Antithetic Multilevel Methods for Elliptic and Hypo-Elliptic Diffusions with Applications*

Yuga Iguchi†, Ajay Jasra‡, Mohamed Maama§, and Alexandros Beskos†

Abstract. We present a new antithetic multilevel Monte Carlo (MLMC) method for the estimation of expectations with respect to laws of diffusion processes that can be elliptic or hypo-elliptic. In particular, we consider the case where one has to resort to time discretization of the diffusion and numerical simulation of such schemes. Inspired by recent works, we introduce a new MLMC estimator of expectations, which does not require any Lévy area simulation and has a strong error of order 2 and a weak error of order 2. We then show how this approach can be used in the context of the filtering problem associated to partially observed diffusions with discrete time observations. We illustrate that in numerical simulations our new approaches provide efficiency gains for several problems, particularly when the diffusion process is hypo-elliptic, relative to some existing methods.

Key words. Stochastic Differential Equations, Multilevel Monte Carlo, Filtering.

MSC codes. 60H35, 65C05, 65C30

1. Introduction.

We consider N-dimensional stochastic differential equation (SDE):

$$dX_t = \sigma_0(X_t)dt + \sum_{1\le j\le d} \sigma_j(X_t)dB_t^j, \quad X_0 = x \in \mathbb{R}^N, \tag{1.1}$$

where $\{B_t\}_{t\ge 0}$ is the d-dimensional standard Brownian motion defined upon the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\ge 0}, \mathbb{P})$, and $\sigma_j : \mathbb{R}^N \to \mathbb{R}^N$ satisfies some regularity conditions, to be made precise later, with $\sigma_j = [\sigma_j^1, \dots, \sigma_j^N]^\top$ for $0 \le j \le d$. Throughout the paper, the matrix $a = \sigma\sigma^\top$ can be degenerate, with $\sigma \equiv [\sigma_1, \dots, \sigma_d]$. Thus, this class of diffusion process includes certain elliptic and hypo-elliptic diffusion processes that can be found in applications; see for instance [25]. In particular, a lot of interest is shown recently in the literature for numerical analysis and statistical inference methods for hypo-elliptic diffusions (see e.g. [7, 13, 17, 16]). We consider the context that one cannot obtain an exact solution of the SDE, despite its existence, and has to resort to time-discretization of the diffusion and the associated numerical simulation and, again, there are many examples of such processes that are used in practice [25].

The collection of problems that we focus upon in this article is, firstly, the computation of expectations with respect to (w.r.t.) laws of diffusion processes; we call this the *forward problem*. That is, given a function $\varphi : \mathbb{R}^N \to \mathbb{R}$ that is integrable w.r.t. the transition law of the

*Funding: YI is supported by EPSRC (EP/V521917/1) and the Heilbronn Institute for Mathematical Research. AJ was supported by SDS CUHK, Shenzhen.

†Department of Statistical Science, University College London, London, UK. (yuga.iguchi.21@ucl.ac.uk, a.beskos@ucl.ac.uk).

‡School of Data Science, The Chinese University of Hong Kong, Shenzhen, CN. (ajayjasra@cuhk.edu.cn).

§Applied Mathematics and Computational Science Program, Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal, KSA. (maama.mohamed@gmail.com).

diffusion, the objective is the computation of a numerical approximation of $\mathbb{E}[\varphi(X_T)]$ for some given terminal time $T > 0$. Secondly, we consider the *filtering problem* for partially observed diffusion processes that are discretely observed in time. In other words (1.1) is a latent process that is observed through noisy data, only at discrete times (which we take as unit times for simplicity). The objective is then to compute an approximation of the conditional expectation of X_t at each observation time and given all the data available up-to that time. This is a classical problem in engineering, statistics and applied mathematics, see e.g. [2, 4] for further references and applications.

For both aforementioned problems, one must resort to a time discretization of (1.1) whose properties can be critical for any resulting numerical approximation method relying on it. There are several numerical methods in the literature, such as the Euler-Maruyama (E-M) method and the Milstein scheme; see for instance [25]. The main properties that are often of interest to inform the efficiency of the approximation are the weak and strong error, which we shall define, loosely, as follows – a full definition can be found later on. For a time discretization on a regular grid of spacing $\Delta > 0$, and a corresponding numerical approximation $\{\bar{X}_t\}_{t\geq 0}$ the weak error (assuming it exists) is the discrepancy:

$$|\mathbb{E}[\varphi(X_T)] - \mathbb{E}[\varphi(\bar{X}_T)]|$$

for an appropriate test function $\varphi : \mathbb{R}^N \to \mathbb{R}$. We remark that the numerical approximation may be defined in continuous time by interpolation between points on the time grid. The strong error[1] (assuming it exists) is taken as:

$$\mathbb{E}\big[\|X_T - \bar{X}_T\|^2\big]$$

where $\|\cdot\|$ is the L_2-norm. There are several results for well-known discretization methods; e.g., E-M has weak error of $\mathcal{O}(\Delta)$ (weak error 1) and strong error of $\mathcal{O}(\Delta)$ (strong error 1) and the Milstein scheme has weak error 1 and strong error 2. In the context of the methods to be used in this article, one generally would like the order of weak and strong error to be 'large' at a cost of $\mathcal{O}(\Delta^{-1})$ for *directly* simulating the approximation. We note that direct simulation, without for instance solving linear equations of cost of order $\mathcal{O}(N^m)$, $m > 2$, is critical for practical problems, especially filtering.

In this work we consider both elliptic and hypo-elliptic diffusion processes and in the latter case we have $N \geq 2$. In such scenarios, the Milstein method (or the strong 1.5 scheme, see [25], with weak error 2 and strong error 3) cannot often be simulated directly, without a restrictive commutative condition (given later on), as one has to compute an intractable *Lévy area*. In such cases one resorts to the E-M approach, which can be simulated exactly, but the order of weak and strong error is comparatively low. Whilst there are some higher order discretization methods based upon stochastic Runge-Kutta approaches (see e.g. [29]), generally for many Monte Carlo simulation-based methods a strong error of 2 generally suffices for 'optimal' (to be clarified later on) variance properties. An elegant methodology that side-steps sampling

[1] In the literature, $\{\mathbb{E}\big[\|X_T - \bar{X}_T\|^2\big]\}^{1/2}$ is used as the standard definition of strong error. However, we make use of the squared version of the definition because it aligns with the analysis on the variance of couplings in the context of multilevel Monte Carlo.

of Lévy areas but preserves strong error 2 was developed in [11] based upon the multilevel Monte Carlo (MLMC) approach [9, 10, 14].

MLMC works with a hierarchy of time-discretised diffusions, that is with a collection of step-sizes $0 < \Delta_0 < \cdots < \Delta_L$, $L \in \mathbb{N}$. Then one rewrites the expectation of interest as a decomposition of the difference of the exact (no time discretization) expectation and the one with the finest time discretization and then a telescoping sum of differences of expectations associated to increasingly coarse step-sizes. Then, if one can appropriately simulate dependent (*coupled*) time discretizations for pairs of step-sizes it is possible to reduce the cost of a Monte Carlo based algorithm (e.g. the cost versus a direct simulation of the time discretised diffusion with a single step-size Δ_L) to achieve a pre-specified mean square error (MSE) using MLMC; see e.g. [10] for a review. [11] introduce an *antithetic* MLMC (AMLMC) using the *truncated Milstein scheme* (defined in Section 2.2.2) which has weak error 1 and stong error 1 without requiring the simulation of intractable Lévy areas, but the variance of couplings at each level decays w.r.t. the step-size at the same rate as the case of a time discretization having strong error 2, which leads to an optimal computational complexity.

In this article we develop a new method (multilevel-based) for time discretization which is effective in both the elliptic and hypo-elliptic contexts. Motivated by the work in [11], we derive a new AMLMC based on the numerical scheme proposed in [16] achieving weak error 2 and strong error 1 (the latter is proven in this article), which still gives an optimal computational complexity (for the forward problem). The method can also be simulated directly with a cost of $\mathcal{O}(\Delta^{-1})$ per-pair of levels $0 < \Delta < \Delta'$. An AMLMC with a weak error 2 has also been investigated in [1], where they used an alternative numerical scheme with a weak error 2 and emphasized its efficiency due to the reduction of the number of time-discretizations, which is an advantage over the AMLMC that uses the truncated Milstein scheme (weak error 1). A comparison between our proposed AMLMC and the method by [1] is given later in Section 2.4. In addition, we show that our new methodology can be used for the filtering problem. Some of the state-of-the-art numerical methods for this problem are based upon particle filters (e.g. [4, 6]) related to the MLMC approach, which are termed *multilevel particle filters* (MLPFs) see e.g. [18, 24]. Based upon the methodology developed herein, we derive a new MLPF. To summarize, the main contributions of this article are:

- We introduce a locally non-degenerate scheme of weak error 2 for both elliptic and hypo-elliptic contexts, inspired by [16]. We prove that the scheme has strong error 1.
- We then develop a new AMLMC method that does not contain Lévy areas and prove that the variance of the AMLMC estimator decays (w.r.t the step-size) at the same rate as for a discretization scheme that would achieve a strong error 2.
- We show how to use the new AMLMC method for filtering within the context of MLPFs.
- We present numerical results to show that our method can out-perform some competing approaches.

We further elaborate on some of the bullet points above. In the case of the forward problem, the second bullet point leads to the new AMLMC estimator having a cost of $\mathcal{O}(\epsilon^{-2})$ to give a MSE of $\mathcal{O}(\epsilon^2)$, $\epsilon > 0$, i.e. the method attains the optimal cost for (stochastic) Monte Carlo based methods. Such a MSE is also achieved by [11], however the higher rate of weak error is expected to provide efficiency gains – verified in our numerical experiments – due to the necessity of the use of a finite L (the most precise level) in simulations. In the case of filtering, we compare to the MLPF approaches in [18, 24]. In [18] the authors prove that, in the elliptic case, to achieve a MSE of $\mathcal{O}(\epsilon^2)$ there is a cost of $\mathcal{O}(\epsilon^{-2.5})$. In [24] the authors show that in simulations to achieve a MSE of $\mathcal{O}(\epsilon^2)$ there is a cost of $\mathcal{O}(\epsilon^{-2}\log(\epsilon)^2)$; this latter MLPF corresponds to an embedding of the multilevel approach of [11] within the filtering problem. We verify in our simulations that, as one expects based upon [18], our new MLPF has costs consistent with the anticipated rate $\mathcal{O}(\epsilon^{-2}\log(\epsilon)^2)$ to achieve a MSE of $\mathcal{O}(\epsilon^2)$. However, as the discretization schemes underpinning the methods in [18, 24] have weak error 1, we again observe efficiency gains for finite L. Finally, we note that our numerical scheme is locally *non-degenerate* under a hypo-elliptic setting, while this is not the case for the truncated Milstein scheme. The existence of the density (non-degeneracy) is important in the filtering problem when utilising guided proposals [5] to improve the performance of particle filters.

This paper is structured as follows. In Section 2 we consider several numerical schemes for SDEs and introduce our approach. In Section 3 we describe how our idea can be used in the context of the filtering problem and derive the new MLPF. In Section 4 we present our numerical results to illustrate our theoretical derivations. The mathematical proofs of our main results are given in the Appendix.

Notation: Let $C_b^K(\mathbb{R}^n;\mathbb{R}^m)$, $n,m,K\in\mathbb{N}$, be the space of K-times differentiable functions $f:\mathbb{R}^n\to\mathbb{R}^m$ such that partial derivatives up to order K are bounded. For a vector $y\in\mathbb{R}^N$, we define the norm $\|\cdot\|$ as $\|y\|\equiv\sqrt{\sum_{1\le i\le N} y_i^2}$.

2. Numerical schemes.

2.1. Basic assumptions and error. To study a broad class of SDEs including the case where the matrix $a=\sigma\sigma^\top$ is degenerate, we consider the following structure for model (1.1):

$$dX_t=\begin{bmatrix}dX_{S,t}\\ dX_{R,t}\end{bmatrix}=\begin{bmatrix}\sigma_{S,0}(X_t)\\ \sigma_{R,0}(X_t)\end{bmatrix}dt+\sum_{1\le j\le d}\begin{bmatrix}\mathbf{0}_{N_S}\\ \sigma_{R,j}(X_t)\end{bmatrix}dB_t^j,\qquad X_0=x\in\mathbb{R}^N, \tag{2.1}$$

where we have set $\sigma_{S,0}:\mathbb{R}^N\to\mathbb{R}^{N_S}$, $\sigma_{R,j}:\mathbb{R}^N\to\mathbb{R}^{N_R}$, $0\le j\le d$, with integers $N_S\ge 0$, $N_R\ge 1$ such that $N_S+N_R=N$. We write for $x\in\mathbb{R}^N$:

$$\sigma_0(x)=\left[\sigma_{S,0}(x)^\top,\sigma_{R,0}(x)^\top\right]^\top,\qquad \sigma_j(x)=\left[\mathbf{0}_{N_S}^\top,\sigma_{R,j}(x)^\top\right]^\top,\quad 1\le j\le d,$$

and $a\equiv\sigma\sigma^\top$ with $\sigma\equiv[\sigma_1,\ldots,\sigma_d]$. Notice that when $N_S\ge 1$, the matrix a is degenerate. We write $[\sigma_0,\sigma_j](x)\equiv\sum_{1\le k\le N}\{\widetilde{\sigma}_0^k(x)\partial_{x_k}\sigma_j(x)-\sigma_j^k(x)\partial_{x_k}\widetilde{\sigma}_0(x)\}$, $1\le j\le d$, where $\widetilde{\sigma}_0:\mathbb{R}^N\to\mathbb{R}^N$ is the drift function when the Itô-type SDE (2.1) is written as a Stratonovich one, specifically, $\widetilde{\sigma}_0(x)\equiv\sigma_0(x)-\frac{1}{2}\sum_{1\le i\le N}\sum_{1\le j\le d}\sigma_j^i(x)\partial_i\sigma_j(x)$.

We introduce the following assumptions related to *Hörmander's condition* (see e.g. [28]).

Assumption 2.1. $\sigma_j\in C_b^\infty(\mathbb{R}^N;\mathbb{R}^N)$, $0\le j\le d$.

Assumption 2.2. (i) Ellipticity. When $N_S=0$, it holds that for any $x\in\mathbb{R}^N$:

$$\mathrm{Span}\{\sigma_1(x),\ldots,\sigma_d(x)\}=\mathbb{R}^N.$$

(ii) Hypo-ellipticity. When $N_S\ge 1$, it holds that for any $x\in\mathbb{R}^N$:

$$\mathrm{Span}\{\sigma_{R,1}(x),\ldots,\sigma_{R,d}(x)\}=\mathbb{R}^{N_R},\ \mathrm{Span}\{\sigma_1(x),\ldots,\sigma_d(x),[\sigma_0,\sigma_1](x),\ldots,[\sigma_0,\sigma_d](x)\}=\mathbb{R}^N.$$

For a numerical scheme $\{\bar{X}_{k\Delta}\}_{k=0,1,\dots,2^\ell}$ with constant step-size $\Delta = T/2^\ell$ and non-negative integer ℓ, weak and strong error of order $m \geq 1$, respectively, are given as follows:

$$\left|\mathbb{E}\left[\varphi(X_T)\right] - \mathbb{E}\left[\varphi(\bar{X}_T)\right]\right| = \mathcal{O}(\Delta^m), \qquad \mathbb{E}\Big[\max_{0\leq k\leq 2^\ell}\left\|X_{k\Delta} - \bar{X}_{k\Delta}\right\|^{2p}\Big] = \mathcal{O}(\Delta^{mp}), \quad p \geq 1,$$

for some appropriate test function $\varphi : \mathbb{R}^N \to \mathbb{R}$.

