l \circ \cdots \circ I_k^l(i) = I_1^{l-1} \circ I_2^{l-1} \circ \cdots \circ I_k^{l-1}(i) = I_1^{l,a} \circ I_2^{l,a} \circ \cdots \circ I_k^{l,a}(i) \big\}.$$

We use the convention that $\mathsf{S}_0^l = \{1, \ldots, N_l\}$. Denote the σ-field generated by the simulated samples, resampled samples, and resampled indices up to time k by $\hat{\mathcal{F}}_k^l$, and denote the σ-field which does the same, except excluding the resampled samples and indices, by $\mathcal{F}_k^l$.

Lemma C.1. *Assume* (A1)–(A2). *Then for any* $(p, k) \in [1, \infty) \times \mathbb{N}$ *there exists a* $C < +\infty$ *such that for any* $(l, N_l) \in \mathbb{N}^2$*,*

$$\mathbb{E}\left[\frac{1}{N_l} \sum_{i \in \mathsf{S}_{k-1}^l} \left\{ \left\| X_k^{i,l} - X_k^{i,l-1} \right\|^p + \left\| X_k^{i,l,a} - X_k^{i,l-1} \right\|^p + \left\| X_k^{i,l} - X_k^{i,l,a} \right\|^p \right\} \right] \leq C \Delta_l^{p/2}.$$

Proof. The case $k = 1$ follows from the work in [11] (the case $p = [1, 2)$ can be adapted from that paper), so we assume $k \geq 2$. By conditioning on $\hat{\mathcal{F}}_{k-1}^l$ and applying Lemma B.1 (see also Remark B.1), we have

$$\mathbb{E}\left[\frac{1}{N_l} \sum_{i \in \mathsf{S}_{k-1}^l} \left\{ \left\| X_k^{i,l} - X_k^{i,l-1} \right\|^p + \left\| X_k^{i,l,a} - X_k^{i,l-1} \right\|^p + \left\| X_k^{i,l} - X_k^{i,l,a} \right\|^p \right\} \right]$$

$$\leq C\left(\Delta_l^{p/2} + \mathbb{E}\left[\frac{1}{N_l} \sum_{i \in \mathsf{S}_{k-1}^l} \left\{ \left\| \hat{X}_{k-1}^{i,l} - \hat{X}_{k-1}^{i,l-1} \right\|^p + \left\| \hat{X}_{k-1}^{i,l,a} - \hat{X}_{k-1}^{i,l-1} \right\|^p + \left\| \hat{X}_{k-1}^{i,l} - \hat{X}_{k-1}^{i,l,a} \right\|^p \right\} \right] \right).$$

One can exchangeably write the expectation on the right-hand side as

$$\mathbb{E}\left[\frac{1}{N_l} \sum_{i \in \mathsf{S}_{k-1}^l} \left\{ \left\| \hat{X}_{k-1}^{I_{k-1}^{i,l},l} - \hat{X}_{k-1}^{I_{k-1}^{i,l-1},l-1} \right\|^p + \left\| \hat{X}_{k-1}^{I_{k-1}^{i,l,a},l,a} - \hat{X}_{k-1}^{I_{k-1}^{i,l-1},l-1} \right\|^p + \left\| \hat{X}_{k-1}^{I_{k-1}^{i,l},l} - \hat{X}_{k-1}^{I_{k-1}^{i,l,a},l,a} \right\|^p \right\} \right].$$

The proof from here is then essentially identical (up to the fact that one has three indices instead of two) to that of [14, Lemma D.3] and is hence omitted. □

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

Lemma C.2. *Assume* (A1)–(A2). *Then for any* $(p, k) \in [1, \infty) \times \mathbb{N}$ *there exists a* $C < +\infty$ *such that for any* $(l, N_l) \in \mathbb{N}^2$,

$$\mathbb{E}\left[\frac{1}{N_l}\sum_{i \in \mathsf{S}_{k-1}^l} \left\|\overline{X}_k^{i,l} - X_k^{i,l-1}\right\|^p\right] \le C\Delta_l^p.$$

Proof. Following the start of the proof of Lemma C.1, except using Proposition B.1 instead of Lemma B.1, one can deduce that

$$\mathbb{E}\left[\frac{1}{N_l}\sum_{i \in \mathsf{S}_{k-1}^l} \left\|\overline{X}_k^{i,l} - X_k^{i,l-1}\right\|^p\right] \le$$

$$C\Bigg(\Delta_l^p + \mathbb{E}\left[\frac{1}{N_l}\sum_{i \in \mathsf{S}_{k-1}^l}\left\{\left\|\hat{X}_{k-1}^{I_{k-1}^{i,l},l} - \hat{X}_{k-1}^{I_{k-1}^{i,l-1},l-1}\right\|^{2p} + \left\|\hat{X}_{k-1}^{I_{k-1}^{i,l,a},l,a} - \hat{X}_{k-1}^{I_{k-1}^{i,l-1},l-1}\right\|^{2p}\right\}\right]$$

$$+\Delta_l^{p/2}\mathbb{E}\left[\frac{1}{N_l}\sum_{i \in \mathsf{S}_{k-1}^l}\left\{\left\|\hat{X}_{k-1}^{I_{k-1}^{i,l},l} - \hat{X}_{k-1}^{I_{k-1}^{i,l-1},l-1}\right\|^{p} + \left\|\hat{X}_{k-1}^{I_{k-1}^{i,l,a},l,a} - \hat{X}_{k-1}^{I_{k-1}^{i,l-1},l-1}\right\|^{p}\right\}\right]\Bigg).$$

From here, one can follow the calculations in [14, Lemma D.3] to deduce that

$$\mathbb{E}\left[\frac{1}{N_l}\sum_{i \in \mathsf{S}_{k-1}^l} \left\|\overline{X}_k^{i,l} - X_k^{i,l-1}\right\|^p\right] \le$$

$$C\Bigg(\Delta_l^p + \mathbb{E}\left[\frac{1}{N_l}\sum_{i \in \mathsf{S}_{k-2}^l}\left\{\left\|X_{k-1}^{i,l} - X_{k-1}^{i,l-1}\right\|^{2p} + \left\|X_{k-1}^{i,l,a} - X_{k-1}^{i,l-1}\right\|^{2p} + \left\|X_{k-1}^{i,l} - X_{k-1}^{i,l,a}\right\|^{2p}\right\}\right]$$

$$+\Delta_l^{p/2}\mathbb{E}\left[\frac{1}{N_l}\sum_{i \in \mathsf{S}_{k-2}^l}\left\{\left\|X_{k-1}^{i,l} - X_{k-1}^{i,l-1}\right\|^{p} + \left\|X_{k-1}^{i,l,a} - X_{k-1}^{i,l-1}\right\|^{p} + \left\|X_{k-1}^{i,l} - X_{k-1}^{i,l,a}\right\|^{p}\right\}\right]\Bigg).$$

Application of Lemma C.1 concludes the result.

Lemma C.3. *Assume* (A1)–(A2). *Then for any* $k \in \mathbb{N}$ *there exists a* $C < +\infty$ *such that for any* $(l, N_l) \in \mathbb{N}^2$,

$$\mathbb{E}\left[\mathbb{I}_{\mathsf{S}_{k-1}^l}(i)\left|\left\{\frac{g_k\big(X_k^{i,l}\big)}{\sum_{j=1}^{N_l} g_k\big(X_k^{j,l}\big)} - \frac{g_k\big(X_k^{i,l,a}\big)}{\sum_{j=1}^{N_l} g_k\big(X_k^{j,l,a}\big)}\right\}\right|\right]$$
$$\le C\left(\frac{\Delta_l^{1/2}}{N_l} + \frac{1}{N_l}\left\{1 - \mathbb{E}\left[\frac{Card\big(\mathsf{S}_{k-1}^l\big)}{N_l}\right]\right\}\right).$$

Proof. We have the elementary decomposition

$$\frac{g_k(X_k^{i,l})}{\sum_{j=1}^{N_l} g_k(X_k^{j,l})} - \frac{g_k(X_k^{i,l,a})}{\sum_{j=1}^{N_l} g_k(X_k^{j,l,a})} = \frac{g_k(X_k^{i,l}) - g_k(X_k^{i,l,a})}{\sum_{j=1}^{N_l} g_k(X_k^{j,l})}$$

$$+ g_k(X_k^{i,l,a}) \left(\frac{\sum_{j=1}^{N_l} g_k(X_k^{j,l,a}) - \sum_{j=1}^{N_l} g_k(X_k^{j,l})}{\sum_{j=1}^{N_l} g_k(X_k^{j,l,a}) \sum_{j=1}^{N_l} g_k(X_k^{j,l})} \right).$$

So application of the triangle inequality along with the lower and upper bounds on g_k (which are uniform in x) yields

$$\mathbb{E}\left[\mathbb{I}_{\mathsf{S}_{k-1}^l}(i) \left| \left\{ \frac{g_k(X_k^{i,l})}{\sum_{j=1}^{N_l} g_k(X_k^{j,l})} - \frac{g_k(X_k^{i,l,a})}{\sum_{j=1}^{N_l} g_k(X_k^{j,l,a})} \right\} \right| \right] \leq$$

$$\frac{C}{N_l}\left(\mathbb{E}\Big[\mathbb{I}_{\mathsf{S}_{k-1}^l}(i) \Big| g_k(X_k^{i,l}) - g_k(X_k^{i,l,a}) \Big| \Big] + \frac{1}{N_l}\mathbb{E}\left[\sum_{j \in \mathsf{S}_{k-1}^l} \left| g_k\Big(X_k^{j,l}\Big) - g_k\Big(X_k^{j,l,a}\Big) \right| \right] \right.$$
$$\left. + 1 - \mathbb{E}\left[\frac{\mathrm{Card}(\mathsf{S}_{k-1}^l)}{N_l} \right] \right).$$

Using the Lipschitz property of g_k along with Lemma C.1 allows one to conclude the result. □

The following is a useful lemma that we will need below. Its proof is a matter of direct computation and is hence omitted.

