Iterative Particle Approximation for McKean-Vlasov SDEs with application to Multilevel Monte Carlo estimation

Lukasz Szpruch\textsuperscript{1,2}, Shuren Tan\textsuperscript{1}, and Alvin Tse\textsuperscript{1}

\textsuperscript{1}School of Mathematics, University of Edinburgh
\textsuperscript{2}The Alan Turing Institute, London

Abstract

The mean field limits of systems of interacting diffusions (also called stochastic interacting particle systems (SIPS)) have been intensively studied since McKean [McKean Jr, 1966]. The interacting diffusions pave a way to probabilistic representations for many important nonlinear/nonlocal PDEs, but provide a great challenge for Monte Carlo simulations. This is due to the nonlinear dependence of the bias on the statistical error arising through the approximation of the law of the process. This and the fact that particles/diffusions are not independent render classical variance reduction techniques not directly applicable and consequently make simulations of interacting diffusions prohibitive.

In this article, we provide an alternative iterative particle representation, inspired by the fixed point argument by Sznitman [Sznitman, 1991]. This new representation has the same mean field limit as the classical SIPS. However, unlike classical SIPS, it also allows decomposing the statistical error and the approximation bias. We develop a general framework to study integrability and regularity properties of the iterated particle system. Moreover, we establish its weak convergence to the McKean-Vlasov SDEs (MVSDEs). One of the immediate advantages of iterative particle system is that it can be combined with the Multilevel Monte Carlo (MLMC) approach for the simulation of MVSDEs. We proved that the MLMC approach reduces the computational complexity of calculating expectations by an order of magnitude. Another perspective on this work is that we analyse the error of nested Multilevel Monte Carlo estimators, which is of independent interest. Furthermore, we work with state dependent functionals, unlike scalar outputs which are common in literature on MLMC. The error analysis is carried out in uniform, and what seems to be new, weighted norms.

2010 AMS subject classifications: Primary: 65C30, 60H35; secondary: 60H30.

Keywords: McKean-Vlasov SDEs, Stochastic Interacting Particle Systems, Non-linear Fokker-Planck equations, Probabilistic Numerical Analysis
1 Introduction

The theory of mean field interacting particle systems was pioneered by the work of H. McKean [McKean Jr, 1966], where he gave a probabilistic interpretation of a class of nonlinear (due to the dependence on the coefficients of the solution itself) nonlocal PDEs. We consider functions $b : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times r}$, and denote $a(x, y) = \sigma(x, y)\sigma(x, y)^T$. Let $G \in C^2_b(\mathbb{R}^d)$. The nonlinear forward Kolmogorov/Fokker-Planck equation that we consider in this paper is given by

$$\frac{\partial}{\partial t} \langle \mu_t, G \rangle = \left\langle \mu_t, \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2 G}{\partial x_i \partial x_j}(x) \int_{\mathbb{R}^d} a_{ij}(x, y) \mu_t(dy) + \sum_{i=1}^d \frac{\partial G}{\partial x_i}(x) \int_{\mathbb{R}^d} b_i(x, y) \mu_t(dy) \right\rangle,$$

(1.1)

where $\langle \mu_t, G \rangle = \int_{\mathbb{R}^d} G(x) \mu_t(dx)$. We are interested in a probabilistic representation of (1.1). The idea is to interpret the solution of (1.1) as the equation for the marginal law of a certain stochastic process. This allows us to study existence, uniqueness and regularity of (1.1) using probabilistic tools (see [Antonelli et al., 2002, Chassagneux et al., Sznitman, 1991]). Furthermore, probabilistic representations lead to Monte-Carlo methods for approximation and simulation of solutions. They are particularly efficient in high dimensions. Let $\{W_t\}_{t \geq 0}$ be an $r$-dimensional Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The stochastic process that corresponds to the time marginals described in (1.1)
is a non-linear in the sense of McKean stochastic differential equation, called McKean-Vlasov SDEs (MVSDEs), given by
\[
\begin{aligned}
dX_t &= \left( \int_{\mathbb{R}^d} b(X_t, y) \mu^X_t(dy) \right) dt + \left( \int_{\mathbb{R}^d} \sigma(X_t, y) \mu^X_t(dy) \right) dW_t, \\
\mu^X_t &= \mathbb{P} \circ X_t^{-1} = \text{Law}(X_t), \quad t \in [0, T].
\end{aligned}
\] (1.2)

Notice that \((X_t)\) is not a Markov process. The theory of propagation of chaos, [Sznitman, 1991], states that (1.2) arises as a limiting equation (under appropriate conditions) of the system of stochastic interacting particles/diffusions \(X_{i,N}^t\), each of which is the solution to the \((\mathbb{R}^d)^N\)-dimensional SDE
\[
\begin{aligned}
dX_{i,N}^t &= \left( \int_{\mathbb{R}^d} b(X_t, y) \mu_{i,N}^X(dy) \right) dt + \left( \int_{\mathbb{R}^d} \sigma(X_t, y) \mu_{i,N}^X(dy) \right) dW^i_t, \quad i = 1, \ldots, N, \\
\mu_{i,N}^X_t &= \frac{1}{N} \sum_{j=1}^N \delta_{X_{i,N}^t}, \quad t \geq 0.
\end{aligned}
\] (1.3)

where \(\{X_{0,N}^i\}_{i=1,\ldots,N}\) are i.i.d samples with law \(\mu_0\) and \(\{W^i_t\}_{i=1,\ldots,N}\) are independent Brownian motions. It can be shown, under sufficient regularity conditions on the coefficients, that the convergence of the empirical measures (of \(\{X_{i,N}^t\}_t\)) on the path space holds in law, i.e. \(\mu_{X,N}^\cdot = \{\mu_{X,N}^t: t \in [0, T]\} \rightarrow \mu^X, \quad N \rightarrow \infty\) (see [Méléard, 1996]). This is a not trivial result as the particles are not independent and the standard law of large numbers does not apply. Moreover, (1.3) can be interpreted as a first step towards numerical schemes for (1.2). We remark that, the analysis of stochastic particles systems is of independent interest, as it is used as models in molecular dynamics; physical particles in fluid dynamics [Pope, 2000]; behaviour of interacting agents in economics or social networks [Carmona et al., 2013] or interacting neurons in biology [Delarue et al., 2015].

### 1.1 Numerical analysis of interacting diffusions

To obtain a fully implementable approximation of (1.2), one needs to both, approximate probability measure \(\mu^X \in \mathcal{P}(C([0, T], \mathbb{R}^d))\) and discretise the dynamics of the time interval \([0, T]\). The study of probabilistic numerical schemes for (1.1) (or (1.2)) has been initiated in a series of articles by Bossy and Talay [Bossy et al., 1996, Bossy and Talay, 1997]. Relying on the propagation of chaos, the authors consider the corresponding Euler scheme to (1.3) with time-step \(h = T/M\) as follows,
\[
Y_{k+1}^{i,N} = Y_k^{i,N} + \frac{1}{N} \sum_{j=1}^N b(Y_k^{i,N}, Y_k^{j,N}) h + \frac{1}{N} \sum_{j=1}^N \sigma(Y_k^{i,N}, Y_k^{j,N}) \Delta W_{k+1}^i, \quad i = 1, \ldots, N.
\] (1.4)

Note that due to interactions between discretised diffusions/particles, implementation of (1.4) requires \(N^2\) arithmetic operations at each step of the scheme, which makes simulations of (1.4) very costly. This should not come as a surprise as the aim is to approximate non linear/non local PDEs (1.1) for which the deterministic schemes based on space
discretisation, typically, are also computationally very demanding [Bossy et al., 1997]. In [Bossy and Talay, 1997], it has been proven that the empirical distribution function of $N$ particles (1.4) converges, in a weak sense, to the distribution of the corresponding McKean-Vlasov limiting equation with the rate $O((\sqrt{N})^{-1} + \sqrt{h})$. The authors remarked from numerical tests (using the Burger’s equation) that while $(\sqrt{N})^{-1}$ seems optimal, one observes the rate $h$ for time discretisation. This is in agreement with the weak rate for time-discretization of linear (in a sense of McKean) SDEs with the Euler scheme. Assuming additional smoothness, [Antonelli et al., 2002] proved that indeed the rate is $O((\sqrt{N})^{-1} + h)$. Similar results for the one dimensional case (but assuming less regularity on the coefficients) have been proven in [Bossy, 2004, Bossy et al., 2002]. Although in this article we work with an iterated particle system (see precise definition below), we also use the Euler scheme and recover the weak rate $h$ for the general multidimensional case following the techniques from [Bossy, 2004]. However, the aim is to obtain better (compared to the works mentioned above) overall convergence rate (depending on the number of discretisation points and the number of particles) using the Multilevel Monte Carlo approach of Giles and Heinrich [Giles et al., 2015, Heinrich, 2001, Kebaier et al., 2005].

We are not aware of any rigorous work on variance reduction techniques for the approximations that is based on propagation of chaos. One of the challenges is the nonlinear relation between the statistical error and the approximation bias. This renders the application of variance reduction, and, in particular, Multilevel Monte Carlo method to MVSDEs non-trivial. To shed more light into this problem, we consider the Euler scheme in the space variable (but not on the measure) of (1.5) with a time step of $h > 0$, $k = 1, \ldots, M$,

$$\bar{X}_{k+1} = \bar{X}_k + b(\bar{X}_k, \mathbb{P}_{kh})h + \sigma(\bar{X}_k, \mathbb{P}_{kh})\Delta W_{k+1}, \quad \mathbb{P}_{kh} = \mathbb{P} \circ (\bar{X}_k)^{-1} = Law(\bar{X}_k).$$

(1.5)

The above scheme is not directly implementable (one cannot sample directly from (1.5) without introducing bias), but it provides a useful initial step for the error analysis. For any function $P \in C^2_b(\mathbb{R}^d, \mathbb{R})$, we consider mean-square error

$$MSE(P) := \mathbb{E} \left[ \left( \mathbb{E}[P(X_T)] - \frac{1}{N} \sum_{i=1}^{N} P(Y_{T,i}^N) \right)^2 \right].$$

By (1.5) and (1.4), we can decompose

$$\mathbb{E}[P(X_T)] - \frac{1}{N} \sum_{i=1}^{N} P(Y_{T,i}^N) = \left( \mathbb{E}[P(X_T)] - \mathbb{E}[P(\bar{X}_T)] \right) + \left( \mathbb{E}[P(\bar{X}_T)] - \frac{1}{N} \sum_{i=1}^{N} P(\bar{X}_{T,i}^N) \right) + \frac{1}{N} \sum_{i=1}^{N} \left( P(\bar{X}_{T,i}^N) - P(Y_{T,i}^N) \right).$$

The first term is the discretization error; the second is the statistical error; and the third stems from the propagation of chaos. Note that the 3rd term is not present in the simulations of standard SDEs and can be upper bounded by $O(1/\sqrt{N})$ following the proof of
Theorem 1.4 from [Sznitman, 1991]. The third term is really the crux of the matter. The superior performance of the MLMC for linear SDEs (in the sense of McKean) is due to the linear dependence of the statistical error and bias. More precisely, MLMC reduces the computational cost of simulation by carefully combining many simulations on grids with low accuracy (at a corresponding low cost); with relatively few simulations computed with high accuracy (and at a high cost) on very fine grids. There are (at least) two issues pertaining to the direct application of MLMC methodology to (1.4): i) the telescopic property needed for MLMC identity [Giles, 2008] does not hold in general; ii) a small number of simulations (particles) on fine time steps (a reason for the improved computational cost in MLMC setting) would lead to a poor approximation of the measure, leading to a high bias.