2.2. Discretizations. We introduce a discretization scheme of weak order 2 and also mention other popular discretization schemes (e.g. the Milstein scheme) for comparison. Let $T > 0$ be a terminal time and $\Delta_\ell = T/2^\ell$ be a step-size of discretization with a non-negative integer ℓ. We make use of the notation $t_k = k\Delta_\ell$, $1 \leq k \leq 2^\ell$, and $\Delta B_{t,s} = B_t - B_s$, $0 \leq s \leq t$. For a sufficiently smooth function $\varphi : \mathbb{R}^N \to \mathbb{R}$, we set:

$$\mathcal{L}_0\varphi(x) = \sum_{1\leq k\leq N} \sigma_0^k(x)\partial_{x_k}\varphi(x) + \tfrac{1}{2}\sum_{1\leq j\leq d}\sum_{1\leq k_1,k_2\leq N}\sigma_j^{k_1}(x)\sigma_j^{k_2}(x)\partial_{x_{k_1}}\partial_{x_{k_2}}\varphi(x);$$

$$\mathcal{L}_i\varphi(x) = \sum_{1\leq k\leq N}\sigma_i^k(x)\partial_{x_k}\varphi(x), \qquad 1 \leq i \leq d.$$

We define, for $1 \leq i_1, i_2 \leq d$:

$$[\sigma_{i_1}, \sigma_{i_2}](x) = \mathcal{L}_{i_1}\sigma_{i_2}(x) - \mathcal{L}_{i_2}\sigma_{i_1}(x), \qquad x \in \mathbb{R}^N,$$

where $\mathcal{L}_{i_1}\sigma_{i_2}(x) = [\mathcal{L}_{i_1}\sigma_{i_2}^1(x), \dots, \mathcal{L}_{i_1}\sigma_{i_2}^N(x)]^\top \in \mathbb{R}^N$.

2.2.1. Milstein scheme. The Milstein scheme, of weak order 1 and strong order 2, writes as follows, for $0 \leq k \leq 2^\ell - 1$:

$$\begin{aligned}\bar{X}^{\text{Mil}}_{t_{k+1}} &= \bar{X}^{\text{Mil}}_{t_k} + \sigma_0(\bar{X}^{\text{Mil}}_{t_k})\Delta_\ell + \sum_{1\leq j\leq d}\sigma_j(\bar{X}^{\text{Mil}}_{t_k})\Delta B^j_{t_{k+1},t_k} \\ &\quad + \tfrac{1}{2}\sum_{1\leq j_1,j_2\leq d}\mathcal{L}_{j_2}\sigma_{j_1}(\bar{X}^{\text{Mil}}_{t_k})\big(\Delta B^{j_1}_{t_{k+1},t_k}\Delta B^{j_2}_{t_{k+1},t_k} - \Delta_\ell\cdot\mathbf{1}_{j_1=j_2} - \Delta A^{j_1j_2}_{t_{k+1},t_k}\big),\end{aligned} \tag{Milstein}$$

with $\bar{X}^{\text{Mil}}_0 = x$, where $\Delta A^{j_1j_2}_{t_{k+1},t_k}$ is a Lévy area specified as $\Delta A^{j_1j_2}_{t_{k+1},t_k} \equiv \int_{t_k}^{t_{k+1}}\int_{t_k}^{s} dB_u^{j_1}dB_s^{j_2} - \int_{t_k}^{t_{k+1}}\int_{t_k}^{s} dB_u^{j_2}dB_s^{j_1}$. Note that in general there is no effective way to directly simulate the Lévy area. However, if the *commutative condition* holds, i.e. for any $x \in \mathbb{R}^N$,

$$\big[\sigma_{j_1}, \sigma_{j_2}\big](x) = 0, \qquad 1 \leq j_1, j_2 \leq d, \quad j_1 \neq j_2, \tag{2.2}$$

then the Lévy area does not appear in the Milstein scheme and the latter becomes tractable.

2.2.2. Truncated Milstein scheme. The truncated Milstein scheme, used by the AMLMC method of [11], has weak and strong errors both of order equal to 1, and writes as follows, for $0 \leq k \leq 2^\ell - 1$:

$$\begin{aligned}\bar{X}^{\text{T-Mil}}_{t_{k+1}} &= \bar{X}^{\text{T-Mil}}_{t_k} + \sigma_0(\bar{X}^{\text{T-Mil}}_{t_k})\Delta_\ell + \sum_{1\leq j\leq d}\sigma_j(\bar{X}^{\text{T-Mil}}_{t_k})\Delta B^j_{t_{k+1},t_k} \\ &\quad + \tfrac{1}{2}\sum_{1\leq j_1,j_2\leq d}\mathcal{L}_{j_1}\sigma_{j_2}(\bar{X}^{\text{T-Mil}}_{t_k})\big(\Delta B^{j_1}_{t_{k+1},t_k}\Delta B^{j_2}_{t_{k+1},t_k} - \Delta_\ell\cdot\mathbf{1}_{j_1=j_2}\big),\end{aligned} \tag{T-Milstein}$$

with $\bar{X}^{\text{T-Mil}}_0 = x$. Since the scheme omits the Lévy area $\Delta A^{j_1j_2}_{t_{k+1},t_k}$, the strong convergence rate is the same as for the E-M scheme unless the commutative condition holds.

2.2.3. Second order weak scheme. Motivated from [16], we introduce two nondegenerate discretization schemes for elliptic ($N_S = 0$) and hypo-elliptic ($N_S \geq 1$) cases, separately. That is, for $0 \leq k \leq 2^\ell - 1$:

$$\begin{aligned} \bar{X}_{t_{k+1}} &= \bar{X}_{t_k} + \sigma_0(\bar{X}_{t_k})\Delta_\ell + \sum_{1 \leq j \leq d} \sigma_j(\bar{X}_{t_k})\Delta B^j_{t_{k+1},t_k} \\ &\quad + \sum_{0 \leq j_1, j_2 \leq d} \mathcal{L}_{j_1}\sigma_{j_2}(\bar{X}_{t_k})\Delta\eta^{j_1 j_2}_{t_{k+1},t_k} + \tfrac{1}{2} \sum_{1 \leq j_1 < j_2 \leq d} [\sigma_{j_1}, \sigma_{j_2}](\bar{X}_{t_k})\Delta\widetilde{A}^{j_1 j_2}_{t_{k+1},t_k}, \end{aligned} \tag{Weak-2}$$

with $\bar{X}_0 = x$, where the random variables $\Delta\eta^{j_1 j_2}_{t_{k+1},t_k}$ and $\Delta\widetilde{A}^{j_1 j_2}_{t_{k+1},t_k}$ are given as:

$$\Delta\eta^{j_1 j_2}_{t_{k+1},t_k} = \begin{cases} \Delta\eta^{\text{Ell},j_1 j_2}_{t_{k+1},t_k} & (N_S = 0); \\ \Delta\eta^{\text{H-Ell},j_1 j_2}_{t_{k+1},t_k} & (N_S \geq 1), \end{cases} \qquad \Delta\widetilde{A}^{j_1 j_2}_{t_{k+1},t_k} = \Delta B^{j_1}_{t_{k+1},t_k}\Delta\widetilde{B}^{j_2}_{t_{k+1},t_k},$$

where $\widetilde{B}_t = (\widetilde{B}^2_t, \ldots, \widetilde{B}^d_t)$, $t \geq 0$, is a standard $(d-1)$-dimensional Brownian motion independent of $\{B_t\}_{t \geq 0}$ and:

$$\Delta\eta^{\text{Ell},j_1 j_2}_{t_{k+1},t_k} = \tfrac{1}{2}\big(\Delta B^{j_1}_{t_{k+1},t_k}\Delta B^{j_2}_{t_{k+1},t_k} - \Delta_\ell \cdot \mathbf{1}_{j_1 = j_2 \neq 0}\big), \quad 0 \leq j_1, j_2 \leq d;$$

$$\Delta\eta^{\text{H-Ell},j_1 j_2}_{t_{k+1},t_k} = \begin{cases} \Delta\eta^{\text{Ell},j_1 j_2}_{t_{k+1},t_k} & (1 \leq j_1, j_2 \leq d \text{ or } j_1 = j_2 = 0); \\ \int_{t_k}^{t_{k+1}} \int_{t_k}^{s} du dB^{j_2}_s & (j_1 = 0, 1 \leq j_2 \leq d); \\ \int_{t_k}^{t_{k+1}} \int_{t_k}^{s} dB^{j_1}_u ds & (1 \leq j_1 \leq d, j_2 = 0). \end{cases}$$

In the above specification of $\Delta\eta^{j_1 j_2}_{t_{k+1},t_k}$, we use the interpretation $\Delta B^0_{t_{k+1},t_k} = \Delta_\ell$. The definition of the scheme under the hypo-elliptic setting slightly differs from the original one given in [16]. In particular, the latter includes additional random variables in the approximation of the smooth component $X_{S,t}$ for the purpose of improving the performance of parameter estimation. Without such additional variables, it is shown that (Weak-2) achieves a weak error 2 since the random variables used in the scheme satisfy the moment conditions outlined in [26, Lemma 2.1.5] that are sufficient for the attained order of weak convergence.

We give several remarks on scheme (Weak-2). First, comparing with the truncated Milstein scheme, we observe that the scheme contains the terms $\Delta\widetilde{A}$ and random variables of size $\mathcal{O}(\Delta_\ell^{3/2})$ or $\mathcal{O}(\Delta_\ell^2)$. Due to the inclusion of these terms, (Weak-2) is shown to achieve weak error 2. In particular, variable $\Delta\widetilde{A}$ is interpreted as a proxy to the Lévy area in the distributional (but not pathwise) sense. Thus, as we will show in Section 2.3, the order of strong convergence for (Weak-2) is not as good as that of the Milstein scheme which uses the true Lévy area (though the latter cannot be exactly simulated in general). Second, under the hypo-elliptic setting ($N_S \geq 1$), the scheme, in particular $\Delta\eta^{\text{H-Ell}}_{t_{k+1},t_k}$, involves $\int_{t_k}^{t_{k+1}} \int_{t_k}^{s} du dB^{j_2}_s$, $\int_{t_k}^{t_{k+1}} \int_{t_k}^{s} dB^{j_1}_u ds$ that can be directly simulated by Gaussian variables that preserve the covariance structure between these integrals and the Brownian motion. Together with Assumption 2.2, use of these variables leads to the current state $\bar{X}_{t_{k+1}}$ given $\bar{X}_{t_k}$ containing a locally Gaussian approximation with non-degenerate covariance, that is:

$$\bar{X}_{S,t_{k+1}} \approx \bar{X}_{S,t_k} + \sigma_{S,0}(\bar{X}_{t_k})\Delta_\ell + \sum_{1\le j\le d} \mathcal{L}_j\sigma_{S,0}(\bar{X}_{t_k})\Delta\eta^{\text{H-Ell},j0}_{t_{k+1},t_k};$$
$$\bar{X}_{R,t_{k+1}} \approx \bar{X}_{R,t_k} + \sigma_{R,0}(\bar{X}_{t_k})\Delta_\ell + \sum_{1\le j\le d} \sigma_{R,j}(\bar{X}_{t_k})\Delta B^j_{t_{k+1},t_k}. \tag{2.3}$$

Note that if $\Delta\eta^{\text{H-Ell},j0}_{t_{k+1},t_k}$ above is replaced with $\Delta\eta^{\text{Ell},j0}_{t_{k+1},t_k} \equiv \frac{1}{2}\Delta B^j_{t_{k+1},t_k}\Delta_\ell$ which is used in the elliptic setting ($N_S = 0$), then the covariance of the right hand side (R.H.S.) of (2.3) is no longer positive definite.

2.3. Strong convergence of the weak second order scheme and summary. The strong error rate of scheme (Weak-2) is the same as for the truncated Milstein and the E-M scheme. The proof of the following result is in Appendix A.

Proposition 2.3. *For any $p \ge 1$, there exists a constant $C > 0$ such that*

$$\mathbb{E}\Big[\max_{0\le k\le 2^\ell} \|X_{t_k} - \bar{X}_{t_k}\|^{2p}\Big] \le C\Delta_\ell^p.$$

Table 1 summarises the weak and strong errors for some of the most popular discretization schemes. The result for the strong error of scheme (Weak-2) is new.

Table 1

Numerical scheme for general SDEs (i.e. commutative condition (2.2) not assumed to hold).

Scheme	Rate of weak/strong convergence	Is Lévy area required?
Milstein	1.0 / 2.0	Yes
T-Milstein	1.0 / 1.0	No
Weak-2	2.0 / 1.0	No

2.4. Antithetic MLMC with weak second order scheme. The aim is to combine the weak order 2 method (Weak-2) with the ideas of [11] and consider a new antithetic MLMC (AMLMC) estimator so that the variance of couplings at each level decays w.r.t. the step-size at the same rate as the case of a time discretization having strong error 2. Throughout this subsection, let $\ell = 0, \dots, L$ be the level of discretization ($2^{-\ell}$), where $L \in \mathbb{N}$ is the finest level of discretization. We write $T > 0$ for the length of the time interval and $\Delta_\ell = T/2^\ell$ for the discretization step-size. To define the antithetic estimator, we design discretizations on coarse/fine grids based upon scheme (Weak-2). For a fixed $\ell \le L$, we define the coarse grids $\mathbf{g}^{c,[\ell-1]} = \{t_k\}_{k=0,1,\dots,2^{\ell-1}}$ and the fine grids $\mathbf{g}^{f,[\ell]} = \{t_k, t_{k+1/2}\}_{k=0,1,\dots,2^{\ell-1}-1} \cup \{T\}$, where $t_k = k\Delta_{\ell-1}$, $t_{k+1/2} = \big(k+1/2\big)\Delta_{\ell-1}$. For notational simplicity, we introduce two integrators $\overline{\mathcal{I}}_{t,s}, \widetilde{\mathcal{I}}_{t,s} : \mathbb{R}^N \to \mathbb{R}^N$, $0 \le s \le t$, associated with scheme (Weak-2), so that for $x \in \mathbb{R}^N$:

$$\overline{\mathcal{I}}_{t,s}(x) \equiv x + \sum_{0\le j\le d}\sigma_j(x)\Delta B^j_{t,s} + \sum_{0\le j_1,j_2\le d}\mathcal{L}_{j_1}\sigma_{j_2}(x)\Delta\eta^{j_1j_2}_{t,s} + \frac{1}{2}\sum_{1\le j_1<j_2\le d}\big[\sigma_{j_1},\sigma_{j_2}\big](x)\Delta\widetilde{A}^{j_1j_2}_{t,s};$$

$$\widetilde{\mathcal{I}}_{t,s}(x) \equiv x + \sum_{0\le j\le d}\sigma_j(x)\Delta B^j_{t,s} + \sum_{0\le j_1,j_2\le d}\mathcal{L}_{j_1}\sigma_{j_2}(x)\Delta\eta^{j_1j_2}_{t,s} - \frac{1}{2}\sum_{1\le j_1<j_2\le d}\big[\sigma_{j_1},\sigma_{j_2}\big](x)\Delta\widetilde{A}^{j_1j_2}_{t,s}.$$

Notice that the difference between the above two integrators is in the sign of the last term. On the coarse grids $\mathbf{g}^{c,[\ell-1]}$, we define a discretization scheme $\{\bar{X}_t^{c,[\ell-1]}\}_{t\in\mathbf{g}^{c,[\ell-1]}}$ and its *antithetic version* $\{\widetilde{X}_t^{c,[\ell-1]}\}_{t\in\mathbf{g}^{c,[\ell-1]}}$ as follows. For $0 \le k \le 2^{\ell-1}-1$:

$$\bar{X}_{t_0}^{c,[\ell-1]} = x, \quad \bar{X}_{t_{k+1}}^{c,[\ell-1]} = \overline{\mathcal{I}}_{t_{k+1},t_k}(\bar{X}_{t_k}^{c,[\ell-1]}); \tag{2.4}$$

$$\widetilde{X}_{t_0}^{c,[\ell-1]} = x, \quad \widetilde{X}_{t_{k+1}}^{c,[\ell-1]} = \widetilde{\mathcal{I}}_{t_{k+1},t_k}(\widetilde{X}_{t_k}^{c,[\ell-1]}). \tag{2.5}$$