Lemma C.4. *Let* $(G^l, f^l, G^a, f^a, G^{l-1}, f^{l-1}) \in \mathbb{R}^6$ *with* (f^l, f^a, f^{l-1}) *non-zero; then*

$$\frac{\frac{1}{2}G^l}{f^l} + \frac{\frac{1}{2}G^a}{f^a} - \frac{G^{l-1}}{f^{l-1}} = \frac{1}{f^{l-1}}\Big(\tfrac{1}{2}G^l + \tfrac{1}{2}G^a - G^{l-1}\Big) + \frac{1}{f^a f^l f^{l-1}}\Big(\tfrac{1}{2}(G^l - G^a)(f^{l-1} - f^l)f^a +$$
$$\tfrac{1}{2}G^a\Big[(f^a - f^l)(f^{l-1} - f^l) - 2f^l\Big\{\tfrac{1}{2}f^l + \tfrac{1}{2}f^a - f^{l-1}\Big\}\Big]\Big).$$

Lemma C.5. *Assume* (A1)–(A2). *Then for any* $k \in \mathbb{N}$ *there exists a* $C < +\infty$ *such that for any* $(l, N_l) \in \mathbb{N}^2$,

$$\mathbb{E}\left[\mathbb{I}_{\mathsf{S}_{k-1}^l}(i) \left| \left\{ \frac{\frac{1}{2}g_k(X_k^{i,l})}{\sum_{j=1}^{N_l} g_k(X_k^{j,l})} + \frac{\frac{1}{2}g_k(X_k^{i,l,a})}{\sum_{j=1}^{N_l} g_k(X_k^{j,l,a})} - \frac{g_k(X_k^{i,l-1})}{\sum_{j=1}^{N_l} g_k(X_k^{j,l-1})} \right\} \right| \right]$$

$$\leq C\left(\frac{\Delta_l}{N_l} + \frac{1}{N_l}\left\{ 1 - \mathbb{E}\left[\frac{Card(\mathsf{S}_{k-1}^l)}{N_l} \right] \right\} \right).$$

Proof. One can apply Lemma C.4 to see that

$$\frac{\frac{1}{2}g_k(X_k^{i,l})}{\sum_{j=1}^{N_l} g_k(X_k^{j,l})} + \frac{\frac{1}{2}g_k(X_k^{i,l,a})}{\sum_{j=1}^{N_l} g_k(X_k^{j,l,a})} - \frac{g_k(X_k^{i,l-1})}{\sum_{j=1}^{N_l} g_k(X_k^{j,l-1})} = \sum_{j=1}^{4} T_j,$$

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

where

$$T_1 = \frac{1}{\sum_{j=1}^{N_l} g_k\left(X_k^{j,l-1}\right)} \left\{\tfrac{1}{2} g_k\left(X_k^{i,l}\right) + \tfrac{1}{2} g_k\left(X_k^{i,l,a}\right) - g_k\left(X_k^{i,l-1}\right)\right\},$$

$$T_2 = \frac{1}{2\left[\sum_{j=1}^{N_l} g_k\left(X_k^{j,l}\right)\right]\left[\sum_{j=1}^{N_l} g_k\left(X_k^{j,l-1}\right)\right]} \left(g_k\left(X_k^{i,l}\right) - g_k\left(X_k^{i,l,a}\right)\right)\left(\sum_{j=1}^{N_l}\left[g_k\left(X_k^{j,l-1}\right) - g_k\left(X_k^{j,l}\right)\right]\right),$$

$$T_3 = \frac{g_k\left(X_k^{i,l,a}\right)}{2\left[\sum_{j=1}^{N_l} g_k\left(X_k^{j,l,a}\right)\right]\left[\sum_{j=1}^{N_l} g_k\left(X_k^{j,l}\right)\right]\left[\sum_{j=1}^{N_l} g_k\left(X_k^{j,l-1}\right)\right]}\left(\sum_{j=1}^{N_l}\left[g_k\left(X_k^{j,l,a}\right) - g_k\left(X_k^{j,l}\right)\right]\right) \times \left(\sum_{j=1}^{N_l}\left[g_k\left(X_k^{j,l-1}\right) - g_k\left(X_k^{j,l}\right)\right]\right),$$

$$T_4 = -\frac{g_k\left(X_k^{i,l,a}\right)}{\left[\sum_{j=1}^{N_l} g_k\left(X_k^{j,l,a}\right)\right]\left[\sum_{j=1}^{N_l} g_k\left(X_k^{j,l-1}\right)\right]}\left\{\sum_{j=1}^{N_l}\left[\tfrac{1}{2} g_k\left(X_k^{j,l}\right) + \tfrac{1}{2} g_k\left(X_k^{j,l,a}\right) - g_k\left(X_k^{j,l-1}\right)\right]\right\}.$$

The terms T_1 (resp. T_2) and T_4 (resp. T_3) can be dealt with in a similar manner, so we only consider T_1 (resp. T_2).

For the case of T_1 we have the upper bound

$$\mathbb{E}\left[\mathbb{I}_{\mathsf{S}_{k-1}^l}(i) T_1\right] \le \frac{C}{N_l}\mathbb{E}\left[\mathbb{I}_{\mathsf{S}_{k-1}^l}(i)\left|\tfrac{1}{2} g_k\big(X_k^{i,l}\big) + \tfrac{1}{2} g_k\big(X_k^{i,l,a}\big) - g_k\big(X_k^{i,l-1}\big)\right|\right].$$

Then one can apply [11, Lemma 2.2] along with Lemmata C.1 and C.2 to deduce that

$$\mathbb{E}\left[\mathbb{I}_{\mathsf{S}_{k-1}^l}(i) T_1\right] \le \frac{C\Delta_l}{N_l}.$$

For the case of T_2 we have the upper bound

$$\mathbb{E}\left[\mathbb{I}_{\mathsf{S}_{k-1}^l}(i) T_2\right] \le \frac{C}{N_l^2}\left(\mathbb{E}\left[\mathbb{I}_{\mathsf{S}_{k-1}^l}(i)\left|g_k\big(X_k^{i,l,a}\big) - g_k\big(X_k^{i,l}\big)\right|\sum_{j\in\mathsf{S}_{k-1}^l}\left[g_k\big(X_k^{j,l-1}\big) - g_k\big(X_k^{j,l,a}\big)\right]\right|\right] + \mathbb{E}\left[\mathbb{I}_{\mathsf{S}_{k-1}^l}(i)\left|g_k\big(X_k^{i,l,a}\big) - g_k\big(X_k^{i,l}\big)\right|\left(N_l - \mathrm{Card}\big(\mathsf{S}_{k-1}^l\big)\right)\right]\right).$$

For the first term on the right-hand side one can use the Lipschitz property of g_k, Cauchy–Schwarz, and Lemma C.1, and for the second term one can use the boundedness of g_k, to obtain

$$\mathbb{E}\left[\mathbb{I}_{\mathsf{S}_{k-1}^l}(i) T_2\right] \le C\left(\frac{\Delta_l}{N_l} + \frac{1}{N_l}\left\{1 - \mathbb{E}\left[\frac{\mathrm{Card}\big(\mathsf{S}_{k-1}^l\big)}{N_l}\right]\right\}\right),$$

which concludes the proof. □

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

Lemma C.6. *Assume* (A1)–(A2). *Then for any* $k \in \mathbb{N}_0$ *there exists a* $C < +\infty$ *such that for any* $(l, N_l) \in \mathbb{N}^2$,

$$1 - \mathbb{E}\left[\frac{Card\big(\mathsf{S}_k^l\big)}{N_l}\right] \le C\Delta_l^{1/2}.$$

Proof. We note that

$$\begin{aligned}
1 - \mathbb{E}\left[\frac{\mathrm{Card}\big(\mathsf{S}_k^l\big)}{N_l}\right] &= 1 - \mathbb{E}\left[\sum_{i=1}^{N_l} \min\left\{\frac{g_k\big(X_k^{i,l}\big)}{\sum_{j=1}^{N_l} g_k\big(X_k^{j,l}\big)}\frac{g_k\big(X_k^{i,l,a}\big)}{\sum_{j=1}^{N_l} g_k\big(X_k^{j,l,a}\big)}, \frac{g_k\big(X_k^{i,l-1}\big)}{\sum_{j=1}^{N_l} g_k\big(X_k^{j,l-1}\big)}\right\}\right] \\
&\quad + \mathbb{E}\left[\sum_{i \notin \mathsf{S}_{k-1}^l} \min\left\{\frac{g_k\big(X_k^{i,l}\big)}{\sum_{j=1}^{N_l} g_k\big(X_k^{j,l}\big)}\frac{g_k\big(X_k^{i,l,a}\big)}{\sum_{j=1}^{N_l} g_k\big(X_k^{j,l,a}\big)}, \frac{g_k\big(X_k^{i,l-1}\big)}{\sum_{j=1}^{N_l} g_k\big(X_k^{j,l-1}\big)}\right\}\right] \\
&\le 1 - \mathbb{E}\left[\sum_{i=1}^{N_l} \min\left\{\frac{g_k\big(X_k^{i,l}\big)}{\sum_{j=1}^{N_l} g_k\big(X_k^{j,l}\big)}\frac{g_k\big(X_k^{i,l,a}\big)}{\sum_{j=1}^{N_l} g_k\big(X_k^{j,l,a}\big)}, \frac{g_k\big(X_k^{i,l-1}\big)}{\sum_{j=1}^{N_l} g_k\big(X_k^{j,l-1}\big)}\right\}\right] \\
&\quad + C\left(1 - \mathbb{E}\left[\frac{\mathrm{Card}\big(\mathsf{S}_{k-1}^l\big)}{N_l}\right]\right). \qquad \text{(C.1)}
\end{aligned}$$

As our proof strategy is via induction (the initialization is trivially true since $\mathsf{S}_0^l = \{1, \ldots, N_l\}$), we need to focus on the first term on the right-hand side of (C.1).

Now, applying twice the result that $\min\{a, b\} = \frac{1}{2}(a + b - |a - b|)$ for any $(a, b) \in \mathbb{R}^2$, we easily obtain that

$$1 - \sum_{i=1}^{N_l} \min\left\{\frac{g_k\big(X_k^{i,l}\big)}{\sum_{j=1}^{N_l} g_k\big(X_k^{j,l}\big)}\frac{g_k\big(X_k^{i,l,a}\big)}{\sum_{j=1}^{N_l} g_k\big(X_k^{j,l,a}\big)}, \frac{g_k\big(X_k^{i,l-1}\big)}{\sum_{j=1}^{N_l} g_k\big(X_k^{j,l-1}\big)}\right\}$$

$$\begin{aligned}
= \frac{1}{2}\sum_{i=1}^{N_l}\Bigg\{\frac{1}{2}\left|\frac{g_k\big(X_k^{i,l}\big)}{\sum_{j=1}^{N_l} g_k\big(X_k^{j,l}\big)} - \frac{g_k\big(X_k^{i,l,a}\big)}{\sum_{j=1}^{N_l} g_k\big(X_k^{j,l,a}\big)}\right| + \Bigg|\frac{\frac{1}{2}g_k\big(X_k^{i,l}\big)}{\sum_{j=1}^{N_l} g_k\big(X_k^{j,l}\big)} + \frac{\frac{1}{2}g_k\big(X_k^{i,l,a}\big)}{\sum_{j=1}^{N_l} g_k\big(X_k^{j,l,a}\big)} \\
- \frac{g_k\big(X_k^{i,l-1}\big)}{\sum_{j=1}^{N_l} g_k\big(X_k^{j,l-1}\big)} - \frac{1}{2}\left|\frac{g_k\big(X_k^{i,l}\big)}{\sum_{j=1}^{N_l} g_k\big(X_k^{j,l}\big)} - \frac{g_k\big(X_k^{i,l,a}\big)}{\sum_{j=1}^{N_l} g_k\big(X_k^{j,l,a}\big)}\right|\Bigg|\Bigg\}.
\end{aligned}$$

To shorten the subsequent notation, we set

$$\begin{aligned}
\alpha_i &= \frac{1}{2}\left|\frac{g_k\big(X_k^{i,l}\big)}{\sum_{j=1}^{N_l} g_k\big(X_k^{j,l}\big)} - \frac{g_k\big(X_k^{i,l,a}\big)}{\sum_{j=1}^{N_l} g_k\big(X_k^{j,l,a}\big)}\right| \\
\beta_i &= \frac{\frac{1}{2}g_k\big(X_k^{i,l}\big)}{\sum_{j=1}^{N_l} g_k\big(X_k^{j,l}\big)} + \frac{\frac{1}{2}g_k\big(X_k^{i,l,a}\big)}{\sum_{j=1}^{N_l} g_k\big(X_k^{j,l,a}\big)} - \frac{g_k\big(X_k^{i,l-1}\big)}{\sum_{j=1}^{N_l} g_k\big(X_k^{j,l-1}\big)}.
\end{aligned}$$

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

In this notation, we have that

$$1 - \mathbb{E}\left[\sum_{i=1}^{N_l} \min\left\{\frac{g_k\left(X_k^{i,l}\right)}{\sum_{j=1}^{N_l} g_k\left(X_k^{j,l}\right)} \frac{g_k\left(X_k^{i,l,a}\right)}{\sum_{j=1}^{N_l} g_k\left(X_k^{j,l,a}\right)}, \frac{g_k\left(X_k^{i,l-1}\right)}{\sum_{j=1}^{N_l} g_k\left(X_k^{j,l-1}\right)}\right\}\right]$$

$$= \frac{1}{2}\mathbb{E}\left[\sum_{i=1}^{N_l}\{\alpha_i + |\beta_i - \alpha_i|\}\right] \tag{C.2}$$

$$\le \frac{1}{2}\mathbb{E}\left[\sum_{i \in \mathsf{S}_{k-1}^l} \{\alpha_i + |\beta_i - \alpha_i|\}\right] + C\left(1 - \mathbb{E}\left[\frac{\mathrm{Card}\left(\mathsf{S}_{k-1}^l\right)}{N_l}\right]\right). \tag{C.3}$$

Then by induction, via Lemmata C.3 and C.5, it trivially follows that

$$1 - \mathbb{E}\left[\frac{\mathrm{Card}\left(\mathsf{S}_k^l\right)}{N_l}\right] \le C\Delta_l^{1/2}.$$

□

Remark C.1. Lemma C.6 is the place where the rate is lost. It is clear that this particular quantity cannot fall faster than $\mathcal{O}\left(\Delta_l^{1/2}\right)$, as the right-hand side of (C.2) is exactly equal to

$$\sum_{i=1}^{N_l} \max\{\alpha_i, \beta_i\}.$$

The expectation of this quantity falls exactly at $\mathcal{O}\left(\Delta_l^{1/2}\right)$. We could not find a way to enhance the proof to deal with this point, so we leave it to future work.