To show that telescopic sum does not hold in general, consider a collection of discretisations of $[0, T]$ with different resolutions. To this end, we fix $L \in \mathbb{N}$ and introduce a family of time grids, $\Pi^\ell = \{0 = t_{\ell}^0, \ldots, t_{\ell}^k, \ldots T = t_{\ell}^L\}$, $\ell = 0, \ldots, L$, where $t_{\ell}^k - t_{\ell}^{k-1} = h_{\ell} = T2^{-\ell}$. Then $Y_t^{i,\ell,N_{\ell}}$, $\ell = 1, \ldots, L$, denotes for each $i$ a particle corresponding to (1.4) with timestep $h_{\ell}$, where $N_{\ell}$ is the total number of particles. Let $P : \mathbb{R}^d \to \mathbb{R}$ be any Borel-measurable function. One cannot directly apply MLMC to $Y_t^{i,\ell,N_{\ell}}$, since, in general,

$$
\mathbb{E}\left[P(Y_t^{1,\ell,N_{\ell}})\right] \neq \mathbb{E}\left[P(Y_t^{1,\ell,N_{\ell+1}})\right],
$$

and consequently

$$
\mathbb{E}\left[P(Y_t^{1,L,N_L})\right] \neq \mathbb{E}\left[P(Y_t^{1,0,N_0})\right] + \sum_{\ell=0}^{L} \mathbb{E}\left[P(Y_t^{1,\ell,N_{\ell}}) - P(Y_t^{1,\ell-1,N_{\ell}})\right].
$$

On the contrary, if we required the number of particles for all the levels to be the same, then the telescopic sum would hold, but clearly, there would be no computational gain from doing MLMC. We are aware of two articles that tackle the aforementioned issue. The case of linear coefficients is treated in [Ricketson, 2015], in which particles from all levels are used to approximate the mean field at the final (most accurate) approximation level. It is not clear how this approach could be extended to general McKean-Vlasov equations. A numerical study of a "multi-cloud" approach is presented in [Haji-Ali and Tempone, 2016]. The algorithm resembles the MLMC approach to the nested simulation problem in [Giles et al., 2015, Bujok et al., 2013, Lemaire et al., 2017]. Their approach is very natural but because particles in each cloud are not independent one faces similar challenges as with classical particle system. In this article, we develop a generic approach that allows decomposing statistical error and bias for the simulation of interacting diffusions. We also provide error analysis for a general class of MVSDEs. It is worth pointing out that the idea of combining iteration method and MLMC to solve non-linear PDEs has very recently been proposed in [Hutzenthaler et al., 2016]. However, their interest is on BSDEs and their connections to semi-linear PDEs.
1.2 Iterated particle (MLMC) method

The main idea is to approximate (1.2) with a sequence of linear (in a sense of McKean) SDEs defined as

\[ dX_t^m = \int_{\mathbb{R}^d} b(x_t^m, y)\mu_t^{X_{t-1}^m}(dy)dt + \int_{\mathbb{R}^d} \sigma(x_t^m, y)\mu_t^{X_{t-1}^m}(dy)dW_t^m, \quad \mu^0_0 = \mu^X_0, \]  

where \( W^m \) and \( X_0^m \) are independent. This independence is the crux of our approach and is exactly what differs from the proof of existence of solutions by Sznitman [Sznitman, 1991], where the same Brownian motion and initial condition are used at every iteration. Furthermore, we can define \( \eta(t) := t^\ell \) if \( t \in [t^\ell, t^{\ell+1}) \) (we simply write \( \eta(t) = \eta(t) \) when it does not lead to ambiguity). With the above notation in hand, a continuous-time extension of the Euler scheme with time discretisation \( h_\ell \) is defined as follows,

\[ dX_t^{m,\ell} = \int_{\mathbb{R}^d} b(x_{\eta(t)}^m, y)\mu_{\eta(t)}^{X_{t-1}^{m-1}}(dy)dt + \int_{\mathbb{R}^d} \sigma(x_{\eta(t)}^m, y)\mu_{\eta(t)}^{X_{t-1}^{m-1}}(dy)dW_t^m, \quad \mu^0_0 = \mu^X_0. \]  

In order to be able to implement the above scheme at every time step of the Euler scheme (at the iteration step \( m \)), one needs to compute the integral with respect to the measure from the previous iteration \( m - 1 \). This integral is calculated by approximating measure \( \mu_{\eta(t)}^{X_{t-1}^{m-1}} \) by the empirical measure \( \mu_{\eta(t)}^{m-1, N_{m-1}} := \frac{1}{N_{m-1}} \sum_{i=1}^{N_{m-1}} \delta_{Y_{\eta(t)}^{i, m-1, \ell}} \). We define, for \( 1 \leq \ell \leq N_m \),

\[ \begin{aligned} dY_t^{i,m,\ell} &= \int_{\mathbb{R}^d} b(Y_t^{i,m,\ell}, y)\mu_{\eta(t)}^{m-1, N_{m-1}}(dy)dt + \int_{\mathbb{R}^d} \sigma(Y_t^{i,m,\ell}, y)\mu_{\eta(t)}^{m-1, N_{m-1}}(dy)dW_t^m, \\
\mu^0_0 &= \mu^X_0, \end{aligned} \]  

and call it an iterative particle system. By this construction, the particles \( (Y_t^{i,m,\ell})_{i,\ell} \) are independent upon conditioning on \( \mathcal{F}^{m-1} := \sigma\left\{ Y_t^{i,m-1,\ell} \mid 1 \leq i \leq N_{m-1,\ell} : t \in [0, T] \right\} \). Surprisingly, perhaps, we will show that the computational cost of simulating (1.8) is the same as for standard particle system (1.2) (at least asymptotically). Nonetheless, we will improve the computational performance of this method by approximating the aforementioned integrals using an MLMC approach. Having defined processes (1.8), we want to construct MLMC estimators at every time point of the finest discretisation \( \Pi^L \). However, since for \( \ell < L \), \( Y_t^{i,m,\ell} \) is not defined at every timepoint in \( \Pi^L \), we introduce a linear-interpolated measure (in time) \( \tilde{\mu}^{Y^{m,\ell}}_t \), which agrees with the empirical measure \( \frac{1}{N} \sum_{i=1}^{N} \delta_{Y_t^{i,m,\ell}} \), for \( t \in \Pi^L \) (see (2.8) for precise definition). This is in line with the original development of MLMC by Heinrich [Heinrich, 2001]. We define the MLMC particle system by

\[ dY_t^{i,m,\ell} = \mathcal{M}^{(m-1)}_{\eta(t)}(b(Y_t^{i,m,\ell}, \cdot)) dt + \mathcal{M}^{(m-1)}_{\eta(t)}(\sigma(Y_t^{i,m,\ell}, \cdot)) dW_t^{i,m}, \]
for $1 \leq i \leq N_{m,\ell}$, $0 \leq \ell \leq L$, where the MLMC operator $\mathcal{M}_t^{(m)}$ is defined by

$$
\mathcal{M}_t^{(m)}(G(x, \cdot)) = \sum_{\ell=0}^{L} \left( \hat{\mu}_t^{Y_{m,\ell},N_{m,\ell}} - \tilde{\mu}_t^{Y_{m,\ell-1},N_{m,\ell}} \right)(G(x, \cdot)), \quad \tilde{\mu}_t^{Y_{m,0},N_{m,0}} := 0, \quad x \in \mathbb{R}^d.
$$

Moreover, we require the initial conditions $Y_{0,m,\ell}$ to be i.i.d., for $1 \leq i \leq N_{m,\ell}$, $0 \leq \ell \leq L$. We set $Y_{i,0,\ell} = X_0$ for $1 \leq i \leq N_{0,\ell}$, $0 \leq \ell \leq L$. We interpret the MLMC operator in a componentwise sense. The fact that MLMC operator acts on the functional that depends on spatial variable is another hurdle this work overcomes. To allow for flexibility, we work with uniform norms, but also introduce weighted norms. While the study of MLMC in uniform norms has already appeared in literature [Heinrich, 2001, Giles et al., 2015], we are not aware of the use of weighted norms. We emphasize that, as in original MLMC developments, the coupling at each level of the telescopic sum is achieved by using the underlying Brownian path to simulate two particle systems with different time steps [Giles, 2008].

The complete algorithm of MLMC particle approximation is presented in a schematic form below, where for brevity, the index corresponding to the particle is suppressed. Note that there are two equivalent implementations of the algorithm. The first one for each $0 \leq m \leq M$ performs "standard" MLMC on $[0, T]$. The second one for each $k \in \{0, \ldots, 2^l - 1\}$, $\ell \in \{1, \ldots, L\}$ performs $M$ iterations. We present the first one at the end of this section.

For classical particle systems, the computational cost of achieving a mean-square-error of order $\epsilon^2 > 0$, i.e. $MSE = \mathcal{O}(h^2 + N^{-1}) \leq \epsilon^2$, is $\mathcal{O}(\epsilon^{-5})$ (assuming cost of $N^2$ at the every step of the Euler approximation).

**A remark on notations.** We use $\|A\|$ to denote the Hilbert-Schmidt norm while $|v|$ is used to denote the Euclidean norm. We adopt the standard shorthand notation in measure theory and define $\mu(f) = \int_X f \, d\mu$, for any function $f$ defined on a measurable space $(X, \mathcal{F})$. For any Polish space $E$, we use $\mathcal{R}(E)$ to denote the set of random measures on $E$ (with respect to some given probability space $(\Omega, \mathcal{F}, \mathbb{P})$). Moreover, for any Polish spaces $E$ and $F$, we use the notation $C(E, F)$ to denote the set of continuous functions from $E$ to $F$. For any stochastic process $R = \{R_t\}_{t \in \mathbb{I}}$, the law of $R_t$ at any time point $t \in \mathbb{I}$ is denoted by $\mu_t^R$.

For any function $f : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}$, if it is twice partially differentiable with respect to all variables in the first argument and partially differentiable with respect to all variables in the second argument, with all its partial derivatives continuous and bounded, then we indicate this condition by the notation $f \in C_b^{2,1}(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R})$. By interpreting this definition componentwise, it is straightforward to extend this notation and to define $C_b^{2,1}(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R}^k)$ and $C_b^{2,1}(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R}^{k\otimes l})$. We use similar notations $C_b^2(\mathbb{R}^m, \mathbb{R}^k)$ and $C_b^2(\mathbb{R}^m, \mathbb{R}^{k\otimes l})$ whenever the second argument vanishes.

Finally, we denote by $C_b^2(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R})$ the set of functions $G$ from $\mathbb{R}^m \times \mathbb{R}^n$ to $\mathbb{R}$ that are twice-differentiable in the second argument, for which there exists a constant $L$ such that for each $x \in \mathbb{R}^m$, $y \in \mathbb{R}^n$, $i, j \in \{1, \ldots, n\}$,

$$
|D_y G(x, y)| \leq L(1 + |y|^p), \quad |D_{y_i, y_j} G(x, y)| \leq L(1 + |y|^p),
$$

for $1 \leq i \leq N_m$, $0 \leq \ell \leq L$, where the MLMC operator $\mathcal{M}_t^{(m)}$ is defined by
where $D_y$ and $D_{y_i}y_j$ denote respectively the first and second order partial derivatives w.r.t. the second argument. Similarly, $C^2_b(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R})$ denotes the set of all twice-differentiable functions in the second argument such that all their first and second order partial derivatives in the second argument are bounded. As before, we interpret this definition in a componentwise sense to define $C^2_b(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R}^k)$ and $C^2_b(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R}^k \otimes \ell)$.