Similarly, on the fine grids $\mathbf{g}^{f,[\ell]}$, we define a numerical scheme $\{\bar{X}_t^{f,[\ell]}\}_{t\in\mathbf{g}^{f,[\ell]}}$ and its antithetic version $\{\widetilde{X}_t^{f,[\ell]}\}_{t\in\mathbf{g}^{f,[\ell]}}$ as follows. For $0 \le k \le 2^{\ell-1}-1$:

$$\bar{X}_{t_0}^{f,[\ell]} = x, \quad \bar{X}_{t_{k+1/2}}^{f,[\ell]} = \overline{\mathcal{I}}_{t_{k+1/2},t_k}(\bar{X}_{t_k}^{f,[\ell]}), \quad \bar{X}_{t_{k+1}}^{f,[\ell]} = \overline{\mathcal{I}}_{t_{k+1},t_{k+1/2}}(\bar{X}_{t_{k+1/2}}^{f,[\ell]}); \tag{2.6}$$

$$\widetilde{X}_{t_0}^{f,[\ell]} = x, \quad \widetilde{X}_{t_{k+1/2}}^{f,[\ell]} = \widetilde{\mathcal{I}}_{t_{k+1},t_{k+1/2}}(\widetilde{X}_{t_k}^{f,[\ell]}), \quad \widetilde{X}_{t_{k+1}}^{f,[\ell]} = \widetilde{\mathcal{I}}_{t_{k+1/2},t_k}(\widetilde{X}_{t_{k+1/2}}^{f,[\ell]}). \tag{2.7}$$

The antithetic scheme (2.7) features the following two key properties: 1. On the interval $[t_k, t_{k+1}]$, the Gaussian increments used in the standard discretisation (2.6) are exchanged between the first and second halfs; 2. The sign of the last term in the integrator is *minus*. We note that the first point (exchange of Gaussian increments) is featured in the antithetic truncated Milstein scheme proposed in [9] as well, but the second point (change of sign) is not. Let $\varphi : \mathbb{R}^N \to \mathbb{R}$ be some suitable test function. In the next section we will define an antithetic estimator based upon the weak second order scheme (Weak-2) and use the identity:

$$\mathbb{E}[\varphi(\bar{X}_T^{f,[L]})] = \mathbb{E}[\mathcal{P}_0^{\varphi}] + \sum_{1\le\ell\le L} \mathbb{E}\big[\mathcal{P}_{f,\ell}^{\varphi} - \mathcal{P}_{c,\ell-1}^{\varphi}\big], \tag{2.8}$$

where we have set, for $1 \le \ell \le L$:

$$\mathcal{P}_{f,\ell}^{\varphi} \equiv \tfrac{1}{2}\big(\varphi(\bar{X}_T^{f,[\ell]}) + \varphi(\widetilde{X}_T^{f,[\ell]})\big),\ \mathcal{P}_{c,\ell-1}^{\varphi} \equiv \tfrac{1}{2}\big(\varphi(\bar{X}_T^{c,[\ell-1]}) + \varphi(\widetilde{X}_T^{c,[\ell-1]})\big),\ \mathcal{P}_0^{\varphi} \equiv \varphi(\bar{X}_T^{c,[0]}). \tag{2.9}$$

We notice that $\mathbb{E}[\mathcal{P}_0^{\varphi}] = \mathbb{E}[\mathcal{P}_{c,0}^{\varphi}]$, $\mathbb{E}[\mathcal{P}_{f,\ell}^{\varphi}] = \mathbb{E}[\mathcal{P}_{c,\ell}^{\varphi}]$, $1 \le \ell \le L-1$. For $s_1 \in \mathbf{g}^{c,[\ell-1]}$ and $s_2 \in \mathbf{g}^{f,[\ell]}$ with $\ell = 1, \dots, L$, we define:

$$\hat{X}_{s_1}^{c,[\ell-1]} = \tfrac{1}{2}\big(\bar{X}_{s_1}^{c,[\ell-1]} + \widetilde{X}_{s_1}^{c,[\ell-1]}\big), \qquad \hat{X}_{s_2}^{f,[\ell]} = \tfrac{1}{2}\big(\bar{X}_{s_2}^{f,[\ell]} + \widetilde{X}_{s_2}^{f,[\ell]}\big), \tag{2.10}$$

and study the L^p bound for the coupling $\mathcal{P}_{f,\ell}^{\varphi} - \mathcal{P}_{c,\ell-1}^{\varphi}$.

Lemma 2.4. *Let $\varphi \in C_b^2(\mathbb{R}^N;\mathbb{R})$ and $1 \le \ell \le L$. For any $p \ge 2$, there exist constants $C_1, C_2, C_3 > 0$ such that*

$$\begin{aligned} \mathbb{E}\big[(\mathcal{P}_{f,\ell}^{\varphi} - \mathcal{P}_{c,\ell-1}^{\varphi})^p\big] &\le C_1\mathbb{E}\big[\|\hat{X}_T^{f,[\ell]} - \hat{X}_T^{c,[\ell-1]}\|^p\big] + C_2\mathbb{E}\big[\|\bar{X}_T^{f,[\ell]} - \widetilde{X}_T^{f,[\ell]}\|^{2p}\big] \\ &\quad + C_3\mathbb{E}\big[\|\bar{X}_T^{c,[\ell-1]} - \widetilde{X}_T^{c,[\ell-1]}\|^{2p}\big]. \end{aligned} \tag{2.11}$$

Proof. The second order Taylor expansion yields: $\mathcal{P}^{\varphi}_{f,\ell} = \varphi(\hat{X}^{f,[\ell]}_T) + F^{\xi_1,\xi_2}_1$ and $\mathcal{P}^{\varphi}_{c,\ell-1} = \varphi(\hat{X}^{c,[\ell-1]}_T) + F^{\xi_3,\xi_4}_2$, where we have set:

$$F^{\xi_1,\xi_2}_1 \equiv \tfrac{1}{16}(\bar{X}^{f,[\ell]}_T - \widetilde{X}^{f,[\ell]}_T)^\top\big(\partial^2\varphi(\xi_1) + \partial^2\varphi(\xi_2)\big)(\bar{X}^{f,[\ell]}_T - \widetilde{X}^{f,[\ell]}_T);$$
$$F^{\xi_3,\xi_4}_2 \equiv \tfrac{1}{16}(\bar{X}^{c,[\ell-1]}_T - \widetilde{X}^{c,[\ell-1]}_T)^\top\big(\partial^2\varphi(\xi_3) + \partial^2\varphi(\xi_4)\big)(\bar{X}^{c,[\ell-1]}_T - \widetilde{X}^{c,[\ell-1]}_T),$$

for some $\xi_1, \xi_2, \xi_3, \xi_4 \in \mathbb{R}^N$, where $\partial^2\varphi(\cdot)$ is the $N \times N$ matrix of 2nd derivatives. Thus:

$$\mathcal{P}^{\varphi}_{f,\ell} - \mathcal{P}^{\varphi}_{c,\ell-1} = \partial\varphi(\xi_5)^\top\big(\hat{X}^{f,[\ell]}_T - \hat{X}^{c,[\ell-1]}_T\big) + F^{\xi_1,\xi_2}_1 - F^{\xi_3,\xi_4}_2, \tag{2.12}$$

for some $\xi_5 \in \mathbb{R}^N$, where $\partial\varphi(\cdot)$ is the $N \times 1$ vector of 1st derivatives. Due to the boundedness of the test function φ and the standard inequality given in (A.1), we conclude from (2.12). ∎

Our objective is to derive bounds for each term in the R.H.S. of (2.11) over a coarse time step $\Delta_{\ell-1}$. For the first term, we have the following result:

Theorem 2.5. *Let $1 \le \ell \le L$. For all $p \ge 2$, there exists a constant C such that*

$$\mathbb{E}\Big[\max_{t \in \mathbf{g}^{c,[\ell-1]}} \|\hat{X}^{f,[\ell]}_t - \hat{X}^{c,[\ell-1]}_t\|^p\Big] \le C\,\Delta^p_{\ell-1}.$$

Proof. Let $0 \le n \le 2^{\ell-1}$ and $\hat{\mathcal{S}}_n \equiv \mathbb{E}\big[\max_{0\le k\le n}\|\hat{X}^{f,[\ell]}_{t_k} - \hat{X}^{c,[\ell-1]}_{t_k}\|^p\big]$. It holds that for any $p \ge 2$, there exists a constant $C_p > 0$ such that:

$$\hat{\mathcal{S}}_n \le C_p \sum_{1\le j\le N} \mathbb{E}\Big[\max_{0\le k\le n}|\hat{X}^{f,[\ell],j}_{t_k} - \hat{X}^{c,[\ell-1],j}_{t_k}|^p\Big]. \tag{2.13}$$

Then, it suffices to show that there exists a constant $C > 0$ such that:

$$\mathbb{E}\Big[\max_{0\le k\le n}|\hat{X}^{f,[\ell],j}_{t_k} - \hat{X}^{c,[\ell-1],j}_{t_k}|^p\Big] \le C\Big(\Delta^p_{\ell-1} + \Delta_{\ell-1}\sum_{0\le k\le n-1}\hat{\mathcal{S}}_k\Big), \tag{2.14}$$

which leads to the desired result by applying the discrete Grönwall inequality to (2.13). Recursive application of (B.5) and (B.8) given in Lemmas B.2 and B.3 respectively yields:

$$\begin{aligned}\hat{X}^{f,[\ell],j}_{t_k} - \hat{X}^{c,[\ell-1],j}_{t_k} &= \sum_{\substack{0\le i\le k-1\\0\le m\le d}}\big(\sigma^j_m(\hat{X}^{f,[\ell]}_{t_i}) - \sigma^j_m(\hat{X}^{c,[\ell-1]}_{t_i})\big)\Delta B^m_{t_{i+1},t_i}\\ &+ \sum_{\substack{0\le i\le k-1\\1\le m_1,m_2\le d}}\big(\mathcal{L}_{m_1}\sigma^j_{m_2}(\hat{X}^{f,[\ell]}_{t_i}) - \mathcal{L}_{m_1}\sigma^j_{m_2}(\hat{X}^{c,[\ell-1]}_{t_i})\big)\Delta\eta^{m_1m_2}_{t_{i+1},t_i} + \sum_{0\le i\le k-1}\big(\hat{\mathcal{M}}^j_{t_{i+1},t_i} + \hat{\mathcal{N}}^j_{t_{i+1},t_i}\big),\end{aligned} \tag{2.15}$$

where the remainder terms are such that $\mathbb{E}\big[\hat{\mathcal{M}}^j_{t_{i+1},t_i}|\mathcal{F}_{t_i}\big] = 0,\ 0 \le i \le 2^{\ell-1}-1$, and for any $p \ge 2$ there exist constants $C_1, C_2 > 0$ such that $\max_{0\le i\le 2^{\ell-1}-1}\mathbb{E}\big[|\hat{\mathcal{M}}^j_{t_{i+1},t_i}|^p\big] \le C_1\Delta^{3p/2}_{\ell-1}$ and $\max_{0\le i\le 2^{\ell-1}-1}\mathbb{E}\big[|\hat{\mathcal{N}}^j_{t_{i+1},t_i}|^p\big] \le C_2\Delta^{2p}_{\ell-1}$. Given (2.15), the bound (2.14) holds by following the same argument as in the proof of [11, Theorem 4.10], and we conclude. ∎

Also, we have that, for any $p \geq 1$ there exist constants $C_1, C_2 > 0$ such that:

$$\mathbb{E}\big[\max_{t \in \mathbf{g}^{c,[\ell-1]}} \|\bar{X}_t^{f,[\ell]} - \widetilde{X}_t^{f,[\ell]}\|^{2p}\big] \leq C_1 \Delta_{\ell-1}^p; \tag{2.16}$$
$$\mathbb{E}\big[\max_{t \in \mathbf{g}^{c,[\ell-1]}} \|\bar{X}_t^{c,[\ell-1]} - \widetilde{X}_t^{c,[\ell-1]}\|^{2p}\big] \leq C_2 \Delta_{\ell-1}^p,$$

which are obtained from the strong convergence rate of scheme (Weak-2) and the same argument used in the proof of [11, Lemma 4.6]. Hence, from Theorem 2.5, Lemma 2.4 and (2.16), we obtain the following result.

Corollary 2.6. *Let $\varphi \in C_b^2(\mathbb{R}^N; \mathbb{R})$ and $1 \leq \ell \leq L$. For any $p \geq 2$ there exists constant $C > 0$ such that $\mathbb{E}[(\mathcal{P}_{f,\ell}^{\varphi} - \mathcal{P}_{c,\ell-1}^{\varphi})^p] \leq C\, \Delta_{\ell-1}^p$.*

Remark 2.7. The AMLMC estimator under scheme (Weak-2) is designed to have four different integrations, as given in (2.4)-(2.7), while the antithetic estimator under the truncated Milstein scheme [11] uses three types of integrators without the antithetic coarse approximation $\widetilde{X}^{c,[\ell-1]}$. In the case of (Weak-2), use of only three integrators would lead to no improvement in strong convergence due to the presence of the term involving $\Delta\widetilde{A}_{t_{k+1},t_k}^{j_1 j_2}$ with a size of $\mathcal{O}(\Delta_\ell)$. $\widetilde{X}^{c,[\ell-1]}$ is exploited to deal with the above $\mathcal{O}(\Delta_\ell)$-term and obtain the higher rate of strong convergence (Theorem 2.5).

Remark 2.8. [1] constructed an AMLMC method based on the Ninomiya-Victoir (N-V) scheme [27], an alternative scheme of weak error 2. They showed that the strong error of the N-V scheme is 1 and then improved it with the technique of the antithetic multilevel estimator. The advantages of the proposed AMLMC based on (Weak-2) against that of the N-V scheme are summarized as follows: (i) Scheme (Weak-2) is always explicit while the N-V is a semi-closed scheme in the sense that it requires solving ODEs defined via the SDE coefficients and their solvability depends on the definition of coefficients; (ii) Our antithetic scheme uses four different integrators (2.4)-(2.7), while the antithetic estimator with the N-V scheme uses six integrators; (iii) Our (Weak-2) scheme is designed to be locally non-degenerate for both elliptic/hypo-elliptic settings (Section 2.1) as we explained in Section 2.2.3. Such a non-degenerate scheme is beneficial for the filtering problem as we described in Section 1.

2.5. AMLMC for forward problem. In order to estimate $\mathbb{E}[\varphi(X_T)]$, one simply needs to sample the systems (2.4)-(2.7) using the same source of randomness (i.e. the same Brownian motion and Gaussian variates) as implied in (2.4)-(2.7). We will sample these afore-mentioned systems multiple times (independently) so will use an argument '(i)' to indicate the i^{th}-sample. For instance, from (2.4), we will write $\bar{X}_{t_k}^{c,[\ell]}(i)$ for the i^{th}-sample associated to recursion (2.4) where the associated Brownian motion and Gaussians variates have been generated anew for each sample. Similarly, in the context of (2.9) we will write $\mathcal{P}_{f,\ell}^{\varphi}(i)$, $\mathcal{P}_{c,\ell-1}^{\varphi}(i)$ and $\mathcal{P}_0^{\varphi}(i)$.

The AMLMC procedure is as follows. We first set L and the sample sizes $M_0, \ldots, M_L$ to be used at each pair of levels; we will state below how this can be done. Then one can follow the approach in Algorithm 2.1. The new AMLMC estimator is given in (2.17) that is contained in Algorithm 2.1 and can be computed using any test function of interest when the underlying quantity $\mathbb{E}[\varphi(X_T)]$ is well defined.