Lemma C.7. *Assume* (A1)–(A2). *Then for any* $k \in \mathbb{N}$ *there exists a* $C < +\infty$ *such that for any* $(l, N_l, i) \in \mathbb{N}^2 \times \{1, \dots, N_l\}$,

$$\mathbb{E}\left[\min\left\{\left\|\overline{X}_k^{i,l} - X_k^{i,l-1}\right\|^p, 1\right\}\right] \le C\Delta_l^{1/2}.$$

Proof. This follows from Lemmata C.2 and C.6; see e.g. the corresponding proof in [14, Theorem D.5]. □

Remark C.2. In a similar manner to the proofs of Lemma C.7 and [14, Theorem D.5], one can establish, assuming (A1)–(A2), that the following hold:

- For any $k \in \mathbb{N}$ there exists a $C < +\infty$ such that for any $(l, N_l, i) \in \mathbb{N}^2 \times \{1, \dots, N_l\}$,

$$\mathbb{E}\left[\min\left\{\left\|X_k^{i,l} - X_k^{i,l-1}\right\|^p, 1\right\}\right] \le C\Delta_l^{1/2}.$$

- For any $k \in \mathbb{N}$ there exists a $C < +\infty$ such that for any $(l, N_l, i) \in \mathbb{N}^2 \times \{1, \dots, N_l\}$,

$$\mathbb{E}\left[\min\left\{\left\|X_k^{i,l} - X_k^{i,l,a}\right\|^p, 1\right\}\right] \le C\Delta_l^{1/2}.$$

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

C.2. Particle convergence proofs: $\mathbb{L}_p$-bounds

For a function $\varphi \in \mathcal{B}_b(\mathsf{X}) \cap \mathrm{Lip}(\mathsf{X})$, we set $\overline{\|\varphi\|} = \max\{\|\varphi\|_\infty, \|\varphi\|_{\mathrm{Lip}}\}$, where $\|\cdot\|_{\mathrm{Lip}}$ is the Lipschitz constant. We denote the N_l-empirical measures by

$$\eta_k^{N_l,l}(dx) = \tfrac{1}{N}\sum_{i=1}^{N_l} \delta_{\{X_k^{i,l}\}}(dx),$$

$$\eta_k^{N_l,l-1}(dx) = \tfrac{1}{N}\sum_{i=1}^{N_l} \delta_{\{X_k^{i,l-1}\}}(dx),$$

$$\eta_k^{N_l,l,a}(dx) = \tfrac{1}{N}\sum_{i=1}^{N_l} \delta_{\{X_k^{i,l,a}\}}(dx),$$

where $k \in \mathbb{N}$ and $\delta_A(dx)$ is the Dirac mass on a set A. We define the predictors as follows: for $l \in \mathbb{N}_0$, $\eta_1^l(dx) = P^l(x_0, dx)$, and for $k \geq 2$, $\eta_k^l(dx) = \eta_{k-1}^l(g_{k-1}P^l(dx))/\eta_{k-1}^l(g_{k-1})$.

Lemma C.8. *Assume* (A1)–(A2). *Then for any* $(k, p) \in \mathbb{N} \times [1, \infty)$ *there exists a* $C < +\infty$ *such that for any* $(l, N_l, \varphi) \in \mathbb{N}^2 \times \mathcal{B}_b(\mathsf{X}) \cap Lip(\mathsf{X})$,

$$\mathbb{E}\left[\left|\left[\eta_k^{N_l,l} - \eta_k^{N_l,l-1}\right](\varphi) - \left[\eta_k^l - \eta_k^{l-1}\right](\varphi)\right|^p\right] \leq \frac{C\overline{\|\varphi\|}^p \Delta_l^{\frac{1}{2}}}{N_l^{\frac{p}{2}}}.$$

Proof. The proof is by induction, with initialization following easily from the Marcinkiewicz–Zygmund (MZ) inequality (the case $p \in [1, 2)$ can be dealt with using the bound $p \in [2, \infty)$ and Jensen's inequality) and strong convergence results for Euler discretizations. We therefore proceed to the induction step. We have that

$$\mathbb{E}\left[\left|\left[\eta_k^{N_l,l} - \eta_k^{N_l,l-1}\right](\varphi) - \left[\eta_k^l - \eta_k^{l-1}\right](\varphi)\right|^p\right]^{1/p} \leq T_1 + T_2,$$

where

$$T_1 = \mathbb{E}\left[\left|\left[\eta_k^{N_l,l} - \eta_k^{N_l,l-1}\right](\varphi) - \mathbb{E}\left[\left[\eta_k^{N_l,l} - \eta_k^{N_l,l-1}\right](\varphi)|\mathcal{F}_{k-1}^l\right]\right|^p\right]^{1/p},$$

$$T_2 = \mathbb{E}\left[\left|\mathbb{E}\left[\left[\eta_k^{N_l,l} - \eta_k^{N_l,l-1}\right](\varphi)\middle|\mathcal{F}_{k-1}\right] - \left[\eta_k^l - \eta_k^{l-1}\right](\varphi)\right|^p\right]^{1/p}.$$

We deal with the two terms in turn. For the case of T_1, one can apply the (conditional) MZ inequality to obtain

$$T_1 \leq \frac{C}{\sqrt{N_l}}\mathbb{E}\left[\left|\varphi\left(X_k^{1,l}\right) - \varphi\left(X_k^{1,l-1}\right)\right|^p\right]^{1/p}.$$

Using $\varphi \in \mathcal{B}_b(\mathsf{X}) \cap \mathrm{Lip}(\mathsf{X})$ along with Remark C.2 yields

$$T_1 \leq \frac{C\overline{\|\varphi\|}\Delta_l^{\frac{1}{2p}}}{\sqrt{N_l}}. \tag{C.4}$$

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

For T_2, we first note that

$$\mathbb{E}\Big[\Big[\eta_k^{N_l,l}-\eta_k^{N_l,l-1}\Big](\varphi)|\mathcal{F}_{k-1}^l\Big]-\big[\eta_k^l-\eta_k^{l-1}\big](\varphi)$$

$$=\left\{\frac{\eta_{k-1}^{N_l,l}\big(g_{k-1}P^l(\varphi)\big)}{\eta_{k-1}^{N_l,l}(g_{k-1})}-\frac{\eta_{k-1}^{N_l,l-1}\big(g_{k-1}P^{l-1}(\varphi)\big)}{\eta_{k-1}^{N_l,l-1}(g_{k-1})}\right\}-\left\{\frac{\eta_{k-1}^{l}\big(g_{k-1}P^l(\varphi)\big)}{\eta_{k-1}^{l}(g_{k-1})}-\frac{\eta_{k-1}^{l-1}\big(g_{k-1}P^{l-1}(\varphi)\big)}{\eta_{k-1}^{l-1}(g_{k-1})}\right\},$$

where we recall that P^l (resp. P^{l-1}) represents the truncated Milstein kernel over unit time with discretization level Δ_l (resp. Δ_{l-1}). This can be further decomposed to

$$\mathbb{E}\Big[\Big[\eta_k^{N_l,l}-\eta_k^{N_l,l-1}\Big](\varphi)|\mathcal{F}_{k-1}\Big]-\big[\eta_k^l-\eta_k^{l-1}\big](\varphi)$$

$$\begin{aligned}&=\left\{\frac{\eta_{k-1}^{N_l,l}\big(g_{k-1}[P^l-P](\varphi)\big)}{\eta_{k-1}^{N_l,l}(g_{k-1})}-\frac{\eta_{k-1}^{l}\big(g_{k-1}[P^l-P](\varphi)\big)}{\eta_{k-1}^{l}(g_{k-1})}\right\}\\&-\left\{\frac{\eta_{k-1}^{l}\big(g_{k-1}[P^{l-1}-P](\varphi)\big)}{\eta_{k-1}^{l}(g_{k-1})}-\frac{\eta_{k-1}^{l-1}\big(g_{k-1}[P^{l-1}-P](\varphi)\big)}{\eta_{k-1}^{l-1}(g_{k-1})}\right\}\end{aligned}$$

$$+\left\{\frac{\eta_{k-1}^{N_l,l}\big(g_{k-1}P(\varphi)\big)}{\eta_{k-1}^{N_l,l}(g_{k-1})}-\frac{\eta_{k-1}^{N_l,l-1}\big(g_{k-1}P(\varphi)\big)}{\eta_{k-1}^{N_l,l-1}(g_{k-1})}\right\}-\left\{\frac{\eta_{k-1}^{l}\big(g_{k-1}P(\varphi)\big)}{\eta_{k-1}^{l}(g_{k-1})}-\frac{\eta_{k-1}^{l-1}\big(g_{k-1}P(\varphi)\big)}{\eta_{k-1}^{l-1}(g_{k-1})}\right\}.$$

Therefore we have that

$$T_2\le T_3+T_4+T_5,$$

where

$$T_3=\mathbb{E}\left[\left|\frac{\eta_{k-1}^{N_l,l}\big(g_{k-1}[P^l-P](\varphi)\big)}{\eta_{k-1}^{N_l,l}(g_{k-1})}-\frac{\eta_{k-1}^{l}\big(g_{k-1}[P^l-P](\varphi)\big)}{\eta_{k-1}^{l}(g_{k-1})}\right|^p\right]^{1/p}, \tag{C.5}$$

$$T_4=\mathbb{E}\left[\left|\frac{\eta_{k-1}^{N_l,l-1}\big(g_{k-1}[P^{l-1}-P](\varphi)\big)}{\eta_{k-1}^{N_l,l-1}(g_{k-1})}-\frac{\eta_{k-1}^{l-1}\big(g_{k-1}[P^{l-1}-P](\varphi)\big)}{\eta_{k-1}^{l-1}(g_{k-1})}\right|^p\right]^{1/p},$$

$$\begin{aligned}T_5=\mathbb{E}\Bigg[\Bigg|&\left\{\frac{\eta_{k-1}^{N_l,l}\big(g_{k-1}P(\varphi)\big)}{\eta_{k-1}^{N_l,l}(g_{k-1})}-\frac{\eta_{k-1}^{N_l,l-1}\big(g_{k-1}P(\varphi)\big)}{\eta_{k-1}^{N_l,l-1}(g_{k-1})}\right\}-\\&\left\{\frac{\eta_{k-1}^{l}\big(g_{k-1}P(\varphi)\big)}{\eta_{k-1}^{l}(g_{k-1})}-\frac{\eta_{k-1}^{l-1}\big(g_{k-1}P(\varphi)\big)}{\eta_{k-1}^{l-1}(g_{k-1})}\right\}\Bigg|^p\Bigg]^{1/p}.\end{aligned}$$