Algorithm 1: Nested MLMC with Picard scheme

**Input:** Initial measure $\mu^0$ for $Y^{i,0,\ell}$, global lipschitz payoff function $C^2_P \ni P : \mathbb{R}^d \rightarrow \mathbb{R}$ and accuracy level $\epsilon < e^{-1}$

**Output:** $M_t(\mathcal{M}^{(M)}(P))$, the approximation for our goal $\mathbb{E}[P(X_T)]$.

1. Fix parameters $M$ and $L$ from $\epsilon$;
2. Given $\mu^0 = Law(Y^{i,0,0})$, sample $\{Y^{i,0,0}_{t_k}\}_{k=0}^{2L}$;
3. for $m = 1$ to $M - 1$ do
   4. During $m$th Picard step, given samples $\{Y^{i,m-1,\ell}_{t_k}\}_{k=0}^{2L}$, take (1.9) and run MLMC to obtain $\{Y^{i,m,\ell}_{t_k}\}_{k=0}^{2L}$. This requires calculating
      \[
      (\mathcal{M}^{(m-1)}_{t_0}(b(x,\cdot)), \ldots, \mathcal{M}^{(m-1)}_{t_{2L}}(b(x,\cdot))), (\mathcal{M}^{(m-1)}_{t_0}(\sigma(x,\cdot)), \ldots, \mathcal{M}^{(m-1)}_{t_{2L}}(\sigma(x,\cdot))),
      \]
      where in place of $x$, we put particles $\{Y^{i,m,\ell}_{t_k}\}_{k=0}^{2^{m-1}L}$;
   5. Given samples $\{Y^{i,M-1,\ell}_{t_k}\}_{k=0}^{2L}$, run standard MLMC (with interpolation) to obtain the final vector of approximations $\{\mathcal{M}^{(M)}_{t_0}(P), \ldots, \mathcal{M}^{(M)}_{t_{2L}}(P)\}$;
6. Return $\mathcal{M}^{(M)}_{t_{2L}}(P)$, i.e. $\mathcal{M}^{(M)}(P)$. 

8
2 Analysis of iterated McKean-Vlasov SDEs

Let \((\Omega, \mathcal{F}, P)\) be a probability space with an \(r\)-dimensional Brownian motion \(\{W_t\}_{t \in [0,T]}\). We consider the following dynamics depending on functions \(b: \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d\) and \(\sigma: \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{d \otimes r}\):

\[
dX_t = \left(\int_{\mathbb{R}^d} b(X_t, y) \mu^X_t(dy)\right) dt + \left(\int_{\mathbb{R}^d} \sigma(X_t, y) \mu^X_t(dy)\right) dW_t, \quad t \in [0,T],
\]

and postulate the following assumptions.

(Ker-Reg) The kernels \(b\) and \(\sigma\) belong to the sets \(C^{2,1}_b(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R}^d) \cap C^2_p(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R}^d)\) and \(C^{2,1}_b(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R}^{d \otimes r}) \cap C^2_p(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R}^{d \otimes r})\) respectively.

(\(\mu_0\)-Lp) The initial law \(\mu_0 := \mu^X_0\) satisfies the following condition: for \(q \geq 2\), \(\mu_0 \in L^q(\Omega; \mathbb{R}^d)\), i.e.

\[
\int_{\mathbb{R}^d} |x|^q \mu_0(dx) < \infty.
\]

Note that if (Ker-Reg) holds, then by the mean value theorem,

(Lip) the kernels \(b\) and \(\sigma\) are globally Lipschitz, i.e. for all \(x_1, x_2, y_1, y_2 \in \mathbb{R}^d\), there exists a constant \(L\) such that

\[
|b(x_1, y_1) - b(x_2, y_2)| + \|\sigma(x_1, y_1) - \sigma(x_2, y_2)\| \leq L(|x_1 - x_2| + |y_1 - y_2|).
\]

If (Lip) and (\(\mu_0\)-Lp) hold, then a weak solution to (2.1) exists and pathwise uniqueness holds. In other words \(\{X_t\}_{t \geq 0}\) induces a unique probability measure on \(C([0,T], \mathbb{R}^d)\) (see [Sznitman, 1991]). Furthermore it has a property that

\[
\sup_{0 \leq t \leq T} \mathbb{E}|X_t|^q < \infty.
\]

The additional smoothness stipulated in (Ker-Reg) is needed in the analysis of weak approximation errors.

2.1 Abstract framework - SDEs with random coefficients

To study the convergence of iterated MLMC estimator, it is useful to introduce an abstract framework. This abstract framework in particular allows us to study one iteration of the method described above. More generally, it can also be applied to other algorithms by replacing the drift and diffusion terms (in (2.3)) by other estimators that are of interest.

Let \(\bar{B}: \mathbb{R}^d \times \mathcal{R}(\mathbb{R}^d)^{\otimes \kappa} \to \mathbb{R}^d\) and \(\bar{\Sigma}: \mathbb{R}^d \times \mathcal{R}(\mathbb{R}^d)^{\otimes \kappa} \to \mathbb{R}^{d \otimes r}\) be some measurable functions and \(V_t = (V_t^{(1)}, \ldots, V_t^{(\kappa)}) \in \mathcal{R}(\mathbb{R}^d)^{\otimes \kappa}\) be a \(\kappa\)-dimensional vector of random measures on \(\mathbb{R}^d\), for each \(t \in [0,T]\) (the precise conditions that we impose on \(\bar{B}, \bar{\Sigma}\) and \(V\) will be
presented in Section (2.2)). In the case of iterative particle systems, in place of $V$, we put an empirical measure coming from the previous step of the iteration. We consider SDEs with random coefficients of the form

$$dU_t = \bar{B}(U_t, V_t)dt + \bar{\Sigma}(U_t, V_t)dW_t, \quad \mu^U_0 = \mu^X_0. \quad (2.3)$$

The solution of this SDE is well-defined under the assumptions in Section 2.2, by [Krylov, 2002]. For $\ell = 1, \ldots, L$, we the corresponding Euler approximation of (2.3) at level $\ell$ is given by

$$dZ^\ell_t = \bar{B}(Z^\ell_{\eta(t)}(t), V_{\eta(t)}(t))dt + \bar{\Sigma}(Z^\ell_{\eta(t)}(t), V_{\eta(t)}(t))dW_t, \quad \mu^Z_0 = \mu^X_0. \quad (2.4)$$

We require that $V$ does not depend on $\ell$ and that $(W_t)_{t\in[0,T]}$ is independent of $V$. Subsequently, we define a particle system $\{Z^i, \ell\}$ as follows,

$$dZ^i,\ell_t = \bar{B}(Z^i,\ell_{\eta(t)}(t), V_{\eta(t)}(t))dt + \bar{\Sigma}(Z^i,\ell_{\eta(t)}(t), V_{\eta(t)}(t))dW^i_t, \quad \mu^{Z^i,\ell}_0 = \mu^{X^i}_0, \quad (2.5)$$

where $(W^i)$ are independent Brownian motions that are independent of $V$ for each $i$. Moreover, we require the initial conditions $Z^i,\ell_0$ to be i.i.d., for $1 \leq i \leq N_\ell$, $0 \leq \ell \leq L$.

As a shorthand notation for subsequent analysis, for any measurable function $F : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, we define

$$\int_{\mathbb{R}^d} F(x, y)\nu_t(dy) := \left( \int_{\mathbb{R}^d} F(x, y)\nu^{(1)}_t(dy), \ldots, \int_{\mathbb{R}^d} F(x, y)\nu^{(\kappa)}_t(dy) \right), \quad (2.6)$$

where the Euclidean norm is defined in the usual way as

$$\left| \int_{\mathbb{R}^d} F(x, y)\nu_t(dy) \right| := \sum_{k=1}^{\kappa} \left| \int_{\mathbb{R}^d} F(x, y)\nu^{(k)}_t(dy) \right|. \quad (2.7)$$

### 2.2 Analysis of the abstract framework

Using the notation defined in the previous section, we can formulate the conditions that allow us to study the convergence of the iterated particle system. We consider equations (2.4) and (2.5) and impose the following assumptions:

(V ⊥ (W, Z_0)) Independence: The random measure $\nu$ is independent of $W^i$ and $Z^i,\ell_0$ (or for the generic case, independent of $W$ and $Z^0_0$).

(V-bound) Integrability: For each $p \geq 1$,

$$\sup_{0 \leq s \leq T} \mathbb{E} \left| \int_{\mathbb{R}^d} |y|^p \nu_s(dy) \right| < \infty.$$
There exists a constant $c$ such that
\[
\sup_{x \in \mathbb{R}^d} \sup_{0 \leq s \leq t \leq T} \mathbb{E} \left[ \left| \tilde{B}(x, V_t) - \tilde{B}(x, V_s) \right|^2 + \left| \tilde{\Sigma}(x, V_t) - \tilde{\Sigma}(x, V_s) \right|^2 \right] \leq c(t - s).
\]

The kernels $\tilde{B} : \mathbb{R}^d \times \mathcal{R}(\mathbb{R}^d)^{\otimes \kappa} \rightarrow \mathbb{R}^d$ and $\tilde{\Sigma} : \mathbb{R}^d \times \mathcal{R}(\mathbb{R}^d)^{\otimes \kappa} \rightarrow \mathbb{R}^{d \otimes r}$ satisfy the following relationships with the random measure $V$: for each $t \in [0, T]$, there exists a constant $c$ such that
\[
|\tilde{B}(x_1, V_t) - \tilde{B}(x_2, V_t)| + |\tilde{\Sigma}(x_1, V_t) - \tilde{\Sigma}(x_2, V_t)| \leq c|x_1 - x_2|, \quad \forall x_1, x_2 \in \mathbb{R}^d,
\]
\[
|\tilde{B}(x, V_t)| + |\tilde{\Sigma}(x, V_t)| \leq c \left( 1 + |x| + \int_{\mathbb{R}^d} |y| |V_t(dy)| \right), \quad \forall x \in \mathbb{R}^d.
\]

**Analysis of conditional MLMC variance** From now onwards, we denote by $c$ a generic constant that depends on $T$, but not on $\ell$, $h_t$ or $N_\ell$. We first consider the integrability of process (2.4).