To specify L and $M_0,\dots,M_L$ one can appeal to the results of Theorem 2.5, Corollary 2.6, as well as the weak error of the scheme (Weak-2) and follow standard computations in MLMC (e.g. [10]). That is, when considering the MSE, $\mathbb{E}\big[\big(\widehat{\mathbb{E}[\varphi(X_T)]}-\mathbb{E}[\varphi(X_T)]\big)^2\big]$, then under the assumptions made above, one has an upper-bound on the MSE as $\mathcal{O}\big(\sum_{0\le\ell\le L}\Delta_\ell^2/M_\ell+\Delta_L^4\big)$. Therefore, for $\epsilon>0$ given, one can achieve a MSE of $\mathcal{O}(\epsilon^2)$ by choosing $L=\mathcal{O}(\log(\epsilon^{-1/2}))$ and $M_\ell=\mathcal{O}(\epsilon^{-2}\Delta_\ell^{3/2})$. The cost to achieve this MSE is $\sum_{0\le\ell\le L}\Delta_\ell^{-1}M_\ell=\mathcal{O}(\epsilon^{-2})$ which is the best possible using stochastic Monte Carlo methods and was also obtained in [11]. In most practical simulations, one generally sets L as on standard computing equipment it is not feasible to generate beyond $L=10$ and this determines ϵ. Therefore, as the bias (weak error) of this method is $\mathcal{O}(\Delta_L^2)$, versus $\mathcal{O}(\Delta_L)$ in the antithetic Milstein method in [11], one might expect to see benefits for L's that are used in practice. We consider this in Section 4.

Algorithm 2.1 AMLMC using the weak second order scheme (Weak-2).

1. Input $L\ge1$ and $M_0,\dots,M_L$. Set $\ell=0$ and go to 2..
2. For $i=1,\dots,M_0$ independently simulate (2.4) to produce $\bar{X}_T^{c,[0]}(1),\dots,\bar{X}_T^{c,[0]}(M_0)$. Set $\ell=\ell+1$ and go to 3..
3. For $i=1,\dots,M_\ell$, independently simulate (2.4)-(2.7) to produce $\{\bar{X}_T^{c,[\ell-1]}(i)\}_{i=1}^{M_\ell}$, $\{\widetilde{X}_T^{c,[\ell-1]}(i)\}_{i=1}^{M_\ell}$, $\{\bar{X}_T^{f,[\ell-1]}(i)\}_{i=1}^{M_\ell}$, $\{\widetilde{X}_T^{f,[\ell-1]}(i)\}_{i=1}^{M_\ell}$. If $\ell\le L-1$, set $\ell=\ell+1$ go to the start of 3. otherwise go to 4..
4. Compute the MLMC estimator:

$$\widehat{\mathbb{E}[\varphi(X_T)]}:=\mathcal{P}_0^{\varphi,M_0}+\sum_{1\le\ell\le L}\{\mathcal{P}_{f,\ell}^{\varphi,M_\ell}-\mathcal{P}_{c,\ell-1}^{\varphi,M_\ell}\} \tag{2.17}$$

where $\mathcal{P}_0^{\varphi,M_0}:=\frac{1}{M_0}\sum_{1\le i\le M_0}\mathcal{P}_0^{\varphi}(i)$, $\mathcal{P}_{f,\ell}^{\varphi,M_\ell}:=\frac{1}{M_\ell}\sum_{1\le i\le M_\ell}\mathcal{P}_{f,\ell}^{\varphi}(i)$, $\mathcal{P}_{c,\ell-1}^{\varphi,M_\ell}:=\frac{1}{M_\ell}\sum_{1\le i\le M_\ell}\mathcal{P}_{c,\ell-1}^{\varphi}(i)$. Return (2.17) and stop.

3. Application to filtering.

3.1. State-space model. We consider a sequence of observations obtained sequentially and at unit times, $Y_1,Y_2,\dots,Y_k\in\mathbb{R}^N$, $k\in\mathbb{N}$. The assumption of unit times is mainly for simplicity of notation and any time grid could be considered. Associated to this sequence is an unobserved diffusion process exactly of the type (1.1). For the data, we shall assume that, at any time $k\in\mathbb{N}$, Y_k has a (bounded) positive probability density that depends only on the position, X_k, of the diffusion process at time k and is denoted $g(x_k,y_k)$. We denote the transition kernel of the diffusion process over a unit time and starting at $z\in\mathbb{R}^N$ as $Q(z,\cdot)$, for instance $\mathbb{E}[\varphi(X_1)]=\int_{\mathbb{R}^N}\varphi(x_1)Q(x,dx_1)$, where the expectation on the R.H.S. is w.r.t. the law of the diffusion (1.1), which we recall starts at $x\in\mathbb{R}^N$, and $\varphi:\mathbb{R}^N\to\mathbb{R}$ is bounded, measurable (the collection of such functions is denoted $\mathcal{B}_b(\mathbb{R}^N)$).

The object of interest is the filtering distribution. For any $k\in\mathbb{N}$ we define the filtering expectation:

$$\pi_k(\varphi):=\frac{\mathbb{E}\big[\varphi(X_k)\big\{\prod_{p=1}^k g(X_p,y_p)\big\}\big]}{\mathbb{E}\big[\big\{\prod_{p=1}^k g(X_p,y_p)\big\}\big]}. \tag{3.1}$$

Note that the fact that φ and $g(\cdot, y)$ are bounded (for any $y \in \mathbb{R}^N$) ensure that the filter is well-defined, but these assumptions are not needed in general – again we seek to simplify the discussion. We will compute a numerical approximation of (3.1) sequentially in time, as an exact computation is seldom possible.

In practice we often cannot (i) simulate from $Q(z, \cdot)$ and/or we may not have an (ii) explicit expression for the density of $Q(z, \cdot)$ or (iii) an unbiased estimate of such density. One of the afore-mentioned properties (i)-(iii) is needed in order to deploy numerical methods which are used in the approximation of the filter (3.1) in continuous time (see e.g. [18] for an explanation). Therefore we consider time discretization via the weak second order method (Weak-2), with step-size $\Delta_\ell = 2^{-\ell}$. Now, for any starting point $z \in \mathbb{R}^N$ and ending at a time 1 we denote the time discretised transition kernel as $Q^{[\ell]}(z, \cdot)$, for instance, $\mathbb{E}^{[\ell]}[\varphi(\bar{X}_1)] = \int_{\mathbb{R}^N} \varphi(x_1) Q^{[\ell]}(x, dx_1)$, where we have modified the notation of the expectation operator to $\mathbb{E}^{[\ell]}[\cdot]$ to emphasize dependence on the discretization level. We consider the approximation of the time discretised filter, $k \in \mathbb{N}$:

$$\pi_k^{[\ell]}(\varphi) := \frac{\mathbb{E}^{[\ell]}\big[\varphi(X_k)\big\{\prod_{p=1}^k g(X_p, y_p)\big\}\big]}{\mathbb{E}^{[\ell]}\big[\big\{\prod_{p=1}^k g(X_p, y_p)\big\}\big]}. \tag{3.2}$$

Note, to clarify, the R.H.S. of the above equation can be alternatively written as:

$$\frac{\int_{\mathbb{R}^{Nk}} \varphi(x_k)\big\{\prod_{p=1}^k g(x_p, y_p)\big\}\prod_{p=1}^k Q^{[\ell]}(x_{p-1}, dx_p)}{\int_{\mathbb{R}^{Nk}} \big\{\prod_{p=1}^k g(x_p, y_p)\big\}\prod_{p=1}^k Q^{[\ell]}(x_{p-1}, dx_p)}$$

where $x_0 = x$. Even with time discretization, one still needs to resort to numerical methods to approximate (3.2).

3.2. Multilevel particle filters.

Our objective is now to approximate the time discretised filter (3.2). We start with the ordinary particle filter (PF) which can do exactly the former task and is described in Algorithm 3.1. This algorithm presents the most standard and well-known PF with several possible extensions. Also note that the estimates of the filter, in equation (3.5) of Algorithm 3.1, are typically returned recursively in time.

The PF on its own is typically much less efficient than using a multilevel version, which has been developed and extended in several works; see e.g. [18, 19, 20, 23, 24] and [22] for a review. We describe the method of [18], except replacing the Euler-Maruyama discretization with the weak second order scheme. The basic idea is based upon the identity:

$$\pi_k^{[L]}(\varphi) = \pi_k^{[0]}(\varphi) + \sum_{1 \le \ell \le L} \big\{\pi_k^{[\ell]}(\varphi) - \pi_k^{[\ell-1]}(\varphi)\big\}. \tag{3.3}$$

We remark that on the R.H.S. of (3.3) one need not start at level 0, but we adopt this choice for ease of exposition. The idea is to use the PF to recursively approximate $\pi_k^{[0]}(\varphi)$ and then to use a coupled particle filter (CPF) for the approximation of $\pi_k^{[\ell]}(\varphi) - \pi_k^{[\ell-1]}(\varphi)$, independently for each index ℓ. The coupling is described in Algorithm **??** and then the CPF is given in Algorithm **??**, which are presented in Section **??** in the Supplementary Material.

Algorithm **??** presents a way to simulate a maximal coupling of two positive probability mass functions with the same support. It allows one to couple the resampling operation across two different levels of discretization as is done for a single level in Algorithm 3.1. This is then incorporated in Algorithm ?? which provides a way to approximate $\pi_k^{[\ell]}(\varphi) - \pi_k^{[\ell-1]}(\varphi)$ recursively in time.

The overall multilevel Particle Filter (MLPF) can be summarized as follows, given L the maximum level and the number of samples $M_0, \ldots, M_L$; we show how these parameters can be chosen below.

1. Run Algorithm 3.1 at level $\ell = 0$ with M_0 samples.
2. Independently of 1. for $\ell = 1, \ldots, L$, independently run Algorithm ?? in the Supplementary Material with M_ℓ samples.

Based on this process, a biased approximation of $\pi_k(\varphi)$ is then

$$\widehat{\pi_k(\varphi)} := \pi_k^{[0],M_0}(\varphi) + \sum_{1\le\ell\le L} \{\pi_k^{[\ell],M_\ell}(\varphi) - \pi_k^{[\ell-1],M_\ell}(\varphi)\},$$

where $\pi_k^{[\ell],M}(\varphi)$ is the PF estimate of $\pi_k^{[\ell]}(\varphi)$ with the number of particles M specifically given in (3.5). The bias of this approximation is from the discretization level L and the bias of the PF/CPF approximation, e.g. that in general, $\mathbb{E}\big[\pi_k^{[\ell],M_\ell}(\varphi) - \pi_k^{[\ell-1],M_\ell}(\varphi)\big] \neq \pi_k^{[\ell]}(\varphi) - \pi_k^{[\ell-1]}(\varphi)$, where $\mathbb{E}$ is used to denote the expectation w.r.t. the probability law used in generating our estimators. Now, if one combines the theory in [16] for the weak error, the strong error result in Proposition 2.3 and the results in [18] one can consider the MSE, $\mathbb{E}\big[\big(\widehat{\pi_k(\varphi)} - \pi_k(\varphi)\big)^2\big]$. Under the assumptions in the current paper and in [18] it can be proved that the MSE has an upper-bound which is:

$$\mathcal{O}\Big(\sum_{0\le\ell\le L} \Delta_\ell^{1/2}/M_\ell + \Delta_L^4\Big). \tag{3.4}$$

We do not prove this bound as it is a fairly trivial application of the results in the aforementioned papers. The exponent of Δ_ℓ, in the summand, is 1/2 and this reduction of the strong error of Euler-Maruyama is due to the resampling mechanism that has been employed; we do not know of any general method that can maintain the strong error rate. We also remark that there is an additional additive term on the R.H.S., but this term is much smaller than the term given above, so we need not consider it. Using the standard approach that has been adopted in MLMC (i.e. as discussed in Section 2.5) one can show that for $\epsilon > 0$ given, setting $L = \mathcal{O}(\log(\epsilon^{-1/2}))$, $M_\ell = \epsilon^{-2}\Delta_\ell^{3/4}\Delta_L^{-1/4}$ gives a MSE of $\mathcal{O}(\epsilon^2)$ for a cost (per time step k) of $\mathcal{O}(\epsilon^{-2.25})$. This is lower than the cost of the approach in [18] due to the increased weak error relative to the Euler-Maruyama discretization used in [18].

In the recent work of [24], the authors show how to use the antithetic Milstein scheme within the context of the MLPF; we abbreviate to AMMLPF (antithetic Milstein MLPF). They show empirically that to achieve a MSE (associated to their estimator) of $\mathcal{O}(\epsilon^2)$ there is a cost (per time step k) of $\mathcal{O}\big(\epsilon^{-2}\log(\epsilon)^2\big)$. The objective now is to show how our new antithetic MLMC method can be extended to MLPFs. As in the case of MLMC, we expect for this new method the error-cost calculation to be of the same order as the AMMLPF, but when using smaller L, as would be adopted in practice, that improvements are seen in simulations, due to the decreased weak error.

Algorithm 3.1 Particle Filter using the weak second order scheme (Weak-2). The algorithm is stopped at a time T, but need not be.

1. Input: level of discretization $\ell \in \mathbb{N}_0$, final time $T \in \mathbb{N}$ and number of samples M. Set $\bar{X}_0^{[\ell]}(i) = x$, $i = 1, \dots, M$ and $k = 1$. Go to 2..
2. Sampling: For $i = 1, \dots, M$, simulate $\bar{X}_k^{[\ell]}(i)|\bar{x}_{k-1}^{[\ell]}(i)$ using the dynamics (Weak-2) up-to time 1, with starting point $\bar{x}_{k-1}^{[\ell]}(i)$ and step-size Δ_ℓ. Go to 3..
3. Resampling: For $i = 1, \dots, M$ compute: $w_k^{[\ell]}(i) := \frac{g(\bar{X}_k^{[\ell]}(i), y_k)}{\sum_{j=1}^M g(\bar{X}_k^{[\ell]}(j), y_k)}$. For any $\varphi \in \mathcal{B}_b(\mathbb{R}^N)$ we have the estimate:
$$\pi_k^{[\ell],M}(\varphi) := \sum_{1 \le i \le M} w_k^{[\ell]}(i)\varphi(\bar{X}_k^{[\ell]}(i)). \tag{3.5}$$
For $i = 1, \dots, M$ sample an index $j(i) \in \{1, \dots, M\}$ using the probability mass function $w_k^{[\ell]}(\cdot)$ and set $\check{X}_k^{[\ell]}(i) = \bar{X}_k^{[\ell]}(j(i))$. For $i = 1, \dots, M$, set $\bar{X}_k^{[\ell]}(i) = \check{X}_k^{[\ell]}(i)$, $k = k + 1$, if $k = T + 1$ go to 4. otherwise go to 2..
4. Return the estimates $\pi_1^{[\ell],M}(\varphi), \dots, \pi_T^{[\ell],M}(\varphi)$ from (3.5).

3.3. New multilevel particle filter. Our new MLPF, which we shall call the antithetic multilevel Particle Filter (AMLPF), is similar to the approach that was illustrated in the previous section. At level 0, we shall use a PF to approximate $\pi_k^{[0]}(\varphi)$. To approximate the differences $\pi_k^{[\ell]}(\varphi) - \pi_k^{[\ell-1]}(\varphi)$ we shall use a combination of the antithetic MLMC weak second order scheme of Section 2.4, which will be the 'sampling' part of a PF and a type of 'coupling' for the 'resampling step'. As we have already introduced the former, we introduce the latter as Algorithm **??** in Section **??** in the Supplementary Material. As has been commented by [23] in the context of coupling two probability mass functions (as in Algorithm **??**) there is nothing that is optimal about using Algorithm **??**. Indeed, when used as part of a MLPF, we expect just as in the case of Algorithm **??** when used for Algorithm **??**, the strong error rate from the forward problem is reduced by a factor of two; see (3.4). It remains an open problem to find a general coupling method which can maintain the forward error rate (as was the case in [3] in dimension 1 only) and a linear complexity in terms of the samples M.