The quantities T_3 and T_4 can be treated using similar calculations, so we only consider T_3. For the case of T_3, clearly we have the upper bound

$$T_3\le T_6+T_7,$$

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

where

$$T_6 = \mathbb{E}\left[\left|\frac{\eta_{k-1}^{N_l,l}\big(g_{k-1}\big[P^l - P\big](\varphi)\big) - \eta_{k-1}^{l}\big(g_{k-1}\big[P^l - P\big](\varphi)\big)}{\eta_{k-1}^{N_l,l}(g_{k-1})}\right|^p\right]^{1/p},$$

$$T_7 = \mathbb{E}\left[\left|\frac{\eta_{k-1}^{l}\big(g_{k-1}\big[P^l - P\big](\varphi)\big)\Big[\eta_{k-1}^{l} - \eta_{k-1}^{N_l,l}\Big](g_{k-1})}{\eta_{k-1}^{N_l,l}(g_{k-1})\eta_{k-1}^{l}(g_{k-1})}\right|^p\right]^{1/p}.$$

For T_6 we have that

$$T_6 \le C\mathbb{E}\left[\left|\big[\eta_{k-1}^{N_l,l} - \eta_{k-1}^{l}\big]\big(g_{k-1}\big[P^l - P\big](\varphi)\big)\right|^p\right]^{1/p}.$$

By using standard results for particle filters (e.g. [14, Proposition C.6]) we have

$$T_6 \le \frac{C\|\big[P^l - P\big](\varphi)\|_\infty}{\sqrt{N_l}}.$$

Then, using standard results for weak errors (the truncated Milstein scheme is a first-order method), we have

$$T_7 \le \frac{C\overline{\|\varphi\|}\Delta_l}{\sqrt{N_l}}.$$

For T_7, again using weak errors, we have

$$T_7 \le C\overline{\|\varphi\|}\Delta_l\mathbb{E}\left[\left|\Big[\eta_{k-1}^{l} - \eta_{k-1}^{N_l,l}\Big](g_{k-1})\right|^p\right]^{1/p},$$

and again using standard results for particle filters, we obtain

$$T_7 \le \frac{C\overline{\|\varphi\|}\Delta_l}{\sqrt{N_l}}.$$

Hence we have shown that

$$\max\{T_3, T_4\} \le \frac{C\overline{\|\varphi\|}\Delta_l}{\sqrt{N_l}}. \tag{C.6}$$

For T_5, one can apply [14, Lemma C.5] to show that

$$T_5 \le \sum_{j=1}^{6} T_{j+7},$$

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

where

$$T_8 = \mathbb{E}\left[\left|\frac{1}{\eta_{k-1}^{N_l,l}(g_{k-1})}\left\{\left[\eta_{k-1}^{N_l,l} - \eta_{k-1}^{N_l,l-1}\right]\Big(g_{k-1}P(\varphi)\Big) - \left[\eta_{k-1}^{l} - \eta_{k-1}^{l-1}\right]\Big(g_{k-1}P(\varphi)\Big)\right\}\right|^p\right]^{1/p},$$

$$T_9 = \mathbb{E}\left[\left|\frac{\eta_{k-1}^{N_l,l-1}\Big(g_{k-1}P(\varphi)\Big)}{\eta_{k-1}^{N_l,l}(g_{k-1})\eta_{k-1}^{N_l,l-1}(g_{k-1})}\left\{\left[\eta_{k-1}^{N_l,l} - \eta_{k-1}^{N_l,l-1}\right](g_{k-1}) - \left[\eta_{k-1}^{l} - \eta_{k-1}^{l-1}\right](g_{k-1})\right\}\right|^p\right]^{1/p},$$

$$T_{10} = \mathbb{E}\left[\left|\frac{1}{\eta_{k-1}^{N_l,l}(g_{k-1})\eta_{k-1}^{l}(g_{k-1})}\left\{\left[\eta_{k-1}^{l} - \eta_{k-1}^{N_l,l}\right](g_{k-1})\right\}\left\{\left[\eta_{k-1}^{l} - \eta_{k-1}^{l-1}\right]\Big(g_{k-1}P(\varphi)\Big)\right\}\right|^p\right]^{1/p},$$

$$T_{11} = \mathbb{E}\left[\left|\frac{1}{\eta_{k-1}^{N_l,l}(g_{k-1})\eta_{k-1}^{N_l,l-1}(g_{k-1})}\left\{\left[\eta_{k-1}^{N_l,l-1} - \eta_{k-1}^{l-1}\right]\Big(g_{k-1}P(\varphi)\Big)\right\}\left\{\left[\eta_{k-1}^{l} - \eta_{k-1}^{l-1}\right](g_{k-1})\right\}\right|^p\right]^{1/p},$$

$$T_{12} = \mathbb{E}\left[\left|\frac{\eta_{k-1}^{l-1}\Big(g_{k-1}P(\varphi)\Big)}{\eta_{k-1}^{l}(g_{k-1})\eta_{k-1}^{N_l,l-1}(g_{k-1})\eta_{k-1}^{l-1}(g_{k-1})}\left\{\left[\eta_{k-1}^{N_l,l-1} - \eta_{k-1}^{l-1}\right](g_{k-1})\right\}\left\{\left[\eta_{k-1}^{l} - \eta_{k-1}^{l-1}\right](g_{k-1})\right\}\right|^p\right]^{1/p},$$

$$T_{13} = \mathbb{E}\left[\left|\frac{\eta_{k-1}^{l-1}\Big(g_{k-1}P(\varphi)\Big)}{\eta_{k-1}^{N_l,l}(g_{k-1})\eta_{k-1}^{l}(g_{k-1})\eta_{k-1}^{N_l,l-1}(g_{k-1})}\left\{\left[\eta_{k-1}^{N_l,l} - \eta_{k-1}^{l}\right](g_{k-1})\right\}\left\{\left[\eta_{k-1}^{l} - \eta_{k-1}^{l-1}\right](g_{k-1})\right\}\right|^p\right]^{1/p}.$$

Since T_8 and T_9 can be bounded using similar approaches, we consider only T_8. Similarly, $T_{10}, \ldots, T_{13}$ can be bounded in almost the same way, so we consider only T_{10}. For T_8 we have the upper bound

$$T_8 \leq C\mathbb{E}\left[\left|\left\{\left[\eta_{k-1}^{N_l,l} - \eta_{k-1}^{N_l,l-1}\right]\big(g_{k-1}P(\varphi)\big) - \left[\eta_{k-1}^{l} - \eta_{k-1}^{l-1}\right]\big(g_{k-1}P(\varphi)\big)\right\}\right|^p\right]^{1/p}.$$

Since $g_{k-1}P(\varphi) \in \mathcal{B}_b(\mathsf{X}) \cap \mathrm{Lip}(\mathsf{X})$ (see e.g. [7, Equation (2.6)]), it follows by the induction hypothesis that

$$T_8 \leq \frac{C\overline{\|\varphi\|}\Delta_l^{\frac{1}{2p}}}{\sqrt{N_l}}.$$

For T_{10}, we have the upper bound

$$T_{10} \leq C\left|\left[\eta_{k-1}^{l} - \eta_{k-1}^{l-1}\right]\big(g_{k-1}P(\varphi)\big)\right|\mathbb{E}\left[\left|\left(\left[\eta_{k-1}^{l} - \eta_{k-1}^{N_l,l}\right](g_{k-1})\right)\right|^p\right]^{1/p}.$$

Then, using [14, Lemma D.2] and standard results for particle filters, we have

$$T_{10} \leq \frac{C\overline{\|\varphi\|}\Delta_l}{\sqrt{N_l}}.$$

Therefore we deduce that

$$T_5 \leq \frac{C\overline{\|\varphi\|}\Delta_l^{\frac{1}{2p}}}{\sqrt{N_l}}. \tag{C.7}$$

The proof is then easily completed by combining (C.4), (C.6), and (C.7). □

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

Remark C.3. One can also deduce the following result using an argument that is similar to (but simpler than) the proof of Lemma C.8. Assume (A1)–(A2). Then for any $(k, p) \in \mathbb{N} \times [1, \infty)$ there exists a $C < +\infty$ such that for any $(l, N_l, \varphi) \in \mathbb{N}^2 \times \mathcal{B}_b(\mathsf{X}) \cap Lip(\mathsf{X})$,

$$\mathbb{E}\left[\left|\left[\eta_k^{N_l,l} - \eta_k^{N_l,l,a}\right](\varphi)\right|^p\right] \leq \frac{C\overline{\|\varphi\|}^p \Delta_l^{\frac{1}{2}}}{N_l^{\frac{p}{2}}}.$$

Below, $\mathcal{C}_b^2(\mathsf{X}, \mathbb{R})$ denotes the collection of twice continuously differentiable functions from X to $\mathbb{R}$ with bounded derivatives of orders 1 and 2.

Lemma C.9. *Assume* (A1)–(A2). *Then for any* $(k, p, \varphi) \in \mathbb{N} \times [1, \infty) \times \mathcal{B}_b(\mathsf{X}) \cap \mathcal{C}_b^2(\mathsf{X}, \mathbb{R})$ *there exists a* $C < +\infty$ *such that for any* $(l, N_l) \in \mathbb{N}^2$,

$$\mathbb{E}\left[\left|\left[\tfrac{1}{2}\eta_k^{N_l,l} + \tfrac{1}{2}\eta_k^{N,l,a} - \eta_k^{N_l,l-1}\right](\varphi) - \left[\eta_k^{l} - \eta_k^{l-1}\right](\varphi)\right|^p\right] \leq C\left(\frac{\Delta_l^{\frac{1}{2}}}{N_l^{p/2}} + \frac{\Delta_l^{\frac{1}{2}}}{N_l^{p}}\right).$$

Proof. The proof is by induction. Initialization follows easily by the MZ inequality and [11, Lemma 2.2, Theorem 4.10] (the case $p \in [1, 2)$ can also be covered with the theory of [11]), so we proceed immediately to the induction step.