**Lemma 2.1.** Let $Z_\ell$ be defined as in (2.4). Assume (V-Lip) and ($\mu_0$-Lp). Then for any $p \geq 2$ and $\ell \geq 0$, there exists a constant $c$ such that
\[
\mathbb{E} \left[ \sup_{t \in [0, T]} |Z_\ell|^p \right] \leq c \left( 1 + \mathbb{E} \left[ \int_0^T |\int_{\mathbb{R}^d} |y| V_t(dy)| \right] ds \right).
\]

**Proof.** Given any $\ell$, let us define a sequence of stopping times $\tau_M := \inf \{ t \geq 0 : |Z_\ell^t - Z_0^\ell| \geq M \}$. For any $t \in [0, T]$, we consider the stopped process $Z_{\ell \wedge \tau_M}^t$ and compute
\[
|Z_{\ell \wedge \tau_M}^t|^p \leq c \left( |Z_0^\ell|^p + \int_0^t |\tilde{B}(Z_{\ell \wedge \tau_M}^s, V_{\eta(s)})| ds \right)^p
\]
\[
+ \int_0^t |\tilde{\Sigma}(Z_{\ell \wedge \tau_M}^s, V_{\eta(s)})| ds \right)^p.
\]
By the Burkholder-Davis-Gundy and Hölder inequalities,
\[
\mathbb{E} \left[ \sup_{0 \leq u \leq t} |Z_{\ell \wedge \tau_M}^u|^p \right] \leq c \left( \mathbb{E} \left[ |Z_0^\ell|^p \right] + t^{p-1} \mathbb{E} \left[ \int_0^t |\tilde{B}(Z_{\ell \wedge \tau_M}^s, V_{\eta(s)})|^p ds \right]
\]
\[
+ t^{\frac{p}{2}-1} \mathbb{E} \left[ \int_0^t |\tilde{\Sigma}(Z_{\ell \wedge \tau_M}^s, V_{\eta(s)})|^p ds \right] \right).
\]
Assumptions (V-Lip) and ($\mu_0$-Lp) yield
\[
\mathbb{E} \left[ \sup_{0 \leq u \leq t} |Z_{\ell \wedge \tau_M}^u|^p \right] \leq c \left( 1 + \mathbb{E} \left[ \int_0^t |\int_{\mathbb{R}^d} |y| V_{\eta(s)}(dy)| ds \right] + \int_0^t \mathbb{E} \left[ \sup_{0 \leq u \leq s} |Z_{\ell \wedge \tau_M}^u|^p \right] ds \right).
\]
Note that, by \((\mu_0 - L_p)\),
\[
\mathbb{E}\left[ \sup_{0 \leq u \leq s} |Z_{u \wedge \tau_M}^\ell|^p \right] \leq c\left( \mathbb{E}\left[ \sup_{0 \leq u \leq s} |Z_{u \wedge \tau_M}^\ell - Z_0^\ell|^p \right] + \mathbb{E}|Z_0^\ell|^p \right) \leq c\left( M^p + \mathbb{E}|Z_0^\ell|^p \right) < +\infty.
\]
By Gronwall’s lemma,
\[
\mathbb{E}\left[ \sup_{0 \leq u \leq t} |Z_{u \wedge \tau_M}^\ell|^p \right] \leq c\left( 1 + \mathbb{E}\left[ \int_0^T \int_{\mathbb{R}^d} |y|^p V_{\eta(s)}(dy) |ds \right] \right).
\]
Furthermore, since \(\sup_{0 \leq t \leq T} |Z_{t \wedge \tau_M}^\ell|^p\) is a non-decreasing sequence (in \(M\)) converging pointwise to \(\sup_{0 \leq t \leq T} |Z_th^\ell|^p\), the lemma follows from the monotone convergence theorem. \(\Box\)

The following two lemmas focus on the regularity of \(Z_{\ell}^\ell\) in time and its strong convergence property. The first lemma bounds the difference in \(Z_{\ell}^\ell\) over two time points, at a fixed level \(\ell\). The second lemma bounds the difference in \(Z_{\ell}^\ell\) over adjacent levels, at a fixed time \(t\).

**Lemma 2.2** (Regularity of \(Z_{\ell}^\ell\)). Let \(Z_{\ell}^\ell\) be defined as in (2.4). Assume \((\mathcal{V}\text{-Lip})\) and \((\mathcal{V}\text{-bound})\). Then, for \(p \geq 1\), \(0 \leq u \leq s \leq T\),
\[
\left( \mathbb{E}|Z_{s}^\ell - Z_{u}^\ell|^p \right)^{\frac{1}{p}} \leq c(s - u)^{\frac{1}{2}}.
\]

**Proof.** From the definition of \(Z_{\ell}^\ell\), we obtain
\[
Z_{s}^\ell - Z_{u}^\ell = \int_u^s B(Z_{\eta(t)}^\ell, \nu_{\eta(t)}) dt + \int_u^s \Xi(Z_{\eta(t)}^\ell, \nu_{\eta(t)}) dW_t.
\]
Using the inequality that \(|a + b|^p \leq 2^{p-1}(|a|^p + |b|^p)\), we have
\[
|Z_{s}^\ell - Z_{u}^\ell|^p \leq 2^{p-1}\left[ \int_u^s B(Z_{\eta(t)}^\ell, \nu_{\eta(t)}) dt \right]^p + \int_u^s |\Xi(Z_{\eta(t)}^\ell, \nu_{\eta(t)})| dt\].
\]
Then, the \(p\)th moment inequality from [Mao, 2007] gives
\[
\mathbb{E}|Z_{s}^\ell - Z_{u}^\ell|^p \leq c \left( \mathbb{E}\left[ \int_u^s B(Z_{\eta(t)}^\ell, \nu_{\eta(t)}) dt \right]^p + (s - u) \frac{1}{2^{p-1}} \mathbb{E}\int_u^s \|\Xi(Z_{\eta(t)}^\ell, \nu_{\eta(t)})\|^p dt \right).
\]
Assumption \((\mathcal{V}\text{-Lip})\) leads to the estimate
\[
\mathbb{E}|Z_{s}^\ell - Z_{u}^\ell|^p \leq c \left[ (s - u)^{p-1} \mathbb{E}\int_u^s \left( 1 + |Z_{\eta(t)}^\ell| + \left| \int_{\mathbb{R}^d} y|\nu_{\eta(t)}(dy) \right| \right) dt \right. + \left. (s - u) \frac{1}{2^{p-1}} \mathbb{E}\int_u^s \left( 1 + |Z_{\eta(t)}^\ell| + \left| \int_{\mathbb{R}^d} y|\nu_{\eta(t)}(dy) \right| \right) dt \right].
\]
Using assumption \((\mathcal{V}\text{-bound})\) along with Lemma 2.1 completes the proof. \(\Box\)
Lemma 2.3 (Strong convergence of $Z_t^\ell$). Assume (V-Lip), (V-bound) and (V-Reg). Then for any $\ell \in \{1, 2, \ldots, L\}$, there exists a constant $c > 0$ such that

$$E \left[ \sup_{0 \leq t \leq T} |Z_t^\ell - Z_t^{\ell-1}|^2 \right] \leq ch_\ell.$$

Proof. We express $(Z_t^\ell)_{t \in [0,T]}$ as

$$Z_t^\ell = Z_0 + \int_0^t \bar{B}(Z_{s-1}^\ell, \mathcal{V}_{s-1}(s)) ds + \bar{\Sigma}(Z_{s-1}^\ell, \mathcal{V}_{s-1}(s)) dW_s + T(t),$$

where $T(t)$ is given by

$$T(t) = \int_0^t \left( \bar{B}(Z_{s-1}^\ell, \mathcal{V}_{s-1}(s)) - \bar{B}(Z_{s-1}^{\ell-1}, \mathcal{V}_{s-1}(s)) \right) ds + \int_0^t \left( \bar{\Sigma}(Z_{s-1}^\ell, \mathcal{V}_{s-1}(s)) - \bar{\Sigma}(Z_{s-1}^{\ell-1}, \mathcal{V}_{s-1}(s)) \right) dW_s.$$

Let $\delta Z_t := Z_t^\ell - Z_t^{\ell-1}$ and for $0 \leq u \leq T$, we write

$$\delta Z_u = \int_0^u \left( \bar{B}(Z_{s-1}^\ell, \mathcal{V}_{s-1}(s)) - \bar{B}(Z_{s-1}^{\ell-1}, \mathcal{V}_{s-1}(s)) \right) ds + \int_0^u \left( \bar{\Sigma}(Z_{s-1}^\ell, \mathcal{V}_{s-1}(s)) - \bar{\Sigma}(Z_{s-1}^{\ell-1}, \mathcal{V}_{s-1}(s)) \right) dW_s + T(u).$$

By the Burkholder-Davis-Gundy inequality,

$$E \left[ \sup_{0 \leq u \leq t} |\delta Z_u|^2 \right] \leq c E \left[ \int_0^t \| \bar{B}(Z_{s-1}^\ell, \mathcal{V}_{s-1}(s)) - \bar{B}(Z_{s-1}^{\ell-1}, \mathcal{V}_{s-1}(s)) \|^2 ds + \sup_{u \leq t} |T(u)|^2 \right].$$

By the Gronwall’s lemma, assumption (V-Lip), and the fact that $\sup_{u \leq s} |Z_{s}(u)| \leq \sup_{u \leq s} |Z_u|$, we have

$$E \left[ \sup_{0 \leq u \leq t} |\delta Z_u|^2 \right] \leq c E \left[ \int_0^t \sup_{u \leq s} \| \delta Z_u \|^2 ds + \sup_{u \leq t} |T(u)|^2 \right].$$

Moreover, assumption (V-Lip) and the BDG inequality imply that

$$E \left[ \sup_{0 \leq u \leq t} |T(u)|^2 \right] \leq c E \left[ \int_0^t \| \bar{B}(Z_{s-1}^\ell, \mathcal{V}_{s-1}(s)) - \bar{B}(Z_{s-1}^{\ell-1}, \mathcal{V}_{s-1}(s)) \|^2 ds + \int_0^t \| \bar{\Sigma}(Z_{s-1}^\ell, \mathcal{V}_{s-1}(s)) - \bar{\Sigma}(Z_{s-1}^{\ell-1}, \mathcal{V}_{s-1}(s)) \|^2 ds \right].$$

We conclude the result by using Lemma 2.2 and (V-Reg). \qed
We define the interpolated empirical measures

\[
\tilde{\nu}_{t}^{i,\ell,N} := \begin{cases} 
\frac{1}{N} \sum_{i=1}^{N} \delta_{Z_{i}^{t,\ell}} & , t \in \Pi^{\ell}, \\
\left[\frac{t - \eta_{\ell}(t)}{h_{\ell}}\right] \tilde{\nu}_{\eta_{\ell}(t) + h_{\ell}}^{i,\ell,N} + \left[1 - \frac{t - \eta_{\ell}(t)}{h_{\ell}}\right] \tilde{\nu}_{\eta_{\ell}(t)}^{i,\ell,N} & , t \not\in \Pi^{\ell},
\end{cases}
\tag{2.8}
\]

and the corresponding MLMC operator \( \mathcal{M}_{i} \) as

\[
\mathcal{M}_{i}(F(x, \cdot)) = \sum_{\ell=0}^{L} \left( \tilde{\nu}_{t}^{i,\ell,N} - \tilde{\nu}_{t}^{i,\ell-1,N} \right)(F(x, \cdot)), \quad \tilde{\nu}_{t}^{i,1,N} := 0.
\]

Note that to allow for flexibility of the notation, we apply this definition componentwise. As classically for MLMC, each difference in the telescopic sum involves the same Brownian motion that gives coupling.

We also define the \( \sigma \)-algebra \( \mathcal{F}_{T}^{\mathcal{V}} = \{ \sigma(\mathcal{V}_{s})_{0 \leq s \leq t} \} \). Since samples \( \{Z_{i,\ell}(t)\}_{i=1,\ldots,N_{\ell}, \ell=0,\ldots,L} \) conditioned on \( \mathcal{F}_{T}^{\mathcal{V}} \) are independent, we can bound the conditional MLMC variance as follows.