Given Algorithm **??**, we are now in a position to give our new coupled particle filter in Algorithm 3.2. Just as in the previous section, the AMLPF can be summarized as follows, given L the maximum level and the number of samples $M_0, \dots, M_L$; we show how these parameters can be chosen below.

1. Run Algorithm 3.1 at level $\ell = 0$ with M_0 samples.
2. Independently of 1. for $\ell = 1, \dots, L$, independently run Algorithm 3.2 with M_ℓ samples.

Thus our new approximation of $\pi_k(\varphi)$ is:
$$\widetilde{\pi_k(\varphi)} := \pi_k^{[0],M_0}(\varphi) + \sum_{1 \le \ell \le L} \{\hat{\pi}_k^{[\ell],M_\ell}(\varphi) - \hat{\pi}_k^{[\ell-1],M_l}(\varphi)\}.$$
where we recall that $\hat{\pi}_k^{[\ell],M_\ell}(\varphi) - \hat{\pi}_k^{[\ell-1],M_l}(\varphi)$ is given in (3.6) in Algorithm 3.2.

We can again consider the MSE $\mathbb{E}\big[\big(\widetilde{\pi_k(\varphi)} - \pi_k(\varphi)\big)^2\big]$. As noted in [24], which considers the AMMLPF, although it is fairly easy to establish a bound on the R.H.S. which is of the type (up-to some other terms which are smaller) $\mathcal{O}(\sum_{0\leq\ell\leq L}\Delta_l^{\nu}/M_l+\Delta_L^4)$, obtaining the value of ν that is observed in simulation is not easy to achieve with the current proof method that has been adopted in [18, 24]. As a result, we do not give a theoretical analysis in this paper. However, as we shall see in Section 4, it appears that the correct value of $\nu = 1$ and hence we use this as our guideline to choose $L, M_0, \dots, M_L$. Following the arguments that have been used previously, for $\epsilon > 0$ given, setting $L = \mathcal{O}(\log(\epsilon^{-1/2}))$, $M_\ell = \epsilon^{-2}\Delta_\ell L$ gives a MSE of $\mathcal{O}(\epsilon^2)$ for a cost (per time step k) of $\mathcal{O}\big(\epsilon^{-2}\log(\epsilon)^2\big)$.

4. Numerical results.

In this section, we provide a series of numerical illustrations detailing our methodology for both forward and filtering problems. Specifically, we compare their performance against both multilevel and standard Monte Carlo (Std MC) methods and particle filters. We here summarise the labels of the algorithms that we use in the numerics:

- Forward problem: Std MC, MLMC (standard method with scheme (Weak-2)), AMLMC (the new antithetic method with scheme (Weak-2)) and AMMLMC (the antithetic method of [11] with scheme (T-Milstein)).
- Filtering problem: PF, MLPF (standard method, using scheme (Weak-2)), AMLPF (the new antithetic PF method with scheme (Weak-2)) and AMMLPF (the antithetic PF method studied in [24] with scheme (T-Milstein)).

4.1. Models.

We consider two SDE models in our experiments. The first model is the stochastic FitzHugh-Nagumo (FHN) model, which is a well-known hypo-elliptic model in neuroscience:

$$dX_t = \tfrac{1}{\epsilon}\big(X_t - X_t^3 - Z_t - s\big)\,dt, \quad dZ_t = (\gamma X_t - Z_t + \beta)\,dt + \sigma dB_t^1.$$

The values of the parameters in the simulations are set as follows: $X_{t_0} = 0$, $Y_{t_0} = 0$, $\epsilon = 0.1$, $\sigma = 0.3$, $\gamma = 1.5$, $\beta = 0.3$ and $s = 0.01$. For the forward problem, we estimate the value of $\mathbb{E}[X_T]$ with $T = 100$ time units. For the filtering case, we estimate $\mathbb{E}[X_n|y_{0:n}]$ with $n = 100$. The observation data y_k we choose is $y_k \mid (X_{k\delta}, Z_{k\delta}) \sim \mathcal{N}(X_{k\delta}, \tau^2)$ with $\delta = 1$, $\tau = 0.1$, where $\mathcal{N}(m, \sigma^2)$ denotes the Gaussian distribution of mean m and variance σ^2.

The second model example is the Heston model [15] given as an elliptic SDE not satisfying the commutative condition (2.2):

$$dS_t = rS_t dt + \sqrt{v_t}S_t dB_t^1, \quad dv_t = \alpha(\theta - v_t)dt + \mu\sqrt{v_t}(\rho dB_t^1 + \sqrt{1-\rho^2}dB_t^2).$$

The values of the parameters used in the simulations are set as: $S_{t_0} = 100$, $v_{t_0} = 0.09$, $r = 0.04$, $\alpha = 2.0$, $\theta = 0.09$, $\mu = 0.1$ and $\rho = 0.7$. For the forward problem, our target quantity is $\mathbb{E}[S_T]$ with $T = 1.0$. For the filtering case, we estimate $\mathbb{E}[S_n|y_{0:n}]$ with $n = 100$, where each observation y_k is obtained as $y_k \mid (S_{k\delta}, v_{k\delta}) \sim \mathcal{N}(S_{k\delta}, \tau^2)$ with $\delta = 0.01$ and $\tau = 2$. We stress here that in the above model settings, the test functions are unbounded for the filtering problem, while we have assumed boundedness in Section 3. As we will show in the numerical results below, such a discrepancy can be negligible under suitable scenarios; e.g. the case where the moments of underlying process are uniformly bounded in the time-interval.

4.2. Set-Up and results. For our numerical experiments, we applied our algorithms to obtain the multilevel estimators. Given the unavailability of an analytical solution, we will use std MC and PF with a high-resolution $L = 9$ to approximate the ground truth for the forward and filtering problem, respectively that shall serve as the benchmark solution. For the filtering problem, though we did not discuss about stochastic resampling for our proposed AMLPF in Section 3, we will run particle filters with adaptive resampling to showcase the practical extendability of AMLPF. Specifically, resampling is performed when the effective sample size (ESS) is less than $\frac{1}{2}$ of the particle numbers. For the coupled filters, we use the ESS of the coarse filter as the measurement of discrepancy. The error within the estimators in our simulations will be evaluated using the mean square error (MSE), which will be computed by conducting 50 independent simulations for each method (Std MC, MLMC, AMLMC and AMMLMC) for the forward problem, and (PF, MLPF, AMLPF and AMMLPF) for the filtering case with the ground truth obtained as described above.

The primary target is to compare the costs of these methods at the same MSE level. In the AMLMC and AMLPF, one needs to determine the number of samples to approximate the multilevel estimators at levels ℓ and $\ell - 1$, denoted by M_ℓ. In particular, we set M_ℓ for the AMLMC and AMLPF as $M_\ell = c_{1,\ell} \times \varepsilon^{-2}\Delta_\ell^{3/2}$ and $M_\ell = c_{2,\ell} \times \varepsilon^{-2}\Delta_\ell L$, respectively, for some constants $c_{1,\ell}, c_{2,\ell} > 0$ and a given L to attain a target MSE of $\mathcal{O}(\varepsilon^2)$, $\varepsilon > 0$, with a cost of $\mathcal{O}(\varepsilon^{-2})$ for AMLMC and $\mathcal{O}(\epsilon^{-2}\log(\epsilon)^2)$ for AMLPF. For the AMMLMC and AMMLPF, we also choose M_ℓ as above. In our experiments, we initially simulate the Std MC and PF algorithms with $L \in \{1, 2, 3, 4\}$ and obtain the corresponding MSE and cost values, where the computational cost is computed as $\sum_{\ell=0}^{L} M_\ell/\Delta_\ell$. Subsequently, we use the MLMC and MLPF estimators to achieve identical MSE levels and record their corresponding cost values. Finally, we compute the AMLMC and AMLPF estimators to attain similar MSE levels and note their respective cost values. Due to the lower order of weak convergence, the AMMLMC and AMMLPF estimators are computed with $L = \{2, 4, 6, 8\}$.

We present our numerical simulations to show the benefits of applying AMLMC/AMLPF to the above SDE models, compared to Std MC, MLMC, AMMLMC/PF, MLPF, AMMLPF. Figures 1-2 show the MSE against the cost. The figures show that as we increase the levels from $L = 1$ to $L = 4$, the difference in the cost between the methods also increases. Table 2 presents the estimated change rates of log(cost) against log(MSE) for both problems. The reported rates align with our theoretical expectations. We observe that the computational costs are of sizes consistent to the theoretical ones of $\mathcal{O}(\epsilon^{-3})$ for the Std MC and PF, $\mathcal{O}(\epsilon^{-2})$ for the AMLMC, and $\mathcal{O}(\epsilon^{-2}\log(\epsilon)^2)$ for the AMLPF. Moreover, we see from the bottom two plots of Figures 1-2 that AMLMC/AMLPF (using the weak second order scheme) outperformed AMMLMC/AMMLPF (using the truncated Milstein scheme) in terms of cost vs MSE. We note that when choosing the number of samples M_ℓ in the experiments, the constants $c_{1,\ell}$ and $c_{2,\ell}$ to determine M_ℓ (indicated above) are allowed to be set lower for the case of the weak second order scheme compared with that of the truncated Milstein scheme. We expect this is due to the tighter variance bounds for the couplings of the AMLMC under a small-noise diffusion setting, i.e. the case some small parameter is contained in the diffusion coefficient, which we detail in Section **??** in Supplementary Material.

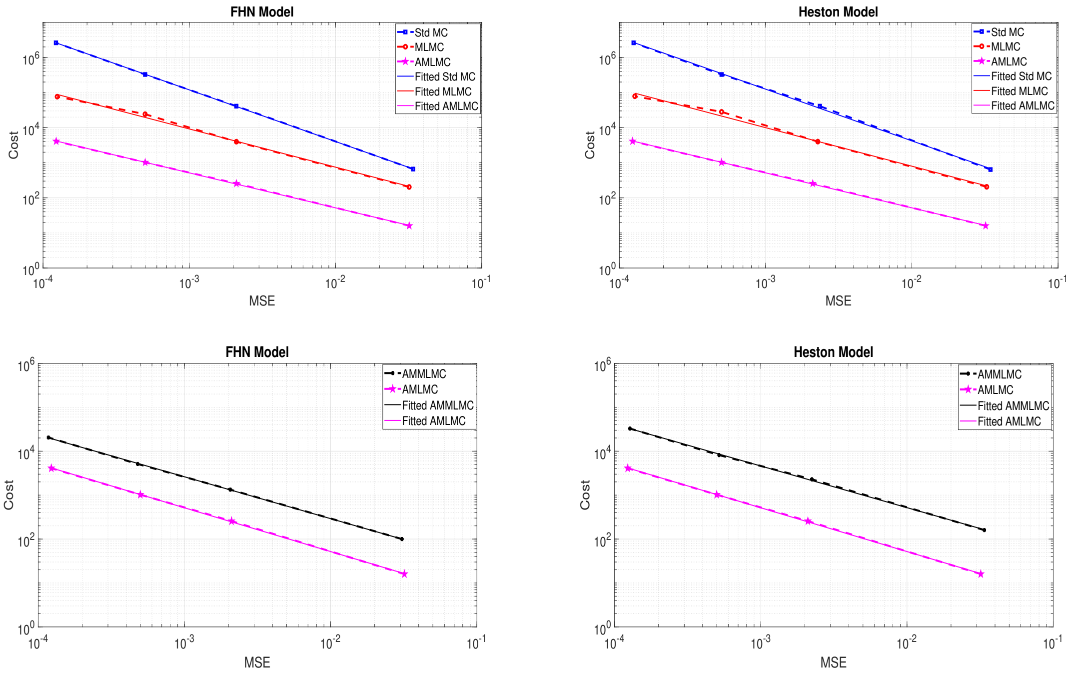

Figure 1. *Cost versus MSE for the forward problem.*

Table 2

Estimated change rate of log(cost) against log(MSE). Left: Forward problem. Right: Filtering problem.

Model	Std MC	MLMC	AMLMC
FHN	-1.48	-1.1	-1.03
Heston	-1.47	-1.11	-1.05

Model	PF	MLPF	AMLPF
FHN	-1.46	-1.17	-1.11
Heston	-1.49	-1.24	-1.14

5. Conclusion.

Our work has investigated the use of a weak second order scheme within the multilevel Monte Carlo (MLMC) framework. We first proved that our scheme has a strong error 1. Then, in the context of MLMC, we developed a new antithetic estimator based on our weak second order scheme which achieves the optimal cost rate $\mathcal{O}(\varepsilon^{-2})$, $\varepsilon > 0$, to obtain a MSE of $\mathcal{O}(\varepsilon^2)$. Such an optimal cost rate is also reported for the different antithetic MLMC approach of [11] which makes use of a truncated Milstein scheme of weak error 1. The new antithetic estimator is shown to possess a benefit versus the one of [11], that is, our estimator is expected to be more efficient for a finite maximum level of discretization L used in practice due to the higher order weak convergence. As an application, we have proposed an antithetic multilevel particle filter (AMLPF) by building upon previous works [18, 24] for the purposes of efficient filtering of diffusion processes from observations. Our simulation studies are in support of the anticipated cost of the proposed AMLPF being $\mathcal{O}(\varepsilon^{-2}\log(\varepsilon)^2)$ to achieve an MSE of $\mathcal{O}(\varepsilon^2)$. Also, all our numerics support the understanding that the new antithetic estimator using the weak second order scheme outperforms the antithetic Milstein scheme-

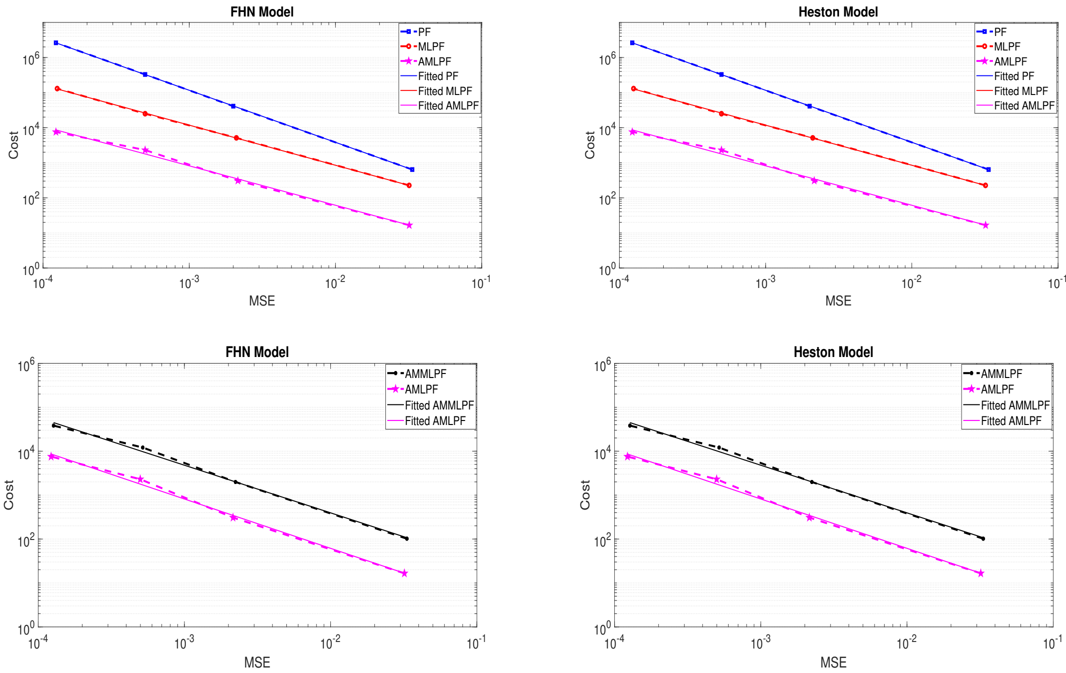

Figure 2. *Cost versus MSE for the filtering problem.*

based estimator in both forward/filtering problems. We emphasize that our numerical scheme is locally non-degenerate under both elliptic/hypo-elliptic settings, whereas the truncated Milstein scheme is degenerate in the hypo-elliptic case. The non-degeneracy of the scheme makes possible its deployment within particle filters with guided proposals so that stochastic weights required to be assigned to particles are well-defined and available as the ratio of products involving the density expression for the numerical scheme and the proposal, though the exploration of this direction is left for future work.