As in the proof of Lemma C.8, one can add and subtract the conditional expectation to obtain an upper bound

$$\mathbb{E}\left[\left|\left[\tfrac{1}{2}\eta_k^{N_l,l} + \tfrac{1}{2}\eta_k^{N,l,a} - \eta_k^{N_l,l-1}\right](\varphi) - \left[\eta_k^{l} - \eta_k^{l-1}\right](\varphi)\right|^p\right]^{1/p} \leq T_1 + T_2,$$

where

$$T_1 = \mathbb{E}\left[\left|\left[\tfrac{1}{2}\eta_k^{N_l,l} + \tfrac{1}{2}\eta_k^{N,l,a} - \eta_k^{N_l,l-1}\right](\varphi) - \mathbb{E}\left[\tfrac{1}{2}\eta_k^{N_l,l} + \tfrac{1}{2}\eta_k^{N,l,a} - \eta_k^{N_l,l-1}\right](\varphi)|\mathcal{F}_{k-1}]\right|^p\right]^{1/p},$$
$$T_2 = \mathbb{E}\left[\left|\mathbb{E}\left[\tfrac{1}{2}\eta_k^{N_l,l} + \tfrac{1}{2}\eta_k^{N,l,a} - \eta_k^{N_l,l-1}\right](\varphi)|\mathcal{F}_{k-1}] - \left[\eta_k^{l} - \eta_k^{l-1}\right](\varphi)\right|^p\right]^{1/p}.$$

For T_1 one can use the conditional MZ inequality along with the boundedness of φ, [11, Lemma 2.2], Lemma C.7, and Remark C.2 to deduce that

$$T_1 \leq \frac{C\Delta_l^{1/(2p)}}{\sqrt{N_l}}. \tag{C.8}$$

The case of T_2 is more challenging. We have the decomposition

$$T_2 \leq T_3 + T_4,$$

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

where

$$T_3 = \mathbb{E}\Bigg[\Bigg|\Bigg\{\frac{\frac{1}{2}\eta_{k-1}^{N_l,l}\big(g_{k-1}\big[P^l-P\big](\varphi)\big)}{\eta_{k-1}^{N_l,l}(g_{k-1})} + \frac{\frac{1}{2}\eta_{k-1}^{N_l,l,a}\big(g_{k-1}\big[P^l-P\big](\varphi)\big)}{\eta_{k-1}^{N_l,l,a}(g_{k-1})} - \frac{\eta_{k-1}^{N_l,l-1}\big(g_{k-1}\big[P^{l-1}-P\big](\varphi)\big)}{\eta_{k-1}^{N_l,l-1}(g_{k-1})}\Bigg\}$$
$$-\Bigg\{\frac{\frac{1}{2}\eta_{k-1}^{l}\big(g_{k-1}\big[P^l-P\big](\varphi)\big)}{\eta_{k-1}^{l}(g_{k-1})} + \frac{\frac{1}{2}\eta_{k-1}^{l}\big(g_{k-1}\big[P^l-P\big](\varphi)\big)}{\eta_{k-1}^{l}(g_{k-1})} - \frac{\eta_{k-1}^{l-1}\big(g_{k-1}\big[P^{l-1}-P\big](\varphi)\big)}{\eta_{k-1}^{l-1}(g_{k-1})}\Bigg\}\Bigg|^p\Bigg]^{1/p},$$
$$T_4 = \mathbb{E}\Bigg[\Bigg|\Bigg\{\frac{\frac{1}{2}\eta_{k-1}^{N_l,l}\big(g_{k-1}P(\varphi)\big)}{\eta_{k-1}^{N_l,l}(g_{k-1})} + \frac{\frac{1}{2}\eta_{k-1}^{N_l,l,a}\big(g_{k-1}P(\varphi)\big)}{\eta_{k-1}^{N_l,l,a}(g_{k-1})} - \frac{\eta_{k-1}^{N_l,l-1}\big(g_{k-1}P(\varphi)\big)}{\eta_{k-1}^{N_l,l-1}(g_{k-1})}\Bigg\}$$
$$-\Bigg\{\frac{\frac{1}{2}\eta_{k-1}^{l}\big(g_{k-1}P(\varphi)\big)}{\eta_{k-1}^{l}(g_{k-1})} + \frac{\frac{1}{2}\eta_{k-1}^{l}\big(g_{k-1}P(\varphi)\big)}{\eta_{k-1}^{l}(g_{k-1})} - \frac{\eta_{k-1}^{l-1}\big(g_{k-1}P(\varphi)\big)}{\eta_{k-1}^{l-1}(g_{k-1})}\Bigg\}\Bigg|^p\Bigg]^{1/p}. \tag{C.9}$$

For T_3 one has the upper bound

$$T_3 \le \sum_{j=1}^{3} T_{j+4},$$

where

$$T_5 = \mathbb{E}\Bigg[\Bigg|\Bigg\{\frac{\frac{1}{2}\eta_{k-1}^{N_l,l}\big(g_{k-1}\big[P^l-P\big](\varphi)\big)}{\eta_{k-1}^{N_l,l}(g_{k-1})} - \frac{\frac{1}{2}\eta_{k-1}^{l}\big(g_{k-1}\big[P^l-P\big](\varphi)\big)}{\eta_{k-1}^{l}(g_{k-1})}\Bigg|^p\Bigg]^{1/p},$$
$$T_6 = \mathbb{E}\Bigg[\Bigg|\Bigg\{\frac{\frac{1}{2}\eta_{k-1}^{N_l,l,a}\big(g_{k-1}\big[P^l-P\big](\varphi)\big)}{\eta_{k-1}^{N_l,l,a}(g_{k-1})} - \frac{\frac{1}{2}\eta_{k-1}^{l}\big(g_{k-1}\big[P^l-P\big](\varphi)\big)}{\eta_{k-1}^{l}(g_{k-1})}\Bigg|^p\Bigg]^{1/p},$$
$$T_7 = \mathbb{E}\Bigg[\Bigg|\Bigg\{\frac{\frac{1}{2}\eta_{k-1}^{N_l,l-1}\big(g_{k-1}\big[P^{l-1}-P\big](\varphi)\big)}{\eta_{k-1}^{N_l,l-1}(g_{k-1})} - \frac{\frac{1}{2}\eta_{k-1}^{l-1}\big(g_{k-1}\big[P^{l-1}-P\big](\varphi)\big)}{\eta_{k-1}^{l-1}(g_{k-1})}\Bigg|^p\Bigg]^{1/p}.$$

Each of these terms can be controlled (almost) exactly as is done for (C.5) in the proof of Lemma C.8, so we do not give the proof; rather, we simply state that

$$T_3 \le \frac{C\Delta_l^{1/2}}{\sqrt{N_l}}. \tag{C.10}$$

For T_4 one can apply Lemma C.4 along with Minkowski to deduce that

$$T_4 \le \sum_{j=1}^{4} T_{j+7},$$

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

where

$$T_8 = \mathbb{E}\Bigg[\Bigg|\frac{1}{\eta_{k-1}^{N_l,l-1}(g_{k-1})}\Big[\tfrac{1}{2}\eta_{k-1}^{N_l,l} + \tfrac{1}{2}\eta_{k-1}^{N_l,l,a} - \eta_{k-1}^{N_l,l-1}\Big]\big(g_{k-1}P(\varphi)\big) - \frac{1}{\eta_{k-1}^{l-1}(g_{k-1})}\Big[\tfrac{1}{2}\eta_{k-1}^{l} + \tfrac{1}{2}\eta_{k-1}^{l} - \eta_{k-1}^{l-1}\Big]\big(g_{k-1}P(\varphi)\big)\Bigg|^p\Bigg]^{1/p},$$

$$T_9 = \mathbb{E}\Bigg[\Bigg|\frac{1}{\eta_{k-1}^{N_l,l}(g_{k-1})\eta_{k-1}^{N_l,l-1}(g_{k-1})}\frac{1}{2}\Big\{\Big[\eta_{k-1}^{N_l,l} - \eta_{k-1}^{N_l,l,a}\Big]\big(g_{k-1}P(\varphi)\big)\Big\}\Big\{\Big[\eta_{k-1}^{N_l,l-1} - \eta_{k-1}^{N_l,l}\Big](g_{k-1})\Big\}\Bigg|^p\Bigg]^{1/p},$$

$$T_{10} = \mathbb{E}\Bigg[\Bigg|\frac{\frac{1}{2}\eta_{k-1}^{N_l,l,a}\big(g_{k-1}P(\varphi)\big)}{\eta_{k-1}^{N_l,l,a}(g_{k-1})\eta_{k-1}^{N_l,l}(g_{k-1})\eta_{k-1}^{N_l,l-1}](g_{k-1})}\Big\{\Big[\eta_{k-1}^{N_l,l,a} - \eta_{k-1}^{N_l,l}\Big](g_{k-1})\Big\}\Big\{\Big[\eta_{k-1}^{N_l,l-1} - \eta_{k-1}^{N_l,l}\Big](g_{k-1})\Big\}\Bigg|^p\Bigg]^{1/p},$$

$$T_{11} = \mathbb{E}\Bigg[\Bigg|\frac{\eta_{k-1}^{N_l,l,a}\big(g_{k-1}P(\varphi)\big)}{\eta_{k-1}^{N_l,l,a}(g_{k-1})\eta_{k-1}^{N_l,l-1}(g_{k-1})}\Big[\tfrac{1}{2}\eta_{k-1}^{N_l,l} + \tfrac{1}{2}\eta_{k-1}^{N_l,l,a} - \eta_{k-1}^{N_l,l-1}\Big](g_{k-1}) - \frac{\eta_{k-1}^{l}\big(g_{k-1}P(\varphi)\big)}{\eta_{k-1}^{l}(g_{k-1})\eta_{k-1}^{l-1}(g_{k-1})} \times \Big[\tfrac{1}{2}\eta_{k-1}^{l} + \tfrac{1}{2}\eta_{k-1}^{l} - \eta_{k-1}^{l-1}\Big](g_{k-1})\Bigg|^p\Bigg]^{1/p}.$$

As the arguments for dealing with T_8 (resp. T_9) and T_{11} (resp. T_{10}) are similar, we only prove bounds for T_8 (resp. T_9). For T_8 we have the upper bound

$$T_8 \le T_{12} + T_{13},$$

where

$$T_{12} = \mathbb{E}\Bigg[\Bigg|\frac{1}{\eta_{k-1}^{N_l,l-1}(g_{k-1})}\Bigg\{\Big[\tfrac{1}{2}\eta_{k-1}^{N_l,l} + \tfrac{1}{2}\eta_{k-1}^{N_l,l,a} - \eta_{k-1}^{N_l,l-1}\Big]\big(g_{k-1}P(\varphi)\big) - \Big[\tfrac{1}{2}\eta_{k-1}^{l} + \tfrac{1}{2}\eta_{k-1}^{l} - \eta_{k-1}^{l-1}\Big]\big(g_{k-1}P(\varphi)\big)\Bigg\}\Bigg|^p\Bigg]^{1/p},$$

$$T_{13} = \mathbb{E}\Bigg[\Bigg|\frac{\Big[\frac{1}{2}\eta_{k-1}^{l} + \frac{1}{2}\eta_{k-1}^{l} - \eta_{k-1}^{l-1}\Big]\big(g_{k-1}P(\varphi)\big)}{\eta_{k-1}^{N_l,l-1}(g_{k-1})\eta_{k-1}^{l-1}(g_{k-1})}\Big[\eta_{k-1}^{l-1} - \eta_{k-1}^{N_l,l-1}\Big](g_{k-1})\Bigg|^p\Bigg]^{1/p}.$$

For T_{12} one can use the lower bound on g_{k-1} along with $g_{k-1}P(\varphi) \in \mathcal{C}_b^2(\mathsf{X}, \mathbb{R})$ (see e.g. [23, Corollary 2.2.8]) and the induction hypothesis to obtain

$$T_{12} \le \frac{C\Delta_l^{1/(2p)}}{\sqrt{N_l}}.$$

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

For T_{13} we apply [14, Lemma D.2] and standard results for particle filters to obtain

$$T_{13} \le \frac{C\Delta_l^{1/2}}{\sqrt{N_l}}.$$

Thus

$$T_8 \le \frac{C\Delta_l^{1/(2p)}}{\sqrt{N_l}}.$$

For T_9 we have the upper bound

$$T_9 \le T_{14} + T_{15},$$

where

$$T_{14} = \mathbb{E}\left[\left|\frac{1}{\eta_{k-1}^{N_l,l}(g_{k-1})\eta_{k-1}^{N_l,l-1}(g_{k-1})}\frac{1}{2}\left\{\left[\eta_{k-1}^{N_l,l} - \eta_{k-1}^{N_l,l,a}\right](g_{k-1}P(\varphi))\right\}\left\{\left[\eta_{k-1}^{N_l,l-1} - \eta_{k-1}^{N_l,l}\right](g_{k-1}) - \left[\eta_{k-1}^{l-1} - \eta_{k-1}^{l}\right](g_{k-1})\right\}\right|^p\right]^{1/p},$$

$$T_{15} = \mathbb{E}\left[\left|\frac{1}{\eta_{k-1}^{N_l,l}(g_{k-1})\eta_{k-1}^{N_l,l-1}(g_{k-1})}\frac{1}{2}\left\{\left[\eta_{k-1}^{N_l,l} - \eta_{k-1}^{N_l,l,a}\right](g_{k-1}P(\varphi))\right\}\left\{\left[\eta_{k-1}^{l-1} - \eta_{k-1}^{l}\right](g_{k-1})\right\}\right|^p\right]^{1/p}.$$