\textbf{Lemma 2.4.} Assume \((\mathcal{V}, \text{Lip}), (\mathcal{V}, \text{bound})\) and \((\mathcal{V}, \text{Reg})\). Let \( \{\mu_{t}\}_{t \in [0, T]} \) be a collection of probability measures on \( \mathbb{R}^{d} \) such that

\[
\sup_{t \in [0, T]} \int_{\mathbb{R}^{d}} |x|^{2} \mu_{t}(dx) < +\infty.
\]

Then for any globally Lipschitz continuous function \( G : \mathbb{R}^{d} \times \mathbb{R}^{d} \to \mathbb{R} \), there exists a constant \( c \) such that

\[
\sup_{0 \leq t \leq T} \int_{\mathbb{R}^{d}} \mathbb{E} \left[ \text{Var} \left( \mathcal{M}_{\eta_{\ell}(t)}(G(x, \cdot)) \bigg| \mathcal{F}_{T}^{\mathcal{V}} \right) \right] \mu_{t}(dx) \leq c \sum_{\ell=0}^{L} \frac{h_{\ell}}{N_{\ell}}, \tag{2.9}
\]

\textbf{Proof.} Assumption \((\mathcal{V} \perp (W, Z_{0}))\) implies that

\[
\mathbb{E} \left[ \text{Var} \left( \mathcal{M}_{\eta_{\ell}(t)}(G(x, \cdot)) \bigg| \mathcal{F}_{T}^{\mathcal{V}} \right) \right] = \sum_{i=1}^{N_{\ell}} \frac{1}{N_{\ell}^{2}} \mathbb{E} \left[ \text{Var} \left( P_{\eta_{\ell}(t)}^{i,0} \mid \mathcal{F}_{T}^{\mathcal{V}} \right) \right] + \sum_{\ell=1}^{L} \sum_{i=1}^{N_{\ell}} \frac{1}{N_{\ell}^{2}} \mathbb{E} \left[ \text{Var} \left( P_{\eta_{\ell}(t)}^{i,\ell} - P_{\eta_{\ell}(t)}^{i,\ell-1} \bigg| \mathcal{F}_{T}^{\mathcal{V}} \right) \right],
\]

where

\[
P_{\eta_{\ell}(t)}^{i,\ell} := (1 - \lambda_{\ell}^{i})G(x, Z_{\eta_{\ell}(t)}^{i,\ell}) + \lambda_{\ell}^{i}G(x, Z_{\eta_{\ell}(t) + h_{\ell}}^{i,\ell}), \tag{2.10}
\]

\( \lambda_{\ell}^{i} = \frac{\eta_{\ell}(t) - \eta_{\ell}(t)}{h_{\ell}} \in [0, 1]. \) Using the fact that \( \mathbb{E}[\text{Var}(X)] \leq \text{Var}(X) \leq \mathbb{E}[X^{2}] \), we obtain the bound

\[
\mathbb{E} \left[ \text{Var} \left( \mathcal{M}_{\eta_{\ell}(t)}(G(x, \cdot)) \bigg| \mathcal{F}_{T}^{\mathcal{V}} \right) \right] \leq \sum_{i=1}^{N_{\ell}} \frac{1}{N_{\ell}^{2}} \mathbb{E} \left[ P_{\eta_{\ell}(t)}^{i,0} \right]^{2} + \sum_{\ell=1}^{L} \sum_{i=1}^{N_{\ell}} \frac{1}{N_{\ell}^{2}} \mathbb{E} \left[ P_{\eta_{\ell}(t)}^{i,\ell} - P_{\eta_{\ell}(t)}^{i,\ell-1} \right]^{2}.
\]
Since $G$ is Lipschitz, it has linear growth. By Lemma 2.1, it follows that
\[ \mathbb{E} \left[ P_{nL}^{i,0} \right]^2 \leq c \sup_{0 \leq t \leq T} \int_{\mathbb{R}^d} \left( x^2 + \mathbb{E} \left[ |Z_{\eta L}(t)|^2 \right] + \mathbb{E} \left[ |Z_{\eta L}(t) + \mu_t|^2 \right] \right) \mu_t(dx) < +\infty. \]

Next, we consider levels $\ell \in \{1, \ldots, L\}$. Recall from (2.10) that
\[
\begin{align*}
P_{nL}^{i,\ell}(t) &= (1 - \lambda^\ell_t) G(x, Z_{\eta L}(t)) + \lambda^\ell_t G(x, Z_{\eta L}(t) + h_t), \\
P_{nL}^{i,\ell-1}(t) &= (1 - \lambda^{\ell-1}_t) G(x, Z_{\eta L}(t)) + \lambda^{\ell-1}_t G(x, Z_{\eta L}(t) + h_{\ell-1}).
\end{align*}
\]

We decompose the error as follows.
\[
|P_{nL}^{i,\ell}(t) - P_{nL}^{i,\ell-1}(t)| \leq (1 - \lambda^{\ell-1}_t) \left| G(x, Z_{\eta L}^{i,\ell}(t)) - G(x, Z_{\eta L}^{i,\ell-1}(t)) \right| + \lambda^{\ell-1}_t \left| G(x, Z_{\eta L}^{i,\ell}(t) + h_t) - G(x, Z_{\eta L}^{i,\ell-1}(t) + h_{\ell-1}) \right|
\]
\[
+ |\lambda^\ell_t - \lambda^{\ell-1}_t| \left| G(x, Z_{\eta L}^{i,\ell}(t) + h_t) - G(x, Z_{\eta L}^{i,\ell}(t)) \right|.
\]

By Lemma 2.3,
\[
\mathbb{E} \left| G(x, Z_{\eta L}^{i,\ell}(t)) - G(x, Z_{\eta L}^{i,\ell-1}(t)) \right|^2 \leq c \ell \tag{2.11}
\]
\[
\mathbb{E} \left| G(x, Z_{\eta L}^{i,\ell}(t) + h_t) - G(x, Z_{\eta L}^{i,\ell-1}(t) + h_{\ell-1}) \right|^2 \leq c h_t. \tag{2.12}
\]

Also, by Lemma 2.2,
\[
\mathbb{E} \left| G(x, Z_{\eta L}^{i,\ell-1}(t)) - G(x, Z_{\eta L}^{i,\ell-1}(t)) \right|^2 \leq c(\eta_t(\eta L(t)) - \eta_{\ell-1}(\eta L(t))) \leq c \ell, \tag{2.13}
\]
\[
\mathbb{E} \left| G(x, Z_{\eta L}^{i,\ell-1}(t) + h_t) - G(x, Z_{\eta L}^{i,\ell-1}(t) + h_{\ell-1}) \right|^2 \leq c(\eta_t(\eta L(t))) \leq c \ell, \tag{2.14}
\]
\[
\mathbb{E} \left| G(x, Z_{\eta L}^{i,\ell}(t) + h_t) - G(x, Z_{\eta L}^{i,\ell}(t)) \right|^2 \leq c(\eta_t(\eta L(t)) + h_t - \eta_{\ell}(\eta L(t))) \leq c \ell. \tag{2.15}
\]

We obtain (2.9) by combining (2.11), (2.13), (2.14) and (2.15). Since $t$ and $x$ are arbitrary, the proof is complete.

\[\square\]

### 2.3 Weak error analysis

The key result of this section is Theorem 2.7. The proofs follow from the weak error analysis of the relevant SDEs. We consider SDE (2.1) and define deterministic functions $B : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $\Sigma : [0, T] \times \mathbb{R}^d \to \mathbb{R}^{d \times r}$ to be $B(t, x) = \mathbb{E}[b(x, X_t)]$ and $\Sigma(t, x) = \mathbb{E}[\sigma(x, X_t)]$ respectively.
Lemma 2.5. Assume $(\mu_0-L_{\mu})$ and (Ker-Reg). Let $X_t$ be defined as in (2.1). Then $B(\cdot, \cdot)$ and $\Sigma(\cdot, \cdot)$ satisfy

$$B(\cdot, \cdot) \in C^{1,2}_b([0, T] \times \mathbb{R}^d, \mathbb{R}^d) \text{ and } \Sigma(\cdot, \cdot) \in C^{1,2}_b([0, T] \times \mathbb{R}^d, \mathbb{R}^{d \otimes r}).$$

Proof. Regularity of $B(\cdot, \cdot)$ and $\Sigma(\cdot, \cdot)$ in $x$ follows from (Ker-Reg). Moreover, for any $x \in \mathbb{R}^d$, $s \in [0, T]$ and $t \in [s, T]$, we apply Itô formula to each coordinate $k \in \{1, \ldots, d\}$ to get

$$b_k(x, X_t) = b_k(x, X_s) + \int_s^t \sum_{j=1}^d \sum_{i=1}^r D_y b_k(x, X_u) \Sigma_{ji}(u, X_u) dW^i_u$$

$$+ \int_s^t \sum_{j=1}^d D_y b_k(x, X_u) B_j(u, X_u) du + \frac{1}{2} \int_s^t \sum_{i,j=1}^d D_y b_k(x, X_u) a_{ij}(u, X_u) du,$$

where $a(t, x) = \Sigma(t, x) \Sigma(t, x)^T$ and $D_y b_k$ indicates the partial derivative w.r.t. the $i$th component in the second argument of $b_k$ (a similar notation for the second order partial derivatives $D_{y_j} b_k$). Assumptions (Ker-Reg), (Lip), $(\mu_0-L_{\mu})$ and (2.2) imply that

$$\mathbb{E}\left[ \int_s^t \sum_{j=1}^d \sum_{i=1}^r D_y b_k(x, X_u) \Sigma_{ji}(u, X_u) dW^i_u \right] = 0.$$

By the fundamental theorem of calculus,

$$\frac{\partial}{\partial t} B_k(t, x) = \frac{\partial}{\partial t} \mathbb{E}[b_k(x, X_t)] = \mathbb{E}\left[ \sum_{j=1}^d D_y b_k(x, X_t) B_j(t, X_t) + \frac{1}{2} \sum_{i,j=1}^d D_{y_i, y_j} b_k(x, X_t) a_{ij}(t, X_t) \right].$$

By (Ker-Reg), we know that $D_y b_k$ and $D_{y_i, y_j} b_k$ are bounded, for any $i, j, k \in \{1, \ldots, d\}$. Moreover, by (Lip), we know that $B_j(t, x)$ and $a_{ij}(t, x)$ are respectively of linear and quadratic polynomial growth in the space variable $x$. Therefore, by (2.2), we conclude that $\frac{\partial}{\partial t} B_k(t, x)$ is bounded. Similarly, we apply the same argument to the kernel $\sigma(\cdot, \cdot)$ to conclude the result.

To begin the weak error analysis, we fix an arbitrary point $t \in [0, T]$. For any $0 \leq s \leq t$, we consider the stochastic flow $((X_u)^{s,x}, u \in [s, t])$, defined by

$$(X_u)^{s,x} = x + \int_s^u B(\theta, (X_{\theta})^{s,x}) d\theta + \int_s^u \Sigma(\theta, (X_{\theta})^{s,x}) dW_{\theta}, \quad u \in [s, t].$$

Notice that (2.17) is a Markov process. Next, for any globally Lipschitz continuous function $G: \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, we consider the function

$$v_y(s, x) := \mathbb{E}[G(y, (X_t)^{s,x})], \quad y \in \mathbb{R}^d \text{ and } (s, x) \in [0, t] \times \mathbb{R}^d.$$
Lemma 2.6. Assume \((\text{Ker-Reg})\) and \((\mu_0\text{-L}_p)\).