Appendix A. Proof of Proposition 2.3.

Proof. Let $1 \le i \le 2^\ell$. We have that

$$\mathcal{S}_i \equiv \mathbb{E}\Big[\max_{0 \le k \le i} \|X_{t_k} - \bar{X}_{t_k}\|^p\Big] \le N^{p-1} \sum_{1 \le j \le N} \mathbb{E}\Big[\max_{0 \le k \le i} |X^j_{t_k} - \bar{X}^j_{t_k}|^p\Big],$$

where we made use of the following inequality:

$$\Big(\sum_{1 \le j \le N} |x_j|\Big)^p \le N^{p-1} \sum_{1 \le j \le N} |x_j|^p, \quad x = (x_1, \dots, x_N) \in \mathbb{R}^N. \tag{A.1}$$

We will show that for any $p \ge 2$, there exists a constant $C > 0$ such that:

$$\mathcal{S}_i \le C\Big(\Delta_\ell^{p/2} + \sum_{0 \le n \le i-1} \mathcal{S}_n \cdot \Delta_\ell\Big), \tag{A.2}$$

which leads to the conclusion due to the discrete Gronwall's inequality. Application of stochastic Taylor expansion for $X^j_{t_n}$, $0\le n\le k-1$, yields that, for $1\le j\le N$:

$$\begin{aligned}
X^j_{t_k} - \bar{X}^j_{t_k} &= \sum_{\substack{0\le n\le k-1\\ 0\le m\le d}} \int_{t_n}^{t_{n+1}} \big(\sigma^j_m(X_s) - \sigma^j_m(\bar{X}_{t_n})\big) dB^m_s \\
&\quad - \sum_{\substack{0\le n\le k-1\\ 0\le m_1,m_2\le d}} \mathcal{L}_{m_1}\sigma^j_{m_2}(\bar{X}_{t_n})\Delta\eta^{m_1m_2}_{t_{n+1},t_n} - \tfrac{1}{2}\sum_{\substack{0\le n\le k-1\\ 1\le m_1<m_2\le d}} \big[\sigma_{m_1},\sigma_{m_2}\big]^j(\bar{X}_{t_n})\Delta\widetilde{A}^{m_1m_2}_{t_{n+1},t_n} \\
&= \sum_{\substack{0\le n\le k-1\\ 0\le m\le d}} \big(\sigma^j_m(X_{t_n}) - \sigma^j_m(\bar{X}_{t_n})\big)\Delta B^m_{t_{n+1},t_n} \\
&\quad + \tfrac{1}{2}\sum_{\substack{0\le n\le k-1\\ 1\le m_1,m_2\le d}} \big(\mathcal{L}_{m_1}\sigma^j_{m_2}(X_{t_n}) - \mathcal{L}_{m_1}\sigma^j_{m_2}(\bar{X}_{t_n})\big)\big\{\Delta B^{m_1}_{t_{n+1},t_n}\Delta B^{m_2}_{t_{n+1},t_n} - \Delta_\ell \mathbf{1}_{m_1=m_2}\big\} \\
&\quad - \tfrac{1}{2}\sum_{\substack{0\le n\le k-1\\ 1\le m_1<m_2\le d}} \big\{[\sigma_{m_1},\sigma_{m_2}]^j(X_{t_n})\Delta A^{m_1m_2}_{t_{n+1},t_n} + [\sigma_{m_1},\sigma_{m_2}]^j(\bar{X}_{t_n})\Delta\widetilde{A}^{m_1m_2}_{t_{n+1},t_n}\big\} \\
&\quad + \sum_{0\le n\le k-1} \big(\mathcal{M}^j_{t_{n+1},t_n} + \mathcal{N}^j_{t_{n+1},t_n}\big),
\end{aligned}$$

where the terms $\mathcal{M}^j_{t_{n+1},t_n}$, $\mathcal{N}^j_{t_{n+1},t_n}$ are such that $\mathbb{E}\big[\mathcal{M}^j_{t_{n+1},t_n}|\mathcal{F}_{t_n}\big]=0$ for $0\le n\le k-1$ and it holds under Assumption 2.1 that for any $p\ge 2$, there exist constants $C_1, C_2>0$ such that

$$\text{(A.3)}\qquad \max_{0\le n\le k-1}\mathbb{E}\big[|\mathcal{M}^j_{t_{n+1},t_n}|^p\big] \le C_1\Delta_\ell^{3p/2}, \qquad \max_{0\le n\le k-1}\mathbb{E}\big[|\mathcal{N}^j_{t_{n+1},t_n}|^p\big] \le C_2\Delta_\ell^{2p}.$$

Thus, inequality (A.1) yields $\mathbb{E}\big[\max_{0\le k\le i}\big|X^j_{t_k} - \bar{X}^j_{t_k}\big|^p\big] \le C_p\sum_{1\le\alpha\le 6}\mathcal{T}^{(\alpha),j}_i$ for some constant $C_p>0$, where we have set:

$$\mathcal{T}^{(1),j}_i = \mathbb{E}\big[\max_{0\le k\le i}\big|\sum_{0\le n\le k-1}\big(\sigma^j_0(X_{t_n}) - \sigma^j_0(\bar{X}_{t_n})\big)\Delta_\ell\big|^p\big];$$

$$\mathcal{T}^{(2),j}_i = \mathbb{E}\big[\max_{0\le k\le i}\big|\sum_{\substack{0\le n\le k-1\\ 1\le m\le d}}\big(\sigma^j_m(X_{t_n}) - \sigma^j_m(\bar{X}_{t_n})\big)\Delta B^m_{t_{n+1},t_n}\big|^p\big];$$

$$\mathcal{T}^{(3),j}_i = \mathbb{E}\big[\max_{0\le k\le i}\big|\sum_{\substack{0\le n\le k-1\\ 1\le m_1,m_2\le d}}\big(\mathcal{L}_{m_1}\sigma^j_{m_2}(X_{t_n}) - \mathcal{L}_{m_1}\sigma^j_{m_2}(\bar{X}_{t_n})\big)\big(\Delta B^{m_1}_{t_{n+1},t_n}\Delta B^{m_2}_{t_{n+1},t_n} - \Delta_\ell\mathbf{1}_{m_1=m_2}\big)\big|^p\big];$$

$$\mathcal{T}^{(4),j}_i = \mathbb{E}\big[\max_{0\le k\le i}\big|\sum_{\substack{0\le n\le k-1\\ 1\le m_1<m_2\le d}}\big\{[\sigma_{m_1},\sigma_{m_2}]^j(X_{t_n})\Delta A^{m_1m_2}_{t_{n+1},t_n} + [\sigma_{m_1},\sigma_{m_2}]^j(\bar{X}_{t_n})\Delta\widetilde{A}^{m_1m_2}_{t_{n+1},t_n}\big\}\big|^p\big];$$

$$\mathcal{T}^{(5),j}_i = \mathbb{E}\big[\max_{0\le k\le i}\big|\sum_{0\le n\le k-1}\mathcal{M}^j_{t_{n+1},t_n}\big|^p\big], \quad \mathcal{T}^{(6),j}_i = \mathbb{E}\big[\max_{0\le k\le i}\big|\sum_{0\le n\le k-1}\mathcal{N}^j_{t_{n+1},t_n}\big|^p\big].$$

Applying inequality (A.1), we have under Assumption 2.1 that:

$$\mathcal{T}^{(1),j}_i \le i^{p-1}\sum_{0\le n\le i-1}\mathbb{E}\big[|\sigma^j_0(X_{t_n}) - \sigma^j_0(\bar{X}_{t_n})|^p\big]\Delta_\ell^p \le c_1T^{p-1}\sum_{0\le n\le i-1}\mathcal{S}_n\cdot\Delta_\ell$$

for some constant $c_1 > 0$ independent of Δ_ℓ since $i\Delta_\ell \le T$. Similarly, we have:

$$\mathcal{T}_i^{(6),j} \le i^{p-1} \sum_{0\le n \le i-1} \mathbb{E}\big[|\mathcal{N}^j_{t_{n+1},t_n}|^p\big] \le c_6 T^p \Delta_\ell^p \tag{A.4}$$

for some constant $c_6 > 0$. We consider the other four terms. Since they involve martingales, we make use of the discrete Burkholder-Davis-Gundy inequality to obtain:

$$\begin{aligned}\mathcal{T}_i^{(2),j} &\le c_{2,1}\mathbb{E}\Big[\Big(\sum_{0\le n\le i-1}\sum_{1\le m\le d}\big\{(\sigma_m^j(X_{t_n}) - \sigma_m^j(\bar{X}_{t_n}))\Delta B^m_{t_{n+1},t_n}\big\}^2\Big)^{p/2}\Big] \\ &\le c_{2,2}\, i^{p/2-1} \sum_{0\le n\le i-1}\sum_{1\le m\le d}\mathbb{E}\big[\big|(\sigma_m^j(X_{t_n}) - \sigma_m^j(\bar{X}_{t_n}))\Delta B^m_{t_{n+1},t_n}\big|^p\big] \\ &\le c_{2,3}\, i^{p/2-1}\sum_{0\le n\le i-1}\mathcal{S}_n\cdot\Delta_\ell^{p/2} \le c_{2,3}\, T^{p/2-1}\sum_{0\le n\le i-1}\mathcal{S}_n\cdot\Delta_\ell\end{aligned}$$

for some constants $c_{2,1}, c_{2,2}, c_{2,3} > 0$, where we applied (A.1) in the second inequality. Similarly, we have that:

$$\begin{aligned}\mathcal{T}_i^{(3),j} &\le c_3 i^{p/2-1}\sum_{0\le n\le i-1}\mathcal{S}_n\cdot\Delta_\ell^p \le c_3 T^{p/2-1}\Delta_\ell^{p/2}\sum_{0\le n\le i-1}\mathcal{S}_n\cdot\Delta_\ell; \\ \mathcal{T}_i^{(5),j} &\le c_{5,1}\, i^{p/2-1}\sum_{0\le n\le i-1}\mathbb{E}\big[\big|\mathcal{M}^j_{t_{n+1},t_n}\big|^p\big] \le c_{5,2}\, i^{p/2}\Delta_\ell^{3p/2} = c_{5,2}\, T^{p/2}\Delta_\ell^p\end{aligned}$$

for some constants $c_3, c_{5,1}, c_{5,2} > 0$. Finally, for $\mathcal{T}_i^{(4),j}$, we obtain:

$$\mathcal{T}_i^{(4),j} \le c_{4,1} i^{p/2-1}\sum_{0\le n\le i-1}\sum_{1\le m_1<m_2\le d}\mathbb{E}\big[|\Delta A^{m_1m_2}_{t_{n+1},t_n}|^p + |\Delta\widetilde{A}^{m_1m_2}_{t_{n+1},t_n}|^p\big] \le c_{4,2}T^{p/2}\Delta_\ell^{p/2} \tag{A.5}$$

for constants $c_{4,1}, c_{4,2} > 0$, where we used that $\mathbb{E}|\Delta A^{m_1m_2}_{t_{n+1},t_n}|^p = \mathcal{O}(\Delta_\ell^p)$, $\mathbb{E}|\Delta\widetilde{A}^{m_1m_2}_{t_{n+1},t_n}|^p = \mathcal{O}(\Delta_\ell^p)$ for any $p \ge 2$. Note that $\sum_{0\le n\le i-1}\mathcal{S}_n$ does not appear in the upper bound of $\mathcal{T}_i^{(4),j}$. Thus, we obtain inequality (A.2) and conclude. ■

Appendix B. Auxiliary results for Theorem 2.5.

Throughout this section, let $1 \le j \le N$, $1 \le \ell \le L$, $0 \le k \le 2^{\ell-1}-1$ and $t_k = k\Delta_{\ell-1}$.

Lemma B.1. *It holds that:*

$$\begin{aligned}\bar{X}^{f,[\ell],j}_{t_{k+1}} &= \bar{X}^{f,[\ell],j}_{t_k} + \sum_{0\le m\le d}\sigma_m^j\big(\bar{X}^{f,[\ell]}_{t_k}\big)\Delta B^m_{t_{k+1},t_k} + \sum_{0\le m_1,m_2\le d}\mathcal{L}_{m_1}\sigma^j_{m_2}\big(\bar{X}^{f,[\ell]}_{t_k}\big)\Delta\eta^{m_1m_2}_{t_{k+1},t_k} \\ &\quad -\tfrac{1}{2}\sum_{1\le m_1,m_2\le d}\mathcal{L}_{m_1}\sigma^j_{m_2}\big(\bar{X}^{f,[\ell]}_{t_k}\big)\big(\Delta B^{m_1}_{t_{k+1},t_{k+1/2}}\Delta B^{m_2}_{t_{k+1/2},t_k} - \Delta B^{m_1}_{t_{k+1/2},t_k}\Delta B^{m_2}_{t_{k+1},t_{k+1/2}}\big) \\ &\quad +\tfrac{1}{2}\sum_{1\le m_1<m_2\le d}\big[\sigma_{m_1},\sigma_{m_2}\big]^j\big(\bar{X}^{f,[\ell]}_{t_k}\big)\big(\Delta\widetilde{A}^{m_1m_2}_{t_{k+1/2},t_k} + \Delta\widetilde{A}^{m_1m_2}_{t_{k+1},t_{k+1/2}}\big) + \bar{\mathcal{M}}^{f,j}_{t_{k+1},t_k} + \bar{\mathcal{N}}^{f,j}_{t_{k+1},t_k},\end{aligned}$$

where the remainder terms are such that $\mathbb{E}\big[\bar{\mathcal{M}}^{f,j}_{t_{k+1},t_k}|\mathcal{F}_{t_k}\big] = 0$, and for any $p \ge 2$ there exist constants $C_1, C_2 > 0$ so that:

$$\max_{0\le k\le 2^{\ell-1}-1}\mathbb{E}\big[|\bar{\mathcal{M}}^{f,j}_{t_{k+1},t_k}|^p\big] \le C_1\Delta_{\ell-1}^{3p/2}, \quad \max_{0\le k\le 2^{\ell-1}-1}\mathbb{E}\big[|\bar{\mathcal{N}}^{f,j}_{t_{k+1},t_k}|^p\big] \le C_2\Delta_{\ell-1}^{2p}.$$

Similarly, it holds that:

$$\widetilde{X}^{f,[\ell],j}_{t_{k+1}} = \widetilde{X}^{f,[\ell],j}_{t_k} + \sum_{0\le m\le d} \sigma^j_m\big(\widetilde{X}^{f,[\ell]}_{t_k}\big)\Delta B^m_{t_{k+1},t_k} + \sum_{0\le m_1,m_2\le d} \mathcal{L}_{m_1}\sigma^j_{m_2}\big(\widetilde{X}^{f,[\ell]}_{t_k}\big)\Delta\eta^{m_1m_2}_{t_{k+1},t_k}$$
$$+ \frac{1}{2}\sum_{1\le m_1,m_2\le d} \mathcal{L}_{m_1}\sigma^j_{m_2}\big(\widetilde{X}^{f,[\ell]}_{t_k}\big)\big(\Delta B^{m_1}_{t_{k+1},t_{k+1/2}}\Delta B^{m_2}_{t_{k+1/2},t_k} - \Delta B^{m_1}_{t_{k+1/2},t_k}\Delta B^{m_2}_{t_{k+1},t_{k+1/2}}\big)$$
$$- \frac{1}{2}\sum_{1\le m_1<m_2\le d} \big[\sigma_{m_1},\sigma_{m_2}\big]^j\big(\widetilde{X}^{f,[\ell]}_{t_k}\big)\big(\widetilde{A}^{m_1m_2}_{t_{k+1/2},t_k} + \widetilde{A}^{m_1m_2}_{t_{k+1},t_{k+1/2}}\big) + \widetilde{\mathcal{M}}^{f,j}_{t_{k+1},t_k} + \widetilde{\mathcal{N}}^{f,j}_{t_{k+1},t_k},$$

where the remainder terms $\widetilde{\mathcal{M}}^{f,j}_{t_{k+1},t_k}$ *and* $\widetilde{\mathcal{N}}^{f,j}_{t_{k+1},t_k}$ *satisfy the same properties as* $\bar{\mathcal{M}}^{f,j}_{t_{k+1},t_k}$ *and* $\bar{\mathcal{N}}^{f,j}_{t_{k+1},t_k}$*, respectively.*