For T_{14} one can use the lower bound on g_{k-1}, Cauchy–Schwarz, Remark C.3, and Lemma C.8 to obtain

$$T_{14} \le \frac{C\Delta_l^{1/(2p)}}{N_l}.$$

For T_{15} one can use the lower bound on g_{k-1}, [14, Lemma D.2], and Remark C.3 to obtain

$$T_{15} \le \frac{C}{\sqrt{N_l}}\left(\Delta_l^{p+1/2}\right)^{1/p};$$

thus,

$$T_9 \le \frac{C\Delta_l^{1/(2p)}}{\sqrt{N_l}} + \frac{C\Delta_l^{1/(2p)}}{N_l}.$$

Therefore we have shown that

$$T_4 \le \frac{C}{\sqrt{N_l}}\left(\Delta_l^{1/2} + \frac{\Delta_l^{\frac{1}{2}}}{N_l^{p/2}}\right)^{1/p}. \tag{C.11}$$

The proof can be completed by combining the bounds (C.8), (C.10), and (C.11). □

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

Theorem 2. *Assume* (A1)–(A2). *Then for any* $(k, p, \varphi) \in \mathbb{N} \times [1, \infty) \times \mathcal{B}_b(\mathsf{X}) \cap \mathcal{C}_b^2(\mathsf{X}, \mathbb{R})$ *there exists a* $C < +\infty$ *such that for any* $(l, N_l) \in \mathbb{N}^2$,

$$\mathbb{E}\left[\left|\left[\pi_k^l - \pi_k^{l-1}\right]^{N_l}(\varphi) - \left[\pi_k^l - \pi_k^{l-1}\right](\varphi)\right|^p\right] \leq C\left(\frac{\Delta_l^{1/2}}{N_l^{p/2}} + \frac{\Delta_l^{\frac{1}{2}}}{N_l^p}\right).$$

Proof. This follows from Lemma C.8, Lemma C.9, Remark C.3, and a similar approach to controlling (C.9) as in the proof of Lemma C.9; we omit the details. □

C.3. Particle convergence proofs: bias bounds

Lemma C.10. *Assume* (A1)–(A2). *Then for any* $k \in \mathbb{N}$ *there exists a* $C < +\infty$ *such that for any* $(l, N_l, \varphi) \in \mathbb{N}^2 \times \mathcal{B}_b(\mathsf{X}) \cap \mathit{Lip}(\mathsf{X})$,

$$\left|\mathbb{E}\left[\left[\eta_k^{N_l,l} - \eta_k^{N_l,l-1}\right](\varphi) - \left[\eta_k^l - \eta_k^{l-1}\right](\varphi)\right]\right| \leq \frac{C\overline{\|\varphi\|}\Delta_l^{\frac{1}{4}}}{N_l}.$$

Proof. The proof is by induction. The case $k = 1$ is trivial as there is no bias, so we proceed to the induction step. Following the proof of Lemma C.8, we have the decomposition

$$\left|\mathbb{E}\left[\left[\eta_k^{N_l,l} - \eta_k^{N_l,l-1}\right](\varphi) - \left[\eta_k^l - \eta_k^{l-1}\right](\varphi)\right]\right| \leq T_1 + T_2 + T_3,$$

where

$$T_1 = \left|\mathbb{E}\left[\frac{\eta_{k-1}^{N_l,l}\left(g_{k-1}\left[P^l - P\right](\varphi)\right)}{\eta_{k-1}^{N_l,l}(g_{k-1})} - \frac{\eta_{k-1}^l\left(g_{k-1}\left[P^l - P\right](\varphi)\right)}{\eta_{k-1}^l(g_{k-1})}\right]\right|,$$

$$T_2 = \left|\mathbb{E}\left[\frac{\eta_{k-1}^{N_l,l-1}\left(g_{k-1}\left[P^{l-1} - P\right](\varphi)\right)}{\eta_{k-1}^{N_l,l-1}(g_{k-1})} - \frac{\eta_{k-1}^{l-1}\left(g_{k-1}\left[P^{l-1} - P\right](\varphi)\right)}{\eta_{k-1}^{l-1}(g_{k-1})}\right]\right|,$$

$$T_3 = \left|\mathbb{E}\left[\left\{\frac{\eta_{k-1}^{N_l,l}\left(g_{k-1}P(\varphi)\right)}{\eta_{k-1}^{N_l,l}(g_{k-1})} - \frac{\eta_{k-1}^{N_l,l-1}\left(g_{k-1}P(\varphi)\right)}{\eta_{k-1}^{N_l,l-1}(g_{k-1})}\right\} - \left\{\frac{\eta_{k-1}^l\left(g_{k-1}P(\varphi)\right)}{\eta_{k-1}^l(g_{k-1})} - \frac{\eta_{k-1}^{l-1}\left(g_{k-1}P(\varphi)\right)}{\eta_{k-1}^{l-1}(g_{k-1})}\right\}\right]\right|.$$

T_1 and T_2 can be bounded using almost the same calculations, so we consider only T_1. The latter can easily be bounded above by $\sum_{j=1}^4 T_{j+3}$, where

$$T_4 = \left|\mathbb{E}\left[\left\{\frac{1}{\eta_{k-1}^{N_l,l}(g_{k-1})} - \frac{1}{\eta_{k-1}^l(g_{k-1})}\right\}\left[\eta_{k-1}^{N_l,l} - \eta_{k-1}^l\right]\left(g_{k-1}[P^l - P](\varphi)\right)\right]\right|,$$

$$T_5 = \left|\mathbb{E}\left[\frac{1}{\eta_{k-1}^l(g_{k-1})}\left[\eta_{k-1}^{N_l,l} - \eta_{k-1}^l\right]\left(g_{k-1}[P^l - P](\varphi)\right)\right]\right|,$$

$$T_6 = \left|\mathbb{E}\left[\eta_{k-1}^l\left(g_{k-1}[P^l - P](\varphi)\right)\left\{\frac{1}{\eta_{k-1}^{N_l,l}(g_{k-1})\eta_{k-1}^l(g_{k-1})} - \frac{1}{\eta_{k-1}^l(g_{k-1})^2}\right\}\left[\eta_{k-1}^l - \eta_{k-1}^{N_l,l}\right](g_{k-1})\right]\right|,$$

$$T_7 = \left|\mathbb{E}\left[\frac{\eta_{k-1}^l\left(g_{k-1}[P^l - P](\varphi)\right)}{\eta_{k-1}^l(g_{k-1})^2}\left[\eta_{k-1}^l - \eta_{k-1}^{N_l,l}\right](g_{k-1})\right]\right|.$$

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

For both T_4 and T_6, one can use the Cauchy–Schwarz inequality, along with [14, Proposition C.6], the lower bound on g_{k-1}, and weak-error results for diffusions to show that

$$\max\{T_4, T_6\} \leq \frac{C\overline{\|\varphi\|}\Delta_l}{N_l}.$$

For T_5 and T_7 one can use standard bias bounds for particle filters (see e.g. the proof of [14, Lemma C.3]), the lower bound on g_{k-1}, and weak-error results for diffusions to obtain

$$\max\{T_5, T_7\} \leq \frac{C\overline{\|\varphi\|}\Delta_l}{N_l};$$

hence

$$\max\{T_1, T_2\} \leq \frac{C\overline{\|\varphi\|}\Delta_l}{N_l}. \tag{C.12}$$

For T_3, using [14, Lemma C.5] we have the upper bound $T_3 \leq \sum_{j=1}^{6} T_{j+7}$, where

$$T_8 = \left|\mathbb{E}\left[\frac{1}{\eta_{k-1}^{N_l,l}(g_{k-1})}\left\{\left[\eta_{k-1}^{N_l,l} - \eta_{k-1}^{N_l,l-1}\right]\big(g_{k-1}P(\varphi)\big) - \left[\eta_{k-1}^{l} - \eta_{k-1}^{l-1}\right]\big(g_{k-1}P(\varphi)\big)\right\}\right]\right|,$$

$$T_9 = \left|\mathbb{E}\left[\frac{\eta_{k-1}^{N_l,l-1}\big(g_{k-1}P(\varphi)\big)}{\eta_{k-1}^{N_l,l}(g_{k-1})\eta_{k-1}^{N_l,l-1}(g_{k-1})}\left\{\left[\eta_{k-1}^{N_l,l} - \eta_{k-1}^{N_l,l-1}\right](g_{k-1}) - \left[\eta_{k-1}^{l} - \eta_{k-1}^{l-1}\right](g_{k-1})\right\}\right]\right|,$$

$$T_{10} = \left|\mathbb{E}\left[\frac{1}{\eta_{k-1}^{N_l,l}(g_{k-1})\eta_{k-1}^{l}(g_{k-1})}\left\{\left[\eta_{k-1}^{l} - \eta_{k-1}^{N_l,l}\right](g_{k-1})\right\}\left\{\left[\eta_{k-1}^{l} - \eta_{k-1}^{l-1}\right]\big(g_{k-1}P(\varphi)\big)\right\}\right]\right|,$$

$$T_{11} = \left|\mathbb{E}\left[\frac{1}{\eta_{k-1}^{N_l,l}(g_{k-1})\eta_{k-1}^{N_l,l-1}(g_{k-1})}\left\{\left[\eta_{k-1}^{N_l,l-1} - \eta_{k-1}^{l-1}\right]\big(g_{k-1}P(\varphi)\big)\right\}\left\{\left[\eta_{k-1}^{l} - \eta_{k-1}^{l-1}\right](g_{k-1})\right\}\right]\right|,$$

$$T_{12} = \left|\mathbb{E}\left[\frac{\eta_{k-1}^{l-1}\big(g_{k-1}P(\varphi)\big)}{\eta_{k-1}^{l}(g_{k-1})\eta_{k-1}^{N_l,l-1}(g_{k-1})\eta_{k-1}^{l-1}(g_{k-1})}\left\{\left[\eta_{k-1}^{N_l,l-1} - \eta_{k-1}^{l-1}\right](g_{k-1})\right\}\left\{\left[\eta_{k-1}^{l} - \eta_{k-1}^{l-1}\right](g_{k-1})\right\}\right]\right|,$$

$$T_{13} = \left|\mathbb{E}\left[\frac{\eta_{k-1}^{l-1}\big(g_{k-1}P(\varphi)\big)}{\eta_{k-1}^{N_l,l}(g_{k-1})\eta_{k-1}^{l}(g_{k-1})\eta_{k-1}^{N_l,l-1}(g_{k-1})}\left\{\left[\eta_{k-1}^{N_l,l} - \eta_{k-1}^{l}\right](g_{k-1})\right\}\left\{\left[\eta_{k-1}^{l} - \eta_{k-1}^{l-1}\right](g_{k-1})\right\}\right]\right|.$$