(a) Suppose that \(G \in C^2_b(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R})\). Then there exists a constant \(L > 0\) such that
\[
\left\{ \begin{array}{l}
\sum_{k=1}^d \sup_{y \in \mathbb{R}^d} \left\| \frac{\partial v_y}{\partial x_k}(s, x) \right\|_\infty \leq L, \quad \forall (s, x) \in [0, t] \times \mathbb{R}^d, \\
\sum_{i,j=1}^d \sup_{y \in \mathbb{R}^d} \left\| \frac{\partial^2 v_y}{\partial x_i \partial x_j}(s, x) \right\|_\infty \leq L, \quad \forall (s, x) \in [0, t] \times \mathbb{R}^d.
\end{array} \right.
\]

(b) Suppose that \(G \in C^2_p(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R})\). Then there exists a constant \(L > 0\) such that
\[
\left\{ \begin{array}{l}
\sup_{y \in \mathbb{R}^d} \sum_{k=1}^d \left| \frac{\partial v_y}{\partial x_k}(s, x) \right| \leq L(1 + |x|^p), \quad \forall (s, x) \in [0, t] \times \mathbb{R}^d, \\
\sup_{y \in \mathbb{R}^d} \sum_{i,j=1}^d \left| \frac{\partial^2 v_y}{\partial x_i \partial x_j}(s, x) \right| \leq L(1 + |x|^p), \quad \forall (s, x) \in [0, t] \times \mathbb{R}^d.
\end{array} \right.
\]

Proof. By Theorem 5.5.3 in [Friedman, 2006], we note that the variational derivatives of \(X_{s,x}^u\) (defined in the sense of Gâteaux derivatives under the \(L^2\) norm) w.r.t. the components of \(x\) satisfy the following system of SDEs:
\[
\frac{\partial}{\partial x_i}(X_{s,x}^u(j)) = \delta_{i,j} + \int_s^t \sum_{k=1}^d D_{y_k} B_j(r, X_{r,x}^u) \frac{\partial}{\partial x_i}(X_{r,x}^u(k)) \, dr \\
+ \int_s^t \sum_{\ell=1}^r \sum_{k=1}^d D_{y_k} \Sigma_{j\ell}(r, X_{r,x}^u) \frac{\partial}{\partial x_i}(X_{r,x}^u(k)) \, dW_{r,\ell}, \quad 1 \leq i, j \leq d.
\]

From this linear SDE with bounded first-order derivatives of \(B\) and \(\Sigma\), it is straightforward to deduce that
\[
\sup_{x \in \mathbb{R}^d} \sup_{s \in [0, t]} \mathbb{E} \left[ \left| \frac{\partial}{\partial x_i}(X_{s,x}^u(j)) \right|^2 \right] < +\infty. \tag{2.19}
\]

Theorem 5.5.5 in [Friedman, 2006] establishes that
\[
\frac{\partial v_y}{\partial x_i}(s, x) = \sum_{j=1}^d \mathbb{E} \left[ D_{y_j} G(y, X_{s,x}^u) \frac{\partial}{\partial x_i}(X_{s,x}^u)(j) \right]. \tag{2.20}
\]

By (2.19), it is clear that the assertion for the first order derivatives in \((\text{v-diff-Reg+})\) holds if \(G \in C^2_b(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R})\). On the other hand, if \(G \in C^2_p(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R})\), then by the Cauchy-Schwarz inequality, (2.2) implies that there exists a constant \(c\) (whose value may change
from line to line) such that

\[
\left| \frac{\partial v_y}{\partial x_i}(s, x) \right| \leq c \sum_{j=1}^{d} \sqrt{E[|D_{y_j} G(y, X^s,x)|^2]}
\]

\[
\leq c \sqrt{E[1 + |X^s,x|^2]}
\]

\[
\leq c \left( 1 + \sup_{t \in [0,T]} E[|X^s,0|^2] + |x|^2 \right)
\]

which proves the assertion for the first order derivatives in \((v\text{-diff-Reg})\). Similarly, we can derive respectively the analogous statements of (2.19) and (2.20) for the second order derivatives and prove the assertions of the second order derivatives in (a) and (b). (For part (b), we use again the assumption that the second order derivatives of \(G\) have polynomial growth only in the second argument.)

By the Feynman-Kac theorem ([Krylov, 1980]), it can be shown that \(v_y(\cdot, \cdot)\) satisfies the following Cauchy problem,

\[
\begin{align*}
\frac{\partial v_y}{\partial s}(s, x) + \frac{1}{2} \sum_{i,j=1}^{d} A_{ij}(s, x) \frac{\partial^2 v_y}{\partial x_i \partial x_j}(s, x) + \sum_{j=1}^{d} B_j(s, x) \frac{\partial v_y}{\partial x_j}(s, x) &= 0, \quad (s, x) \in [0, t] \times \mathbb{R}^d, \\
v_y(t, x) &= G(y, x),
\end{align*}
\]

(2.21)

where \(A = \Sigma(s, x) \Sigma(s, x)^T\).

Recall the notation that \(\mathcal{F}^V_t = \{\sigma(V_s)_{0 \leq s \leq t}\}\). Recall from hypothesis \((V \perp (W, Z_0))\) that \(W\) is independent of \(\mathcal{F}^V_t\).

The following theorem reveals the order of weak convergence of (2.4) to (2.1). From now onwards, we denote by \(\mu_t^{Z^L|\mathcal{F}^V_t}\) the regular conditional probability measure of \(Z^L_t\) given \(\mathcal{F}^V_t\). (See Lemma A.2 for further details.) The existence of regular conditional probability measure follows from the fact that we work on a Polish space with the Borel \(\sigma\)–algebra.

**Theorem 2.7.** Let \(G \in C^2_b(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R})\) be a globally Lipschitz continuous function. Assume that \((\text{Ker-Reg})\), \((\mu_0-L_p)\), \((V \perp (W, Z_0))\), \((V\text{-bound})\) and \((V\text{-Lip})\) hold. Then there exists a
constant \( c \) such that for each \( t \in [0, T] \) and \( x \in \mathbb{R}^d \),

\[
\sup_{0 \leq s \leq t} |\mathbb{E}[G(x, Z^L_s)] - \mathbb{E}[G(x, X_s)]| \\
\leq c \left( h_L + \int_0^t \mathbb{E} \left[ \int_{\mathbb{R}^d} \left| \tilde{B}(x, \mathcal{V}_{t,s}(s)) - \mathbb{E}[b(x, X_{\eta_L}(s))] \right| \mu_{\eta_L}(s) (dx) \right] ds \\
+ \int_0^t \mathbb{E} \left[ \int_{\mathbb{R}^d} \left| \tilde{\Sigma}(x, \mathcal{V}_{t,s}(s)) - \mathbb{E}[\sigma(x, X_{\eta_L}(s))] \right| \mu_{\eta_L}(s) (dx) \right] ds \right).
\]

**Proof.** To lighten the notation, we use \( \eta(s) \) to denote \( \eta_L(s) \) in all instances in the proof. First, we observe that

\[
|\mathbb{E}[G(y, Z^L_s)] - \mathbb{E}[G(y, X_s)]| = |\mathbb{E}[\mathbb{E}[G(y, Z^L_s)|\mathcal{F}^y_s]] - \mathbb{E}[G(y, X_s)]| \\
\leq \mathbb{E}[\mathbb{E}[G(y, Z^L_s)|\mathcal{F}^y_s]] - \mathbb{E}[G(y, X_s)]|
\]

From definition of \( v(\cdot, \cdot) \) in (2.18), we compute

\[
\mathbb{E}[v_y(0, X_0)] = \int_{\mathbb{R}^d} v_y(0, x) \mu_0 (dx) = \int_{\mathbb{R}^d} \mathbb{E}[G(y, (X_t)^{0,x})] \mu_0 (dx) \\
= \int_{\mathbb{R}^d} \mathbb{E}[G(y, (X_t)^{0,x})]|_{X_0 = x} \mu_0 (dx) = \mathbb{E}[\mathbb{E}[G(y, X_t)|X_0]].
\]

The Feynman-Kac theorem, hypothesis \((V \perp (W, Z_0))\) and the fact that \( \mu_0^X = \mu_0^{Z^L} \) give

\[
\mathbb{E}[G(y, Z^L_t)|\mathcal{F}^y_t] - \mathbb{E}[G(y, X_t)] = \mathbb{E}[v_y(t, Z^L_t)|\mathcal{F}^y_t] - \mathbb{E}[v_y(0, Z^L_0)] \\
= \mathbb{E}[v_y(t, Z^L_t)|\mathcal{F}^y_t] - \mathbb{E}[v_y(0, Z^L_0)|\mathcal{F}^y_t] \\
= \sum_{k=0}^{n-1} \mathbb{E} \left[ v_y(t^L_{k+1}, Z^L_{k+1}) - v_y(t^L_k, Z^L_k) \bigg| \mathcal{F}^y_t \right],
\]

where \( n = t/h_L \). By Itô’s formula,

\[
\mathbb{E}[v_y(t, Z^L_t)|\mathcal{F}^y_t] - \mathbb{E}[v_y(0, Z^L_0)] = \sum_{k=0}^{n-1} \mathbb{E} \left[ \int_{t^L_k}^{t^L_{k+1}} \left( \frac{\partial v_y}{\partial t} (s, Z^L_s) + \sum_{j=1}^d \frac{\partial v_y}{\partial x_j} (s, Z^L_s) \tilde{B}_j (Z^L_{\eta_L}(s), \mathcal{V}_{t,s}(s)) \right) ds \\
+ \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2 v_y}{\partial x_i \partial x_j} (s, Z^L_s) a_{ij} (Z^L_{\eta_L}(s), \mathcal{V}_{t,s}(s)) ds \\
+ \sum_{k=1}^{n-1} \sum_{j=1}^r \frac{\partial v_y}{\partial x_j} (s, Z^L_s) \tilde{\Sigma}_{j1} (Z^L_{\eta_L}(s), \mathcal{V}_{t,s}(s)) dW_s^{(i)} \right] \bigg| \mathcal{F}^y_t \right],
\]

\(^1\)For simplicity we assume that \( n \) is an integer.
where \( a(x, \mu) = \Sigma(x, \mu) \Sigma(x, \mu)^T \). Condition (\textbf{v-diff-Reg+}), as well as hypotheses (\textbf{Lip}) , \((\mu_0, L_p)\) and (\textbf{V-bound}) , along with Lemma 2.1 and part (a) of Lemma A.1 (with the filtration \( \{ F_t \}_{t \in [0,T]} \) such that \( F_t = \sigma(F_{t^r}, \{ W_u \}_{0 \leq u \leq t}, \{ Z_u^L \}_{0 \leq u \leq t}) \)) imply that

\[
\mathbb{E} \left[ \int_{t_k}^{t_{k+1}} \sum_{j=1}^d \sum_{i=1}^r \frac{\partial v_y}{\partial x_j}(s, Z_s^L(\Sigma a(\sigma(s, Z_s^L))) dW_s^{(i)} \bigg| F_T^L \right] = 0. \tag{2.22}
\]