Proof. From the definition of the fine discretization scheme (2.6), we have:

$$\text{(B.1)}\quad \bar{X}^{f,[\ell],j}_{t_{k+1}} = \bar{X}^{f,[\ell],j}_{t_k} + \sum_{0\le m\le d}\Big\{\sigma^j_m(\bar{X}^{f,[\ell]}_{t_k})\Delta B^m_{t_{k+1/2},t_k} + \sigma^j_m(\bar{X}^{f,[\ell]}_{t_{k+1/2}})\Delta B^m_{t_{k+1},t_{k+1/2}}\Big\}$$
$$+ \sum_{0\le m_1,m_2\le d}\Big\{\mathcal{L}_{m_1}\sigma^j_{m_2}\big(\bar{X}^{f,[\ell]}_{t_k}\big)\Delta\eta^{m_1m_2}_{t_{k+1/2},t_k} + \mathcal{L}_{m_1}\sigma^j_{m_2}\big(\bar{X}^{f,[\ell]}_{t_{k+1/2}}\big)\Delta\eta^{m_1m_2}_{t_{k+1},t_{k+1/2}}\Big\}$$
$$+ \frac{1}{2}\sum_{1\le m_1<m_2\le d}\Big\{[\sigma_{m_1},\sigma_{m_2}]^j\big(\bar{X}^{f,[\ell]}_{t_k}\big)\widetilde{A}^{m_1m_2}_{t_{k+1/2},t_k} + [\sigma_{m_1},\sigma_{m_2}]^j\big(\bar{X}^{f,[\ell]}_{t_{k+1/2}}\big)\widetilde{A}^{m_1m_2}_{t_{k+1},t_{k+1/2}}\Big\}.$$

The Itô-Taylor expansion gives, for $0 \le m \le d$:

$$\text{(B.2)}\qquad \sigma^j_m(\bar{X}^{f,[\ell]}_{t_{k+1/2}}) = \sigma^j_m(\bar{X}^{f,[\ell]}_{t_k}) + \sum_{0\le m_1\le d}\mathcal{L}_{m_1}\sigma^j_m(\bar{X}^{f,[\ell]}_{t_k})\Delta B^{m_1}_{t_{k+1/2},t_k} + \mathcal{E}^{f,j}_{t_{k+1/2},t_k},$$

where under Assumption 2.1 the remainder term $\mathcal{E}^{f,j}_{t_{k+1/2},t_k}$ is such that, for any $p \ge 2$, there exists a constant $C > 0$ so that $\max_{0\le k\le 2^{\ell-1}-1}\mathbb{E}\big[|\mathcal{E}^{f,j}_{t_{k+1/2},t_k}|^p\big] \le C\Delta^p_\ell$. Furthermore, we note that the standard Taylor expansion gives that for any $f \in C^1_b(\mathbb{R}^N)$:

$$\text{(B.3)}\qquad f(\bar{X}^{f,[\ell]}_{t_{k+1/2}}) = f(\bar{X}^{f,[\ell]}_{t_k}) + \sum_{1\le i\le N}\partial_i f(\xi)\big(\bar{X}^{f,[\ell],i}_{t_{k+1/2}} - \bar{X}^{f,[\ell],i}_{t_k}\big)$$

for some variable $\xi \in \mathbb{R}^N$, and it holds that:

$$\Delta B^{m_1}_{t_{k+1},t_k}\Delta B^{m_2}_{t_{k+1},t_k} = \Delta B^{m_1}_{t_{k+1},t_{k+1/2}}\Delta B^{m_2}_{t_{k+1},t_{k+1/2}} + \Delta B^{m_1}_{t_{k+1},t_{k+1/2}}\Delta B^{m_2}_{t_{k+1/2},t_k}$$
$$\text{(B.4)}\qquad + \Delta B^{m_1}_{t_{k+1/2},t_k}\Delta B^{m_2}_{t_{k+1},t_{k+1/2}} + \Delta B^{m_1}_{t_{k+1/2},t_k}\Delta B^{m_2}_{t_{k+1/2},t_k}.$$

Thus, applying (B.2), (B.3) and (B.4) to (B.1), we obtain that:

$$\bar{X}^{f,[\ell],j}_{t_{k+1}} = \bar{X}^{f,[\ell],j}_{t_k} + \sum_{0\le m\le d}\sigma^j_m(\bar{X}^{f,[\ell]}_{t_k})\Delta B^m_{t_{k+1},t_k} + \sum_{0\le m_1,m_2\le d}\mathcal{L}_{m_1}\sigma^j_{m_2}\big(\bar{X}^{f,[\ell]}_{t_k}\big)\Delta\eta^{m_1m_2}_{t_{k+1},t_k}$$
$$- \frac{1}{2}\sum_{1\le m_1,m_2\le d}\mathcal{L}_{m_1}\sigma^j_{m_2}\big(\bar{X}^{f,[\ell]}_{t_k}\big)\big(\Delta B^{m_1}_{t_{k+1},t_{k+1/2}}\Delta B^{m_2}_{t_{k+1/2},t_k} - \Delta B^{m_1}_{t_{k+1/2,k}}\Delta B^{m_2}_{t_{k+1},t_{k+1/2}}\big)$$
$$+ \frac{1}{2}\sum_{1\le m_1<m_2\le d}[\sigma_{m_1},\sigma_{m_2}]^j\big(\bar{X}^{f,[\ell]}_{t_k}\big)\big(\widetilde{A}^{m_1m_2}_{t_{k+1/2},t_k} + \widetilde{A}^{m_1m_2}_{t_{k+1},t_{k+1/2}}\big) + \bar{\mathcal{M}}^{f,j}_{t_{k+1},t_k} + \bar{\mathcal{N}}^{f,j}_{t_{k+1},t_k},$$

where the remainder terms $\bar{\mathcal{M}}^{f,j}_{t_{k+1},t_k}$ and $\bar{\mathcal{N}}^{f,j}_{t_{k+1},t_k}$ have the properties stated in Lemma B.1. The assertion for $\widetilde{X}^{f,[\ell]}$ follows from the same discussion above, and the proof is now complete. ■

Lemma B.2. *It holds that:*

$$\text{(B.5)}\quad \begin{aligned} \hat{X}^{f,[\ell],j}_{t_{k+1}} &= \hat{X}^{f,[\ell],j}_{t_k} + \sum_{0\le m\le d} \sigma^j_m(\hat{X}^{f,[\ell]}_{t_k})\Delta B^m_{t_{k+1},t_k} + \sum_{1\le m_1,m_2\le d} \mathcal{L}_{m_1}\sigma^j_{m_2}(\hat{X}^{f,[\ell]}_{t_k})\Delta\eta^{m_1m_2}_{t_{k+1},t_k} \\ &\quad + \hat{\mathcal{M}}^{f,j}_{t_{k+1},t_k} + \hat{\mathcal{N}}^{f,j}_{t_{k+1},t_k}, \end{aligned}$$

where the remainder terms $\hat{\mathcal{M}}^{f,j}_{t_{k+1},t_k}$ and $\hat{\mathcal{N}}^{f,j}_{t_{k+1},t_k}$ satisfy the same properties as $\bar{\mathcal{M}}^{f,j}_{t_{k+1},t_k}$ and $\bar{\mathcal{N}}^{f,j}_{t_{k+1},t_k}$ in Lemma B.1, respectively.

Proof. For notational simplicity, we omit the subscript "$[\ell]$" during the proof. Due to Lemma B.1, we get:

$$\text{(B.6)}\quad \begin{aligned} \hat{X}^{f,j}_{t_{k+1}} &= \hat{X}^{f,j}_{t_k} + \sum_{0\le m\le d} \sigma^j_m(\hat{X}^{f}_{t_k})\Delta B^m_{t_{k+1},t_k} + \sum_{1\le m_1,m_2\le d} \mathcal{L}_{m_1}\sigma^j_{m_2}(\hat{X}^{f}_{t_k})\Delta\eta^{m_1m_2}_{t_{k+1},t_k} \\ &\quad + \sum_{1\le i\le 6} \mathcal{E}^{(i),j}_{t_{k+1},t_k} + \tfrac{1}{2}\big\{\bar{\mathcal{M}}^{f,j}_{t_{k+1},t_k} + \widetilde{\mathcal{M}}^{f,j}_{t_{k+1},t_k} + \bar{\mathcal{N}}^{f,j}_{t_{k+1},t_k} + \widetilde{\mathcal{N}}^{f,j}_{t_{k+1},t_k}\big\}, \end{aligned}$$

where we have set:

$$\mathcal{E}^{(1),j}_{t_{k+1},t_k} = \sum_{1\le m\le d} \big(\tfrac{1}{2}\sigma^j_m(\bar{X}^f_{t_k}) + \tfrac{1}{2}\sigma^j_m(\widetilde{X}^f_{t_k}) - \sigma^j_m(\hat{X}^f_{t_k})\big)\Delta B^m_{t_{k+1},t_k};$$

$$\mathcal{E}^{(2),j}_{t_{k+1},t_k} = \big(\tfrac{1}{2}\sigma^j_0(\bar{X}^f_{t_k}) + \tfrac{1}{2}\sigma^j_0(\widetilde{X}^f_{t_k}) - \sigma^j_0(\hat{X}^f_{t_k})\big)\Delta_{\ell-1} + \big(\mathcal{L}_0\sigma^j_0(\bar{X}^f_{t_k}) + \mathcal{L}_0\sigma^j_0(\widetilde{X}^f_{t_k})\big)\tfrac{\Delta^2_{\ell-1}}{4};$$

$$\mathcal{E}^{(3),j}_{t_{k+1},t_k} = \sum_{1\le m_1,m_2\le d} \Big(\tfrac{1}{2}\mathcal{L}_{m_1}\sigma^j_{m_2}(\bar{X}^f_{t_k}) + \tfrac{1}{2}\mathcal{L}_{m_1}\sigma^j_{m_2}(\widetilde{X}^f_{t_k}) - \mathcal{L}_{m_1}\sigma^j_{m_2}(\hat{X}^f_{t_k})\Big)\Delta\eta^{m_1m_2}_{t_{k+1},t_k};$$

$$\mathcal{E}^{(4),j}_{t_{k+1},t_k} = \tfrac{1}{2}\sum_{1\le m\le d}\big\{\big(\mathcal{L}_m\sigma^j_0(\bar{X}^f_{t_k}) + \mathcal{L}_m\sigma^j_0(\widetilde{X}^f_{t_k})\big)\Delta\eta^{m0}_{t_{k+1},t_k} + \big(\mathcal{L}_0\sigma^j_m(\bar{X}^f_{t_k}) + \mathcal{L}_0\sigma^j_m(\widetilde{X}^f_{t_k})\big)\Delta\eta^{0m}_{t_{k+1},t_k}\big\};$$

$$\begin{aligned} \mathcal{E}^{(5),j}_{t_{k+1},t_k} &= -\tfrac{1}{4}\sum_{1\le m_1,m_2\le d}\big\{\big(\mathcal{L}_{m_1}\sigma^j_{m_2}(\bar{X}^f_{t_k}) - \mathcal{L}_{m_1}\sigma^j_{m_2}(\widetilde{X}^f_{t_k})\big) \\ &\qquad\qquad \times \big(\Delta B^{m_1}_{t_{k+1},t_{k+1/2}}\Delta B^{m_2}_{t_{k+1/2},t_k} - \Delta B^{m_1}_{t_{k+1/2},t_k}\Delta B^{m_2}_{t_{k+1},t_{k+1/2}}\big)\big\}; \end{aligned}$$

$$\mathcal{E}^{(6),j}_{t_{k+1},t_k} = \tfrac{1}{4}\sum_{1\le m_1<m_2\le d}\big([\sigma_{m_1},\sigma_{m_2}]^j(\bar{X}^f_{t_k}) - [\sigma_{m_1},\sigma_{m_2}]^j(\widetilde{X}^f_{t_k})\big)\big(\Delta\widetilde{A}^{m_1m_2}_{t_{k+1},t_{k+1/2}} + \Delta\widetilde{A}^{m_1m_2}_{t_{k+1/2},t_k}\big).$$

We immediately have that, $\mathbb{E}\big[\mathcal{E}^{(i),j}_{t_{k+1},t_k}|\mathcal{F}_{t_k}\big] = 0$, $i\in\{1,3,4,5,6\}$. Applying second order Taylor expansion around $\hat{X}^f_{t_k}$, we have under Assumption 2.1 that, for $g\in C^2_b(\mathbb{R}^N;\mathbb{R})$ and $p\ge 2$, there exist constants $C_1, C_2 > 0$ such that for all $0\le k\le 2^{\ell-1}-1$:

$$\mathbb{E}\big[\big|\tfrac{1}{2}\big(g(\bar{X}^f_{t_k}) + g(\widetilde{X}^f_{t_k})\big) - g(\hat{X}^f_{t_k})\big|^p\big] \le C_1\Delta^p_{\ell-1}, \quad \mathbb{E}\big[\big|g(\bar{X}^f_{t_k}) - g(\widetilde{X}^f_{t_k})\big|^p\big] \le C_2\Delta^{p/2}_{\ell-1},$$

where we made use of the following result: for any $p \geq 2$, there exists $C > 0$ such that

$$\max_{0\leq k\leq 2^{\ell-1}-1} \mathbb{E}\big[\|\bar{X}^f_{t_k} - \widetilde{X}^f_{t_k}\|^p\big] \leq C\Delta_{\ell-1}^{p/2}. \tag{B.7}$$

The bound (B.7) is obtained by noticing that

$$\max_{0\leq k\leq 2^{\ell-1}-1} \mathbb{E}\big[\|\bar{X}^f_{t_k} - \widetilde{X}^f_{t_k}\|^p\big] \leq \mathbb{E}\big[\max_{0\leq k\leq 2^{\ell-1}-1} \|\bar{X}^f_{t_k} - \widetilde{X}^f_{t_k}\|^p\big]$$

and applying the same argument used in the proof of [11, Lemma 4.6] with the strong convergence result (Proposition 2.3) to the right-hand-side of the above inequality. Then, we have that: $\max_{0\leq k\leq 2^{\ell-1}-1} \mathbb{E}\big[|\mathcal{E}^{(2),j}_{t_{k+1},t_k}|^p\big] \leq C_1\Delta_{\ell-1}^{2p}$, $\max_{0\leq k\leq 2^{\ell-1}-1} \mathbb{E}\big[|\mathcal{E}^{(i),j}_{t_{k+1},t_k}|^p\big] = C_2\Delta_{\ell-1}^{3p/2}$, $j \in \{1,3,4,5,6\}$ for some positive constants C_1, C_2. Setting $\hat{\mathcal{M}}^{f,j}_{t_{k+1},t_k} \equiv \sum_{i\in\{1,3,4,5,6\}} \mathcal{E}^{(i),j}_{t_{k+1},t_k} + \frac{1}{2}\big(\bar{\mathcal{M}}^{f,j}_{t_{k+1},t_k} + \widetilde{\mathcal{M}}^{f,j}_{t_{k+1},t_k}\big)$ and $\hat{\mathcal{N}}^{f,j}_{t_{k+1},t_k} \equiv \mathcal{E}^{(2),j}_{t_{k+1},t_k} + \frac{1}{2}\big(\bar{\mathcal{N}}^{f,j}_{t_{k+1},t_k} + \widetilde{\mathcal{N}}^{f,j}_{t_{k+1},t_k}\big)$, we conclude. ∎

Lemma B.3. *It holds that:*

$$\begin{aligned}\hat{X}^{c,[\ell-1],j}_{t_{k+1}} &= \hat{X}^{c,[\ell-1],j}_{t_k} + \sum_{0\leq m\leq d} \sigma^j_m(\hat{X}^{c,[\ell-1]}_{t_k})\Delta B^m_{t_{k+1},t_k} \\ &\quad + \sum_{1\leq m_1,m_2\leq d} \mathcal{L}_{m_1}\sigma^j_{m_2}(\hat{X}^{c,[\ell-1]}_{t_k})\Delta\eta^{m_1m_2}_{t_{k+1},t_k} + \hat{\mathcal{M}}^{c,j}_{t_{k+1},t_k} + \hat{\mathcal{N}}^{c,j}_{t_{k+1},t_k},\end{aligned} \tag{B.8}$$

where the remainder terms $\hat{\mathcal{M}}^{c,j}_{t_{k+1},t_k}$ and $\hat{\mathcal{N}}^{c,j}_{t_{k+1},t_k}$ satisfy the same properties as $\bar{\mathcal{M}}^{f,j}_{t_{k+1},t_k}$ and $\bar{\mathcal{N}}^{f,j}_{t_{k+1},t_k}$ in Lemma B.1, respectively.