Since T_8 and T_9 can be bounded using similar approaches, we consider only T_8. Similarly, $T_{10}, \ldots, T_{13}$ can be bounded in almost the same way, so we consider only T_{10}. For T_8 one has $T_8 \leq T_{14} + T_{15}$, where

$$\begin{aligned} T_{14} = \Bigg|\mathbb{E}\Bigg[\left\{\frac{1}{\eta_{k-1}^{N_l,l}(g_{k-1})} - \frac{1}{\eta_{k-1}^{l}(g_{k-1})}\right\}\Bigg\{&\left[\eta_{k-1}^{N_l,l} - \eta_{k-1}^{N_l,l-1}\right]\big(g_{k-1}P(\varphi)\big) \\ &- \left[\eta_{k-1}^{l} - \eta_{k-1}^{l-1}\right]\big(g_{k-1}P(\varphi)\big)\Bigg\}\Bigg]\Bigg|, \end{aligned}$$

$$T_{15} = \left|\mathbb{E}\left[\frac{1}{\eta_{k-1}^{l}(g_{k-1})}\left\{\left[\eta_{k-1}^{N_l,l} - \eta_{k-1}^{N_l,l-1}\right]\big(g_{k-1}P(\varphi)\big) - \left[\eta_{k-1}^{l} - \eta_{k-1}^{l-1}\right]\big(g_{k-1}P(\varphi)\big)\right\}\right]\right|.$$

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

For T_{14} one can use Cauchy–Schwarz, the lower bound on g_{k-1}, [14, Proposition C.6], and Lemma C.8. For T_{15} we can apply the lower bound on g_{k-1} and the induction hypothesis. These two arguments give that

$$\max\{T_{14}, T_{15}\} \leq \frac{C\overline{\|\varphi\|}\Delta_l^{\frac{1}{4}}}{N_l}$$

and hence that

$$T_8 \leq \frac{C\overline{\|\varphi\|}\Delta_l^{\frac{1}{4}}}{N_l}.$$

For T_{10} we have the upper bound $T_{10} \leq T_{16} + T_{17}$, where

$$T_{16} = \Bigg|\mathbb{E}\Bigg[\Bigg\{\frac{1}{\eta_{k-1}^{N_l,l}(g_{k-1})\eta_{k-1}^{l}(g_{k-1})} - \frac{1}{\eta_{k-1}^{l}(g_{k-1})^2}\Bigg\}\Big\{\Big[\eta_{k-1}^{l} - \eta_{k-1}^{N_l,l}\Big](g_{k-1})\Big\}$$
$$\times\Big\{\Big[\eta_{k-1}^{l} - \eta_{k-1}^{l-1}\Big]\big(g_{k-1}P(\varphi)\big)\Big\}\Bigg]\Bigg|,$$
$$T_{17} = \Bigg|\mathbb{E}\Bigg[\frac{1}{\eta_{k-1}^{l}(g_{k-1})^2}\Big\{\Big[\eta_{k-1}^{l} - \eta_{k-1}^{N_l,l}\Big](g_{k-1})\Big\}\Big\{\Big[\eta_{k-1}^{l} - \eta_{k-1}^{l-1}\Big]\big(g_{k-1}P(\varphi)\big)\Big\}\Bigg]\Bigg|.$$

For T_{15} one can use Cauchy–Schwarz, the lower bound on g_{k-1}, [14, Proposition C.6] (twice), and [14, Lemma D.2]. For T_{16} we can apply the lower bound on g_{k-1}, [14, Lemma D.2], and bias results for particle filters. These two arguments give

$$\max\{T_{15}, T_{16}\} \leq \frac{C\overline{\|\varphi\|}\Delta_l}{N_l}$$

and thus

$$T_9 \leq \frac{C\overline{\|\varphi\|}\Delta_l}{N_l}.$$

Thus we have proved that

$$T_2 \leq \frac{C\overline{\|\varphi\|}\Delta_l^{\frac{1}{4}}}{N_l}. \tag{C.13}$$

The proof can be completed by combining (C.12) and (C.13). □

Remark C.4. Using the approach in the proof of Lemma C.10, one can also prove the following result. Assume (A1)–(A2). Then for any $k \in \mathbb{N}$ there exists a $C < +\infty$ such that for any $(l, N_l, \varphi) \in \mathbb{N}^2 \times \mathcal{B}_b(\mathsf{X}) \cap Lip(\mathsf{X})$,

$$\Big|\mathbb{E}\Big[\Big[\eta_k^{N_l,l} - \eta_k^{N_l,a}\Big](\varphi)\Big]\Big| \leq \frac{C\overline{\|\varphi\|}\Delta_l^{\frac{1}{4}}}{N_l}.$$

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

Lemma C.11 . *Assume* (A1)–(A2). *Then for any* $(k, \varphi) \in \mathbb{N} \times \mathcal{B}_b(\mathsf{X}) \cap \mathcal{C}_b^2(\mathsf{X}, \mathbb{R})$ *there exists a* $C < +\infty$ *such that for any* $(l, N_l, \varepsilon) \in \mathbb{N}^2 \times (0, 1/2)$,

$$\left|\mathbb{E}\left[\left[\tfrac{1}{2}\eta_k^{N_l,l} + \tfrac{1}{2}\eta_k^{N,l,a} - \eta_k^{N_l,l-1}\right](\varphi) - \left[\eta_k^l - \eta_k^{l-1}\right](\varphi)\right]\right| \leq \frac{C\Delta_l^{\frac{1}{4}}}{N_l}.$$

Proof. The proof is by induction. The case $k = 1$ is trivial as there is no bias, so we proceed to the induction step. Following the proof of Lemma C.9, we have the decomposition

$$\left|\mathbb{E}\left[\left[\tfrac{1}{2}\eta_k^{N_l,l} + \tfrac{1}{2}\eta_k^{N,l,a} - \eta_k^{N_l,l-1}\right](\varphi) - \left[\eta_k^l - \eta_k^{l-1}\right](\varphi)\right]\right| \leq T_1 + T_2,$$

where

$$T_1 = \left|\mathbb{E}\left[\left\{\frac{\frac{1}{2}\eta_{k-1}^{N_l,l}\left(g_{k-1}[P^l - P](\varphi)\right)}{\eta_{k-1}^{N_l,l}(g_{k-1})} + \frac{\frac{1}{2}\eta_{k-1}^{N_l,l,a}\left(g_{k-1}[P^l - P](\varphi)\right)}{\eta_{k-1}^{N_l,l,a}(g_{k-1})} - \frac{\eta_{k-1}^{N_l,l-1}\left(g_{k-1}[P^{l-1} - P](\varphi)\right)}{\eta_{k-1}^{N_l,l-1}(g_{k-1})}\right\} - \left\{\frac{\frac{1}{2}\eta_{k-1}^{l}\left(g_{k-1}[P^l - P](\varphi)\right)}{\eta_{k-1}^{l}(g_{k-1})} + \frac{\frac{1}{2}\eta_{k-1}^{l}\left(g_{k-1}[P^l - P](\varphi)\right)}{\eta_{k-1}^{l}(g_{k-1})} - \frac{\eta_{k-1}^{l-1}\left(g_{k-1}[P^{l-1} - P](\varphi)\right)}{\eta_{k-1}^{l-1}(g_{k-1})}\right\}\right]\right|,$$

$$T_2 = \left|\mathbb{E}\left[\left\{\frac{\frac{1}{2}\eta_{k-1}^{N_l,l}\left(g_{k-1}P(\varphi)\right)}{\eta_{k-1}^{N_l,l}(g_{k-1})} + \frac{\frac{1}{2}\eta_{k-1}^{N_l,l,a}\left(g_{k-1}P(\varphi)\right)}{\eta_{k-1}^{N_l,l,a}(g_{k-1})} - \frac{\eta_{k-1}^{N_l,l-1}\left(g_{k-1}P(\varphi)\right)}{\eta_{k-1}^{N_l,l-1}(g_{k-1})}\right\} - \left\{\frac{\frac{1}{2}\eta_{k-1}^{l}\left(g_{k-1}P(\varphi)\right)}{\eta_{k-1}^{l}(g_{k-1})} + \frac{\frac{1}{2}\eta_{k-1}^{l}\left(g_{k-1}P(\varphi)\right)}{\eta_{k-1}^{l}(g_{k-1})} - \frac{\eta_{k-1}^{l-1}\left(g_{k-1}P(\varphi)\right)}{\eta_{k-1}^{l-1}(g_{k-1})}\right\}\right]\right|. \tag{C.14}$$

For T_1 one can match the empirical and limit terms across the l, a, $l-1$ and adopt the same proof approach as used for T_1 in the proof of Lemma C.10; since the proof would be repeated, we omit it and remark only that

$$T_1 \leq \frac{C\Delta_l^{1/2}}{N_l}. \tag{C.15}$$

For T_2, using Lemma C.4 and the triangle inequality, we have that $T_2 \leq \sum_{j=1}^4 T_{j+2}$, where

$$T_3 = \left|\mathbb{E}\left[\frac{1}{\eta_{k-1}^{N_l,l-1}(g_{k-1})}\left[\tfrac{1}{2}\eta_{k-1}^{N_l,l} + \tfrac{1}{2}\eta_{k-1}^{N_l,l,a} - \eta_{k-1}^{N_l,l-1}\right]\left(g_{k-1}P(\varphi)\right) - \frac{1}{\eta_{k-1}^{l-1}(g_{k-1})}\left[\tfrac{1}{2}\eta_{k-1}^{l} + \tfrac{1}{2}\eta_{k-1}^{l} - \eta_{k-1}^{l-1}\right]\left(g_{k-1}P(\varphi)\right)\right]\right|,$$

$$T_4 = \left|\mathbb{E}\left[\frac{1}{\eta_{k-1}^{N_l,l}(g_{k-1})\eta_{k-1}^{N_l,l-1}(g_{k-1})}\frac{1}{2}\left\{\left[\eta_{k-1}^{N_l,l} - \eta_{k-1}^{N_l,l,a}\right]\left(g_{k-1}P(\varphi)\right)\right\}\left\{\left[\eta_{k-1}^{N_l,l-1} - \eta_{k-1}^{N_l,l}\right](g_{k-1})\right\}\right]\right|,$$

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

$$T_5 = \left|\mathbb{E}\left[\frac{\frac{1}{2}\eta_{k-1}^{N_l,l,a}\big(g_{k-1}P(\varphi)\big)}{\eta_{k-1}^{N_l,l,a}(g_{k-1})\eta_{k-1}^{N_l,l}(g_{k-1})\eta_{k-1}^{N_l,l-1}](g_{k-1})}\left\{\left[\eta_{k-1}^{N_l,l,a} - \eta_{k-1}^{N_l,l}\right](g_{k-1})\right\}\left\{\left[\eta_{k-1}^{N_l,l-1} - \eta_{k-1}^{N_l,l}\right](g_{k-1})\right\}\right]\right|,$$

$$T_6 = \left|\mathbb{E}\left[\frac{\eta_{k-1}^{N_l,l,a}\big(g_{k-1}P(\varphi)\big)}{\eta_{k-1}^{N_l,l,a}(g_{k-1})\eta_{k-1}^{N_l,l-1}(g_{k-1})}\left[\tfrac{1}{2}\eta_{k-1}^{N_l,l} + \tfrac{1}{2}\eta_{k-1}^{N_l,l,a} - \eta_{k-1}^{N_l,l-1}\right](g_{k-1}) - \frac{\eta_{k-1}^{l}\big(g_{k-1}P(\varphi)\big)}{\eta_{k-1}^{l}(g_{k-1})\eta_{k-1}^{l-1}(g_{k-1})} \times \left[\tfrac{1}{2}\eta_{k-1}^{l} + \tfrac{1}{2}\eta_{k-1}^{l} - \eta_{k-1}^{l-1}\right](g_{k-1})\right]\right|.$$