Subsequently, using the fact that \( v(\cdot, \cdot) \) satisfies PDE (2.21), we have

\[
\mathbb{E}[v_y(t, Z_t^L)|F_T^L] - \mathbb{E}[v_y(0, Z_0^L)] = \sum_{k=0}^{n-1} \int_{t_k}^{t_{k+1}} \mathbb{E} \left[ \sum_{j=1}^d \frac{\partial v_y}{\partial x_j}(s, Z_s^L(\Sigma a(\sigma(s, Z_s^L))) - B_j(s, Z_s^L) \right)
\]
\[
+ \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2 v_y}{\partial x_i \partial x_j}(s, Z_s^L(\Sigma a(\sigma(s, Z_s^L))) - A_{ij}(s, Z_s^L) \bigg| F_T^L] \right) ds,
\]

where \( A(s, x) = \Sigma(s, x) \Sigma(s, x)^T \). Hence,

\[
\mathbb{E}[v_y(t, Z_t^L)|F_T^L] - \mathbb{E}[v_y(0, Z_0^L)] = \sum_{k=0}^{n-1} \int_{t_k}^{t_{k+1}} \mathbb{E} \left[ \sum_{i=1}^d R_i(s) \bigg| F_T^L \right] ds,
\]

where

\[
R_1(s) := \sum_{j=1}^d \frac{\partial v_y}{\partial x_j}(s, Z_s^L(B_j(\eta(s), Z_s^L) - B_j(s, Z_s^L))
\]
\[
R_2(s) := \sum_{j=1}^d \frac{\partial v_y}{\partial x_j}(s, Z_s^L(\Sigma a(\sigma(s, Z_s^L))) - B_j(\eta(s), Z_s^L))
\]
\[
R_3(s) := \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2 v_y}{\partial x_i \partial x_j}(s, Z_s^L(A_{ij}(\eta(s), Z_s^L) - A_{ij}(s, Z_s^L))
\]
\[
R_4(s) := \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2 v_y}{\partial x_i \partial x_j}(s, Z_s^L(A_{ij}(\eta(s), Z_s^L), \Sigma a(\sigma(s, Z_s^L))) - A_{ij}(\eta(s), Z_s^L))].
\]
Using these two bounds along with Lemma \( \text{Lemma A.1} \) and the tower property, we have

\[
R_{\text{Error}} \leq c \sum_{k=1}^{d} \mathbb{E} \left[ \frac{\partial B_k}{\partial u} (u, Z^L_s) \mathbb{E} \left[ \int_{\eta(s)}^{t} \left[ \frac{\partial B_k}{\partial u} (u, Z^L_u) \right] dw + \sum_{i=1}^{d} \frac{\partial B_k}{\partial x_i} (u, Z^L_u) B_i (Z^L_{\eta(u)}, V_{\eta(u)}) \right] du \right].
\]

Let \( \mathcal{F}^Z \) be the sigma-algebra generated by \( \{ Z^L_t \}_{t \in [0, T]} \). From part (a) of Lemma A.1 and the tower property, we have

\[
\mathbb{E} \left[ R_1(s) \mid \mathcal{F}^Z \right] = \sum_{k=1}^{d} \mathbb{E} \left[ \frac{\partial u_k}{\partial x_k} (s, Z^L_s) \mathbb{E} \left[ \int_{\eta(s)}^{t} \left[ \frac{\partial B_k}{\partial u} (u, Z^L_u) \right] dw + \sum_{i=1}^{d} \frac{\partial B_k}{\partial x_i} (u, Z^L_u) B_i (Z^L_{\eta(u)}, V_{\eta(u)}) \right] du \right].
\]

Condition (\text{v-diff-Reg+}) and the conditional Jensen inequality imply that

\[
\mathbb{E} \left\| R_1(s) \right\| \leq c \sum_{k=1}^{d} \left( \int_{\eta(s)}^{t} \left[ \frac{\partial B_k}{\partial u} (u, Z^L_u) \right] du + \sum_{i=1}^{d} \frac{\partial B_k}{\partial x_i} (u, Z^L_u) B_i (Z^L_{\eta(u)}, V_{\eta(u)}) \right) + \frac{1}{2} \sum_{i,j=1}^{d} \frac{\partial^2 B_k}{\partial x_i \partial x_j} (u, Z^L_u) a_{ij} (Z^L_{\eta(u)}, V_{\eta(u)}) \right) du \right). \tag{2.23}
\]

By assumption (\text{V-Lip}),

\[
\left| B_i (Z^L_{\eta(u)}, V_{\eta(u)}) \right| \leq c \left( 1 + \left| Z^L_{\eta(u)} \right| + \int_{\mathbb{R}^d} \left| x \right| V_{\eta(u)}(dx) \right)
\]

and

\[
\left| a_{ij} (Z^L_{\eta(u)}, V_{\eta(u)}) \right| \leq c \left( 1 + \left| Z^L_{\eta(u)} \right|^2 + \int_{\mathbb{R}^d} \left| x \right|^2 V_{\eta(u)}(dx) \right).
\]

Using these two bounds along with Lemma 2.5, we can see that

\[
\mathbb{E} \left\| R_1(s) \right\| \leq c \left( \int_{\eta(s)}^{t} \left( 1 + \left| Z^L_{\eta(u)} \right|^2 + \int_{\mathbb{R}^d} \left| x \right|^2 V_{\eta(u)}(dx) \right) du \right) + c \left( \int_{\eta(s)}^{t} 1 + \sup_{s' \in [0, t]} \mathbb{E} |Z^L_{s'}|^2 + \sup_{s' \in [0, t]} \mathbb{E} \int_{\mathbb{R}^d} \left| x \right|^2 V_{s'}(dx) \right) du \right).
\]

Assumptions (\text{Lip}), (\text{\( \mu_0 \)-Lip}) and (\text{\( \nu \)-bound}) allow us to conclude that

\[
\sup_{0 \leq s \leq t} \mathbb{E} \left\| R_1(s) \right\| \leq c h_L.
\]

21
Error $R_2$: Condition (v-diff-Reg$^+$) implies that

$$
|\mathbb{E}[R_2(s)|\mathcal{F}_T^Y]| = \mathbb{E}\left[\sum_{j=1}^{d} \frac{\partial^2 V}{\partial x_j^2}(s, Z^L_s) \left(B_j(\eta(s), Z^L_{\eta(s)}) - \bar{B}_j(Z^L_{\eta(s)}, \mathcal{V}_{\eta(s)})\right)|\mathcal{F}_T^Y]\right] \\
\leq c \mathbb{E}\left|B(\eta(s), Z^L_{\eta(s)}) - \bar{B}(Z^L_{\eta(s)}, \mathcal{V}_{\eta(s)})\right| |\mathcal{F}_T^Y|. \quad (2.24)
$$

Using the notation of regular conditional probability measures,

$$
\int_0^t \mathbb{E}|\mathbb{E}[R_2(s)|\mathcal{F}_T^Y]| \, ds \leq c \int_0^t \mathbb{E}\left[\int_{\mathbb{R}^d} |\mathbb{E}[b(x, X_{\eta(s)})] - \bar{B}(x, \mathcal{V}_{\eta(s)})| \mu_{\eta(s)}(dx)\right] |\mathcal{F}_T^Y| \, ds.
$$

Similarly, by the condition on the second-order derivatives from (v-diff-Reg$^+$), we can establish that

$$
\sup_{0 \leq s \leq T} \mathbb{E}|\mathbb{E}[R_3(s)|\mathcal{F}_T^Y]| \leq c h_L \quad (2.25)
$$

and

$$
|\mathbb{E}[R_4(s)|\mathcal{F}_T^Y]| \leq c \mathbb{E}\left[\left\|\Sigma(\eta(s), Z^L_{\eta(s)}) - \bar{\Sigma}(Z^L_{\eta(s), \mathcal{V}_{\eta(s)})}\right\| |\mathcal{F}_T^Y|. \quad (2.26)
$$

In the previous theorem, we assumed that the function $G$ is in $C^2_b$. For the rest of the article, we relax the regularity of $G$ to $C^2_p$. Therefore, by Lemma (2.6), we only get condition (v-diff-Reg). We work with this weaker assumption from now on.

Since $\bar{Z}^L_{\eta(s)}$ is a random measure, it is undesirable for our iteration of the MLMC scheme. We therefore introduce an artificial process $\bar{Z}^L$ in order to remove the dependence of $Z^L$ on $\mathcal{F}_T^Y$.

**Lemma 2.8.** Let $G \in C^2_p(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R})$ be a globally Lipschitz continuous function. Assume that (Ker-Reg), ($\mu_0$-L$^p$), (V$^\perp$ (W, Z$^0$)), (V-bound) and (V-Lip) hold. Then there exists a constant $c$ such that for each $t \in [0, T]$ and $x \in \mathbb{R}^d$,

$$
\sup_{0 \leq s \leq t} \mathbb{E}\left[|\mathbb{E}[G(x, Z^L_s)|\mathcal{F}_T^Y] - \mathbb{E}[G(x, X_s)]|^2\right] \\
\leq c \left(h_L^2 + \int_0^t \mathbb{E}\left[\bar{B}(x, \mathcal{V}_{\eta(s)}) - \mathbb{E}[b(x, X_{\eta(s)})]| \mu_{\eta(s)}^Z(dx)\right] ds \\
+ \int_0^t \mathbb{E}\left[\bar{\Sigma}(x, \mathcal{V}_{\eta(s)}) - \mathbb{E}[\sigma(x, X_{\eta(s)})]| \mu_{\eta(s)}^Z(dx)\right] ds\right),
$$

where $\bar{Z}^L$ is a process defined by

$$
d\bar{Z}^L_t = \int_{\mathbb{R}^d} b(\bar{Z}^L_{\eta(t), y}) \mu_{\eta(t)}^X(dy) \, dt + \int_{\mathbb{R}^d} \sigma(\bar{Z}^L_{\eta(t), y}) \mu_{\eta(t)}^X(dy) \, dW_t.
$$
Proof. As in the proof of Theorem 2.7, we use \( \eta(s) \) to denote \( \eta_L(s) \) in all instances. Although the regularity of function \( v_y \) is weakened from (v-diff-Reg+) to (v-diff-Reg), the proof of Theorem 2.7 is still applicable: the stochastic integrand of (2.22) remains square integrable, by Lemma 2.5. Similarly, we establish an analogous estimate of (2.23) as follows. By the Cauchy-Schwarz and Jensen’s inequalities,

\[
E \left[ \left| E[R_1(s) \mathcal{F}_T^V] \right|^2 \right] \\
\leq c \sum_{k=1}^{d} \left[ \frac{\partial v_y}{\partial x_k}(s, Z^L_s) \sum_{i=1}^{d} \frac{\partial B_k(u, Z^L_u)}{\partial u} + \sum_{i=1}^{d} \frac{\partial B_k(u, Z^L_u) \bar{B}_i(Z^L_{\eta(u)}, \mathcal{V}_{\eta(u)})}{\partial x_i} \right] \left( \sigma(\mathcal{F}^{Z^L}_{s_T}, \mathcal{F}^V_{s_T}) \right)^2 \\
+ \frac{1}{2} \sum_{i,j=1}^{d} \frac{\partial^2 B_k}{\partial x_i \partial x_j}(u, Z^L_u) a_{ij} \left( Z^L_{\eta(u)}, \mathcal{V}_{\eta(u)} \right) \left( \sigma(\mathcal{F}^{Z^L}_{s_T}, \mathcal{F}^V_{s_T}) \right)^2, \tag{2.27}
\]

which implies the bound

\[
\sup_{0 \leq s \leq t} E \left[ \left| E[R_1(s) \mathcal{F}_T^V] \right|^2 \right] \leq ch^2_L.
\]