Proof. For notational simplicity, we omit the subscript "$[\ell-1]$" during the proof. From the discretizations (2.4) and (2.5), we have:

$$\hat{X}^{c,j}_{t_{k+1}} = \hat{X}^{c,j}_{t_k} + \sum_{0\leq m\leq d} \sigma^j_m(\hat{X}^c_{t_k})\Delta B^m_{t_{k+1},t_k} + \sum_{1\leq m_1,m_2\leq d} \mathcal{L}_{m_1}\sigma^j_{m_2}(\hat{X}^c_{t_k})\Delta\eta^{m_1m_2}_{t_{k+1},t_k} + \sum_{1\leq i\leq 5} \mathcal{R}^{(i),j}_{t_{k+1},t_k},$$

where we have defined:

$$\mathcal{R}^{(1),j}_{t_{k+1},t_k} = \big(\tfrac{1}{2}\sigma^j_0(\bar{X}^c_{t_k}) + \tfrac{1}{2}\sigma^j_0(\widetilde{X}^c_{t_k}) - \sigma^j_0(\hat{X}^c_{t_k})\big)\Delta_{\ell-1} + \big(\mathcal{L}_0\sigma^j_0(\bar{X}^c_{t_k}) + \mathcal{L}_0\sigma^j_0(\widetilde{X}^c_{t_k})\big)\frac{\Delta^2_{\ell-1}}{4};$$

$$\mathcal{R}^{(2),j}_{t_{k+1},t_k} = \sum_{1\leq m\leq d} \big(\tfrac{1}{2}\sigma^j_m(\bar{X}^c_{t_k}) + \tfrac{1}{2}\sigma^j_m(\widetilde{X}^c_{t_k}) - \sigma^j_m(\hat{X}^c_{t_k})\big)\Delta B^m_{t_{k+1},t_k};$$

$$\mathcal{R}^{(3),j}_{t_{k+1},t_k} = \sum_{1\leq m_1,m_2\leq d} \big(\tfrac{1}{2}\mathcal{L}_{m_1}\sigma^j_{m_2}(\bar{X}^c_{t_k}) + \tfrac{1}{2}\mathcal{L}_{m_1}\sigma^j_{m_2}(\widetilde{X}^c_{t_k}) - \mathcal{L}_{m_1}\sigma^j_{m_2}(\hat{X}^c_{t_k})\big)\Delta\eta^{m_1m_2}_{t_{k+1},t_k};$$

$$\mathcal{R}^{(4),j}_{t_{k+1},t_k} = \frac{1}{4}\sum_{1\leq m_1<m_2\leq d} \big([\sigma_{m_1},\sigma_{m_2}]^j(\bar{X}^c_{t_k}) - [\sigma_{m_1},\sigma_{m_2}]^j(\widetilde{X}^c_{t_k})\big)\Delta B^{m_1}_{t_{k+1},t_k}\Delta\widetilde{B}^{m_2}_{t_{k+1},t_k};$$

$$\mathcal{R}^{(5),j}_{t_{k+1},t_k} = \frac{1}{2}\sum_{1\leq m\leq d}\big\{(\mathcal{L}_m\sigma^j_0(\bar{X}^c_{t_k}) + \mathcal{L}_m\sigma^j_0(\widetilde{X}^c_{t_k}))\Delta\eta^{m0}_{t_{k+1},t_k} + (\mathcal{L}_0\sigma^j_m(\bar{X}^c_{t_k}) + \mathcal{L}_0\sigma^j_m(\widetilde{X}^c_{t_k}))\Delta\eta^{0m}_{t_{k+1},t_k}\big\}.$$

From the argument used in proof of Lemma B.2, we have that $\hat{\mathcal{M}}^{c,j}_{t_{k+1},t_k} = \sum_{2\leq i\leq 5} \mathcal{R}^{(i),j}_{t_{k+1},t_k}$ and $\hat{\mathcal{N}}^{c,j}_{t_{k+1},t_k} = \mathcal{R}^{(1),j}_{t_{k+1},t_k}$ satisfy the properties in the statement of Lemma B.3, and then we conclude. ∎

REFERENCES

[1] Al Gerbi, A., Jourdain, B., & Clément, E. (2016). Ninomiya–Victoir scheme: Strong convergence, antithetic version and application to multilevel estimators. *Monte Carlo Methods Appl.*, **22**, 197-228.

[2] Bain, A., & Crisan, D. (2009). *Fundamentals of Stochastic Filtering*. Springer: New York.

[3] Ballesio, M., Jasra, A., von Schwerin, E. & Tempone,R. (2023). A Wasserstein coupled particle filter for multilevel estimation. *Stoch. Anal. Appl.*, **41**, 820–859.

[4] Cappé, O., Ryden, T, & Moulines, É. (2005). *Inference in Hidden Markov Models*. Springer: New York.

[5] Chopin, N. and Papaspiliopoulos, O. (2020). *An introduction to sequential Monte Carlo*. Springer Cham.

[6] Del Moral, P. (2004). *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer: New York.

[7] Ditlevsen, S. & Samson, A. (2019). Hypoelliptic diffusions: filtering and inference from complete and partial observations. *J. R. Stat. Soc., B: Stat. Methodol.*, **81**, 361–384.

[8] Ermentrout, B. & Terman, D. (2010). *Mathematical foundations of neuroscience*. Springer: New York.

[9] Giles, M. B. (2008). Multilevel Monte Carlo path simulation. *Op. Res.*, **56**, 607-617.

[10] Giles, M. B. (2015) Multilevel Monte Carlo methods. *Acta Numerica*, **24**, 259-328.

[11] Giles, M. B., & Szpruch, L. (2014). Antithetic multilevel Monte Carlo estimation for multidimensional SDEs without Lévy area simulation. *Ann. Appl. Probab.*, **24**, 1585-1620.

[12] Glasserman, P. (2003). *Monte Carlo Methods in Financial Engineering*. Springer: New York.

[13] Gloter, A., & Yoshida, N. (2021). Adaptive estimation for degenerate diffusion processes. *Electron. J. Stat.*, **15**, 1424–1472.

[14] Heinrich, S. (2001). Multilevel Monte Carlo methods. In *Large-Scale Scientific Computing*, (eds. S. Margenov, J. Wasniewski & P. Yalamov), Springer: Berlin.

[15] Heston, S.L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Finan. Stud.*, **6**, 327-343.

[16] Iguchi, Y., Beskos, A., & Graham, M. (2025). Parameter estimation with increased precision for elliptic and hypo-elliptic diffusions. *Bernoulli*. **31**, 333–358.

[17] Iguchi, Y. & Yamada, T. (2021). A second-order discretization for degenerate systems of stochastic differential equations. *IMA J. Numer. Anal.*, **41**, 2782–2829.

[18] Jasra, A., Kamatani, K., Law K. J. H. & Zhou, Y. (2017). Multilevel particle filters. *SIAM J. Numer. Anal.*, **55**, 3068-3096.

[19] Jasra, A., Kamatani, K., Osei, P. P., & Zhou, Y. (2018). Multilevel particle filters: Normalizing Constant Estimation. *Statist. Comp.*, **28**, 47-60.

[20] Jasra, A., Law K. J. H. & Osei, P. P. (2019). Multilevel particle filters for Lévy driven stochastic differential equations. *Statist. Comp.*, **29**, 775-789.

[21] Jasra, A., & Yu, F. (2020). Central limit theorems for coupled particle filters. *Adv. Appl. Probab.*, **52**, 942-1001.

[22] Jasra, A., Law K. J. H. & Suciu, C. (2020). Advanced Multilevel Monte Carlo. *Intl. Stat. Rev.*, **88**, 548-579.

[23] Jasra, A., Law, K. J. H. & Yu, F. (2022). Unbiased filtering of a class of partially observed diffusions. *Adv. Appl. Probab.*, **54**, 661-687.

[24] Jasra, A., Maama, M. & Ombao, H. (2024). Antithetic Multilevel Particle Filters. *Adv. Appl. Probab.* (to appear).

[25] Kloeden, P. & Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations*. Springer: New York.

[26] Milstein, G. N. & Tretyakov, M. V. (2021). *Stochastic Numerics for Mathematical Physics*. Springer: Cham.

[27] Ninomiya, S. & Victoir, N. (2008). Weak approximation of stochastic differential equations and application to derivative pricing. *Appl. Math. Finance*, **15**, 107-121.

[28] Nualart, D. *The Malliavin Calculus and Related Topics*. Springer Berlin, Heidelberg.

[29] Rumelin, W. (1982). Numerical treatment of stochastic differential equations. *SIAM J. Numer. Anal.*, **19**, 604-613.

Algorithm 3.2 New Coupled Particle Filter using the antithetic weak second order scheme. The algorithm is stopped at a time T, but need not be.

1. Input: level of discretization $\ell \in \mathbb{N}$, final time $T \in \mathbb{N}$ and number of samples M. Set $\bar{X}_0^{c,[\ell-1]}(i) = \widetilde{X}_0^{c,[\ell-1]}(i) = \bar{X}_0^{f,[\ell]}(i) = \widetilde{X}_0^{f,[\ell]}(i) = x$, $i = 1,\dots,M$ and $k = 1$. Go to 2..
2. Sampling: For $i = 1,\dots,M$, simulate

$$\big(\bar{X}_k^{c,[\ell-1]}(i), \widetilde{X}_k^{c,[\ell-1]}(i), \bar{X}_k^{f,[\ell]}(i), \widetilde{X}_k^{f,[\ell]}(i)\big)|\big(\bar{x}_{k-1}^{c,[\ell-1]}(i), \widetilde{x}_{k-1}^{c,[\ell-1]}(i), \bar{x}_{k-1}^{f,[\ell]}(i), \widetilde{x}_{k-1}^{f,[\ell]}(i)\big)$$

   using the coupled dynamics (2.4)-(2.7) up-to time 1, with:
   - starting point $\bar{x}_{k-1}^{c,[\ell-1]}(i)$, step-size $\Delta_{\ell-1}$ for (2.4)
   - starting point $\widetilde{x}_{k-1}^{c,[\ell-1]}(i)$, step-size $\Delta_{\ell-1}$ for (2.5)
   - starting point $\bar{x}_{k-1}^{f,[\ell]}(i)$ and step-size Δ_ℓ for (2.6)
   - starting point $\widetilde{x}_{k-1}^{f,[\ell]}(i)$ and step-size Δ_ℓ for (2.7).

   Go to 3..
3. Resampling: For $i = 1,\dots,M$ compute

$$w_k^{c,[\ell-1]}(i) := \frac{g(\bar{X}_k^{c,[\ell-1]}(i),y_k)}{\sum_{j=1}^M g(\bar{X}_k^{c,[\ell-1]}(j),y_k)}, \quad w_k^{f,[\ell]}(i) := \frac{g(\bar{X}_k^{f,[\ell]}(i),y_k)}{\sum_{j=1}^M g(\bar{X}_k^{f,[\ell]}(j),y_k)};$$
$$\widetilde{w}_k^{c,[\ell-1]}(i) := \frac{g(\widetilde{X}_k^{c,[\ell-1]}(i),y_k)}{\sum_{j=1}^M g(\widetilde{X}_k^{c,[\ell-1]}(j),y_k)}, \quad \widetilde{w}_k^{f,[\ell]}(i) := \frac{g(\widetilde{X}_k^{f,[\ell]}(i),y_k)}{\sum_{j=1}^M g(\widetilde{X}_k^{f,[\ell]}(j),y_k)}.$$

   For any $\varphi \in \mathcal{B}_b(\mathbb{R}^N)$ we have the estimate:

$$\hat{\pi}_k^{[\ell],M}(\varphi) - \hat{\pi}_k^{[\ell-1],M}(\varphi) := \tfrac{1}{2}\sum_{1\le i\le M}\Big\{w_k^{f,[\ell]}(i)\varphi(\bar{X}_k^{f,[\ell]}(i)) + \widetilde{w}_k^{f,[\ell]}(i)\varphi(\widetilde{X}_k^{f,[\ell]}(i))\Big\}$$
$$- \tfrac{1}{2}\sum_{1\le i\le M}\Big\{w_k^{c,[\ell-1]}(i)\varphi(\bar{X}_k^{c,[\ell-1]}(i)) + \widetilde{w}_k^{c,[\ell-1]}(i)\varphi(\widetilde{X}_k^{c,[\ell-1]}(i))\Big\}. \tag{3.6}$$

   For $i = 1,\dots,M$ sample indices $\Big(j^{c,[\ell-1]}(i), \widetilde{j}^{c,[\ell-1]}(i), j^{f,[\ell]}(i), \widetilde{j}^{f,[\ell]}(i)\Big) \in \{1,\dots,M\}^4$ using Algorithm **??** in Supplementary Material with probability mass functions $(w_k^{c,[\ell-1]}(\cdot), \widetilde{w}_k^{c,[\ell-1]}(\cdot), w_k^{f,[\ell]}(\cdot), \widetilde{w}_k^{f,[\ell]}(\cdot))$, cardinality M and set

$$\check{X}_k^{c,[\ell-1]}(i) = \bar{X}_k^{c,[\ell-1]}(j^{c,[\ell-1]}(i)), \quad \acute{X}_k^{c,[\ell-1]}(i) = \widetilde{X}_k^{c,[\ell-1]}(\widetilde{j}^{c,[\ell-1]}(i));$$
$$\check{X}_k^{f,[\ell]}(i) = \bar{X}_k^{f,[\ell]}(j^{f,[\ell]}(i)), \quad \acute{X}_k^{f,[\ell]}(i) = \widetilde{X}_k^{f,[\ell]}(\widetilde{j}^{f,[\ell]}(i)).$$

   For $i = 1,\dots,M$, set $\bar{X}_k^{c,[\ell-1]}(i) = \check{X}_k^{c,[\ell-1]}(i)$, $\widetilde{X}_k^{c,[\ell-1]}(i) = \acute{X}_k^{c,[\ell-1]}(i)$, $\bar{X}_k^{f,[\ell]}(i) = \check{X}_k^{f,[\ell]}(i)$, $\widetilde{X}_k^{f,[\ell]}(i) = \acute{X}_k^{f,[\ell]}(i)$. Set $k = k+1$, if $k = T+1$ go to 4. otherwise go to 2..
4. Return the estimates $\hat{\pi}_1^{[\ell],M}(\varphi) - \hat{\pi}_1^{[\ell-1],M}(\varphi), \dots, \hat{\pi}_T^{[\ell],M}(\varphi) - \hat{\pi}_T^{[\ell-1],M}(\varphi)$ from (3.6).