As T_3 (resp. T_4) and T_6 (resp. T_5) can be dealt with using similar arguments, we only prove bounds for T_3 (resp. T_4). For T_3 we have the upper bound $T_3 \leq \sum_{j=1}^{4} T_{j+6}$, where

$$T_7 = \left|\mathbb{E}\left[\left\{\frac{1}{\eta_{k-1}^{N_l,l-1}(g_{k-1})} - \frac{1}{\eta_{k-1}^{l-1}(g_{k-1})}\right\}\left\{\left[\tfrac{1}{2}\eta_{k-1}^{N_l,l} + \tfrac{1}{2}\eta_{k-1}^{N_l,l,a} - \eta_{k-1}^{N_l,l-1}\right]\big(g_{k-1}P(\varphi)\big) - \left[\eta_{k-1}^{l} - \eta_{k-1}^{l-1}\right]\big(g_{k-1}P(\varphi)\big)\right\}\right]\right|,$$

$$T_8 = \left|\mathbb{E}\left[\frac{1}{\eta_{k-1}^{l-1}(g_{k-1})}\left\{\left[\tfrac{1}{2}\eta_{k-1}^{N_l,l} + \tfrac{1}{2}\eta_{k-1}^{N_l,l,a} - \eta_{k-1}^{N_l,l-1}\right]\big(g_{k-1}P(\varphi)\big) - \left[\eta_{k-1}^{l} - \eta_{k-1}^{l-1}\right]\big(g_{k-1}P(\varphi)\big)\right\}\right]\right|,$$

$$T_9 = \left|\mathbb{E}\left[\left[\eta_{k-1}^{l} - \eta_{k-1}^{l-1}\right]\big(g_{k-1}P(\varphi)\big)\left\{\frac{1}{\eta_{k-1}^{N_l,l-1}(g_{k-1})\eta_{k-1}^{l-1}(g_{k-1})} - \frac{1}{\eta_{k-1}^{l-1}(g_{k-1})^2}\right\} \times \left[\eta_{k-1}^{l-1} - \eta_{k-1}^{N_l,l-1}\right](g_{k-1})\right]\right|,$$

$$T_{10} = \left|\mathbb{E}\left[\frac{\left[\eta_{k-1}^{l} - \eta_{k-1}^{l-1}\right]\big(g_{k-1}P(\varphi)\big)}{\eta_{k-1}^{l-1}(g_{k-1})^2}\left[\eta_{k-1}^{l-1} - \eta_{k-1}^{N_l,l-1}\right](g_{k-1})\right]\right|.$$

For T_7 one can use Cauchy–Schwarz, the lower bound on g_{k-1}, [14, Proposition C.6], and Lemma C.9. For T_8 we can apply the lower bound on g_{k-1} and the induction hypothesis. These arguments yield

$$\max\{T_7, T_8\} \leq \frac{C\Delta_l^{\frac{1}{4}}}{N_l}.$$

For T_9 we can use [14, Lemma D.2], Cauchy–Schwarz, the lower bound on g_{k-1}, and [14, Proposition C.6] twice. For T_{10} we can use [14, Lemma D.2], the lower bound on g_{k-1}, and standard bias results for particle filters. This gives

$$\max\{T_9, T_{10}\} \leq \frac{C\Delta_l^{1/2}}{N_l}.$$

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

Collecting the above arguments, we have shown that

$$T_3 \le \frac{C\Delta_l^{\frac{1}{4}}}{N_l}.$$

For T_4 we have the upper bound $T_4 \le \sum_{j=1}^{3} T_{j+10}$, where

$$T_{11} = \left|\mathbb{E}\left[\frac{1}{\eta_{k-1}^{N_l,l}(g_{k-1})\eta_{k-1}^{N_l,l-1}(g_{k-1})}\frac{1}{2}\left\{\left[\eta_{k-1}^{N_l,l} - \eta_{k-1}^{N_l,l,a}\right]\left(g_{k-1}P(\varphi)\right)\right\}\left\{\left[\eta_{k-1}^{N_l,l-1} - \eta_{k-1}^{N_l,l}\right](g_{k-1}) - \left[\eta_{k-1}^{l-1} - \eta_{k-1}^{l}\right](g_{k-1})\right\}\right]\right|,$$

$$T_{12} = \left|\mathbb{E}\left[\frac{1}{2}\left\{\left[\eta_{k-1}^{N_l,l} - \eta_{k-1}^{N_l,l,a}\right]\left(g_{k-1}P(\varphi)\right)\right\}\left[\eta_{k-1}^{l-1} - \eta_{k-1}^{l}\right](g_{k-1})\left\{\frac{1}{\eta_{k-1}^{N_l,l}(g_{k-1})\eta_{k-1}^{N_l,l-1}(g_{k-1})} - \frac{1}{\eta_{k-1}^{l}(g_{k-1})\eta_{k-1}^{l-1}(g_{k-1})}\right\}\right]\right|,$$

$$T_{13} = \left|\mathbb{E}\left[\frac{[\eta_{k-1}^{l-1} - \eta_{k-1}^{l}](g_{k-1})}{\eta_{k-1}^{l}(g_{k-1})\eta_{k-1}^{l-1}(g_{k-1})}\left\{\left[\eta_{k-1}^{N_l,l} - \eta_{k-1}^{N_l,l,a}\right]\left(g_{k-1}P(\varphi)\right)\right\}\right]\right|.$$

For T_{11} one can use the lower bound on g_{k-1}, Cauchy–Schwarz, Remark C.3, and Lemma C.9. For T_{12} we can apply [14, Lemma D.2], Cauchy–Schwarz, the lower bound on g_{k-1}, [14, Proposition C.6], and Remark C.3. For T_{13} we can use the lower bound on g_{k-1}, Remark C.4, and [14, Lemma D.2]. Putting these results together, we have established that

$$\max\{T_{11}, T_{12}, T_{13}\} \le \frac{C\Delta_l^{\frac{1}{4}}}{N_l},$$

and thus we can deduce the same upper bound (up to a constant) for T_4. As a result of the above arguments, we have established that

$$T_2 \le \frac{C\Delta_l^{\frac{1}{4}}}{N_l}. \tag{C.16}$$

The proof can be completed by combining (C.15) and (C.16). □

Theorem 3. *Assume* (A1)–(A2). *Then for any* $(k, \varphi) \in \mathbb{N} \times \mathcal{B}_b(\mathsf{X}) \cap \mathcal{C}_b^2(\mathsf{X}, \mathbb{R})$ *there exists a* $C < +\infty$ *such that for any* $(l, N_l) \in \mathbb{N}^2$,

$$\left|\mathbb{E}\left[\left[\pi_k^l - \pi_k^{l-1}\right]^{N_l}(\varphi) - \left[\pi_k^l - \pi_k^{l-1}\right](\varphi)\right]\right| \le \frac{C\Delta_l^{\frac{1}{4}}}{N_l}.$$

Proof. This can be proved using Lemma C.11, Remark C.4, and the approach used to deal with (C.14) in the proof of Lemma C.11. The proof is omitted for brevity. □

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press

Acknowledgements

All of the authors were supported by KAUST baseline funding. We thank two referees and an editor for their valuable comments, which greatly improved the article.

Funding information

There are no funding bodies to thank in relation to the creation of this article.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] BESKOS, A. AND ROBERTS, G. (2005). Exact simulation of diffusions. *Ann. Appl. Prob.* **15**, 2422–2444.
[2] BESKOS, A. *et al.* (2017). Multilevel sequential Monte Carlo samplers. *Stoch. Process. Appl.* **127**, 1417–1440.
[3] BESKOS, A. *et al.* (2021). Score-based parameter estimation for a class of continuous-time state space models. *SIAM J. Sci. Comput.* **43**, A2555–A2580.
[4] BLANCHET, J. AND ZHANG, F. (2020). Exact simulation for multivariate Itô diffusions. *Adv. Appl. Prob.* **52**, 1003–1034.
[5] CAPPÉ, O., MOULINES, E. AND RYDÉN, T. (2005). *Inference in Hidden Markov Models*. Springer, New York.
[6] DEL MORAL, P. (2013). *Mean Field Simulation for Monte Carlo Integration*. Chapman and Hall, London.
[7] DEL MORAL, P., JACOD, J. AND PROTTER, P. (2001). The Monte-Carlo method for filtering with discrete-time observations. *Prob. Theory Relat. Fields* **120**, 346–368.
[8] FEARNHEAD, P., PAPASPILIOPOULOS, O. AND ROBERTS, G. O. (2008). Particle filters for partially observed diffusions. *J. R. Statist. Soc. B [Statist. Methodology]* **70**, 755–777.
[9] GILES, M. B. (2008). Multilevel Monte Carlo path simulation. *Operat. Res.* **56**, 607–617.
[10] GILES, M. B. (2015) Multilevel Monte Carlo methods. *Acta Numerica* **24**, 259–328.
[11] GILES, M. B. AND SZPRUCH, L. (2014). Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation. *Ann. Appl. Prob.* **24**, 1585–1620.
[12] GOLIGHTLY, A. AND WILKINSON, D. (2008). Bayesian inference for nonlinear multivariate diffusion models observed with error. *Comput. Statist. Data Anal.* **52**, 1674–1693.
[13] HEINRICH, S. (2001). Multilevel Monte Carlo methods. In *LSSC '01: Proceedings of the Third International Conference on Large-Scale Scientific Computing—Revised Papers*, eds S. Margenov, J. Wasniewski and P. Yalamov, Springer, Berlin, Heidelberg, pp. 58–67.
[14] JASRA, A., KAMATANI, K., LAW, K. J. H. AND ZHOU, Y. (2017). Multilevel particle filters. *SIAM J. Numer. Anal.* **55**, 3068–3096.
[15] JASRA, A., KAMATANI, K., OSEI, P. P. AND ZHOU, Y. (2018). Multilevel particle filters: normalizing constant estimation. *Statist. Comput.* **28**, 47–60.
[16] JASRA, A., LAW, K. J. H. AND OSEI, P. P. (2019). Multilevel particle filters for Lévy driven stochastic differential equations. *Statist. Comput.* **29**, 775–789.
[17] JASRA, A., LAW, K. J. H. AND SUCIU, C. (2020). Advanced multilevel Monte Carlo. *Internat. Statist. Rev.* **88**, 548–579.
[18] JASRA, A., LAW, K. J. H. AND YU, F. (2022). Unbiased filtering of a class of partially observed diffusions. *Adv. Appl. Prob.* **54**, 661–687.
[19] JASRA, A. AND YU, F. (2020). Central limit theorems for coupled particle filters. *Adv. Appl. Prob.* **52**, 942–1001.
[20] ROGERS, L. C. G. AND WILLIAMS, D. (2000). *Diffusions, Markov Processes, and Martingales*, Vol. **2**. Cambridge University Press.
[21] RUZAYQAT, H. *et al.* (2023). Unbiased estimation using a class of diffusion processes. *J. Comput. Phys.* **472**, article no. 111643.
[22] RYDER, T., GOLIGHTLY, A., MCGOUGH, S. AND PRANGLE, D. (2018). Black-box variational inference for stochastic differential equations. *Internat. Conf. Mach. Learning* **80**, 4423–4432.
[23] STROOCK, D. W. (2008). *Partial Differential Equations for Probabilists*. Cambridge University Press.

https://doi.org/10.1017/apr.2024.12 Published online by Cambridge University Press