Similarly, by the condition on the second-order derivatives from (v-diff-Reg), we can also establish an analogous bound for (2.25), which becomes

\[
\sup_{0 \leq s \leq t} E \left[ \left| E[R_3(s) \mathcal{F}_T^V] \right|^2 \right] \leq ch^2_L.
\]

Thus, the proof of Theorem 2.7 implies that

\[
\sup_{0 \leq s \leq t} E \left[ \left| E[G(x, Z^L_s) \mathcal{F}_T^V] - E[G(x, X_s)] \right|^2 \right] \leq c \left( h^2_L + \int_0^t E \left( \left| E[R_2(s) \mathcal{F}_T^V] \right|^2 \right) + E \left( \left| E[R_4(s) \mathcal{F}_T^V] \right|^2 \right) ds \right).
\]

In a similar way as in (2.27), we can prove the analogous bounds for (2.24) and (2.26) to be

\[
\left\{ \begin{align*}
\left| E[R_2(s) \mathcal{F}_T^V] \right|^2 & \leq c E \left[ \left| B(\eta(s), Z^L_{\eta(s)}) - \bar{B}(Z^L_{\eta(s)}, \mathcal{V}_{\eta(s)}) \right|^2 \left| \mathcal{F}_T^V \right| \right], \\
\left| E[R_4(s) \mathcal{F}_T^V] \right|^2 & \leq c E \left[ \left| \Sigma(\eta(s), Z^L_{\eta(s)}) - \bar{\Sigma}(Z^L_{\eta(s)}, \mathcal{V}_{\eta(s)}) \right|^2 \left| \mathcal{F}_T^V \right| \right].
\end{align*} \right.
\]

Since we wish to obtain independence with the \( \sigma \)-algebra \( \mathcal{F}_T^V \) by replacing the process \( Z^L \) by \( \bar{Z}^L \), it therefore suffices to estimate the difference (by adding and subtracting the term \( B(\eta(s), Z^L_{\eta(s)}) - \bar{B}(Z^L_{\eta(s)}, \mathcal{V}_{\eta(s)}) \))

\[
E \left[ \left| (B(\eta(s), Z^L_{\eta(s)}) - \bar{B}(Z^L_{\eta(s)}, \mathcal{V}_{\eta(s)}) - (B(\eta(s), \bar{Z}^L_{\eta(s)}) - \bar{B}(\bar{Z}^L_{\eta(s)}, \mathcal{V}_{\eta(s)}) \right|^2 \left| \mathcal{F}_T^V \right| \right].
\]

23
By (Lip) and (V-Lip),
\[
\mathbb{E}\left[\left| (B(\eta(s), Z_{\eta(s)}^L) - \bar{B}(Z_{\eta(s)}^L, \mathcal{V}_{\eta(s)}) - (B(\eta(s), \bar{Z}_{\eta(s)}^L) - \bar{B}(\bar{Z}_{\eta(s)}^L, \mathcal{V}_{\eta(s)}) \right|^2 \right| \mathcal{F}_T^V \\
\leq c \mathbb{E}\left[ |Z_{\eta(s)}^L - \bar{Z}_{\eta(s)}^L|^2 \right| \mathcal{F}_T^V].
\]

We further decompose the error as follows.
\[
\mathbb{E}\left[ |Z_{\eta(s)}^L - \bar{Z}_{\eta(s)}^L|^2 \right| \mathcal{F}_T^V \leq 2 \left( \mathbb{E}\left[ \left| \int_0^s \left( B(\eta(u), Z_{\eta(u)}^L) - \bar{B}(Z_{\eta(u)}^L, \mathcal{V}_{\eta(u)}) \right) \, du \right|^2 \right| \mathcal{F}_T^V \right] \\
+ \mathbb{E}\left[ \left| \int_0^s \left( \Sigma(\eta(u), \bar{Z}_{\eta(u)}^L) - \Sigma(Z_{\eta(u)}^L, \mathcal{V}_{\eta(u)}) \right) \, dW_u \right|^2 \right| \mathcal{F}_T^V \right) \\
=: 2(R_{21}(s) + R_{22}(s)).
\]

By the conditional Fubini’s theorem and the Cauchy-Schwarz inequality, there exists a constant $K > 0$ such that
\[
R_{21}(s) \leq c \left( \int_0^s \mathbb{E}\left[ \left| B(\eta(u), Z_{\eta(u)}^L) - \bar{B}(Z_{\eta(u)}^L, \mathcal{V}_{\eta(u)}) \right|^2 \right| \mathcal{F}_T^V \right] \\
+ \mathbb{E}\left[ \left| \int_0^s \left( \Sigma(\eta(u), \bar{Z}_{\eta(u)}^L) - \Sigma(Z_{\eta(u)}^L, \mathcal{V}_{\eta(u)}) \right) \, dW_u \right|^2 \right| \mathcal{F}_T^V \right) \, du,
\]
where assumption (V-Lip) is used in the final inequality. Since $\bar{Z}^L$ is independent of $\mathcal{F}_T^V$ and that $\mu^X_{\eta(u)}$ is a non-random measure, we use the properties of regular conditional distributions as outlined in Lemma A.2 to prove that for each $\omega \in \Omega$,
\[
\mathbb{E}\left[ \left| B(\eta(u), Z_{\eta(u)}^L) - \bar{B}(Z_{\eta(u)}^L, \mathcal{V}_{\eta(u)}) \right|^2 \right| \mathcal{F}_T^V \right) \right) (\omega) \\
= \int_{\mathbb{R}^d} \left| B(\eta(u), x) - \bar{B}(x, \mathcal{V}_{\eta(u)}(\omega)) \right|^2 \mu^X_{\eta(u)}(dx).
\]
Therefore,
\[
R_{21}(s) \leq c \left( \int_0^s \mathbb{E}\left[ \left| Z_{\eta(u)}^L - \bar{Z}_{\eta(u)}^L \right|^2 \right| \mathcal{F}_T^V \right] \right) + \int_{\mathbb{R}^d} \left| B(\eta(u), x) - \bar{B}(x, \mathcal{V}_{\eta(u)}) \right|^2 \mu^X_{\eta(u)}(dx) \, du.
\]
We proceed similarly as $R_{22}(s)$ and apply part (b) of Lemma A.1 (with the filtration $\{\mathcal{F}_t\}_{t \in [0,T]}$ such that $\mathcal{F}_t = \sigma(F_t^V, \{W_u\}_{0 \leq u \leq t}, Z_0^L)$) to get
\[
R_{22}(s) \leq c \left( \int_0^s \mathbb{E}\left[ \left| Z_{\eta(u)}^L - \bar{Z}_{\eta(u)}^L \right|^2 \right| \mathcal{F}_T^V \right] \right) + \int_{\mathbb{R}^d} \left| \Sigma(\eta(u), x) - \Sigma(x, \mathcal{V}_{\eta(u)}) \right|^2 \mu^X_{\eta(u)}(dx) \, du.
\]

24
Combining both bounds gives

\[
\mathbb{E}[|Z_{\eta(s)}^L - \bar{Z}_{\eta(s)}^L|^2 | \mathcal{F}_T^s] \leq c \left( \int_0^s \mathbb{E}[|Z_{\eta(u)}^L - \bar{Z}_{\eta(u)}^L|^2 | \mathcal{F}_T^u] + \int_{\mathbb{R}^d} \left| B(\eta(u), x) - \bar{B}(x, \nu_{\eta(u)}) \right|^2 \mu_{\eta(u)}^L(dx) \right)
+ \int_{\mathbb{R}^d} \left| \Sigma(\eta(u), x) - \Sigma(x, \nu_{\eta(u)}) \right|^2 \mu_{\eta(u)}^L(dx) du \),
\]

for any \( s \in [0, t] \). By Gronwall’s lemma,

\[
\mathbb{E}[|Z_{\eta(s)}^L - \bar{Z}_{\eta(s)}^L|^2 | \mathcal{F}_T^s] \leq c \left( \int_0^s \left[ \int_{\mathbb{R}^d} \left| B(\eta(u), x) - \bar{B}(x, \nu_{\eta(u)}) \right|^2 \mu_{\eta(u)}^L(dx) \right] du \right)
+ \int_{\mathbb{R}^d} \left| \Sigma(\eta(s), x) - \Sigma(x, \nu_{\eta(s)}) \right|^2 \mu_{\eta(s)}^L(dx) du \),
\]

for any \( s \in [0, t] \). Integrating \( s \) from 0 to \( t \) gives

\[
\int_0^t \mathbb{E}[|Z_{\eta(s)}^L - \bar{Z}_{\eta(s)}^L|^2 | \mathcal{F}_T^s] ds \leq c \left( \int_0^t \left[ \int_{\mathbb{R}^d} \left| B(\eta(s), x) - \bar{B}(x, \nu_{\eta(s)}) \right|^2 \mu_{\eta(s)}^L(dx) \right] ds \right)
+ \int_{\mathbb{R}^d} \left| \Sigma(\eta(s), x) - \Sigma(x, \nu_{\eta(s)}) \right|^2 \mu_{\eta(s)}^L(dx) ds \).
\]

By (2.24) and (2.28), it is clear that

\[
\int_0^t \left| \mathbb{E}[R_2(s), \mathcal{F}_T^s] \right|^2 ds \leq c \left( \int_0^t \mathbb{E}[|Z_{\eta(s)}^L - \bar{Z}_{\eta(s)}^L|^2 | \mathcal{F}_T^s] + \mathbb{E}[\left| B(\eta(s), \bar{Z}_{\eta(s)}^L) - \bar{B}(\bar{Z}_{\eta(s)}^L, \nu_{\eta(s)}) \right|^2 | \mathcal{F}_T^s] \right) ds \).
\]

We note by the above that

\[
\mathbb{E}[|B(\eta(s), \bar{Z}_{\eta(s)}^L) - \bar{B}(\bar{Z}_{\eta(s)}^L, \nu_{\eta(s)})|^2 | \mathcal{F}_T^s] \leq \int_{\mathbb{R}^d} \left| B(\eta(s), x) - \bar{B}(x, \nu_{\eta(s)}) \right|^2 \mu_{\eta(s)}^L(dx).
\]

This shows that

\[
\int_0^t \left| \mathbb{E}[R_2(s), \mathcal{F}_T^s] \right|^2 ds \leq c \left( \int_0^t \left[ \int_{\mathbb{R}^d} \left| B(\eta(s), x) - \bar{B}(x, \nu_{\eta(s)}) \right|^2 \mu_{\eta(s)}^L(dx) \right] ds \right)
+ \int_{\mathbb{R}^d} \left| \Sigma(\eta(s), x) - \Sigma(x, \nu_{\eta(s)}) \right|^2 \mu_{\eta(s)}^L(dx) ds \).
\]

We repeat the same argument for \( R_4(s) \) and conclude that

\[
\int_0^t \left| \mathbb{E}[R_4(s), \mathcal{F}_T^s] \right|^2 ds \leq c \left( \int_0^t \left[ \int_{\mathbb{R}^d} \left| B(\eta(s), x) - \bar{B}(x, \nu_{\eta(s)}) \right|^2 \mu_{\eta(s)}^L(dx) \right] ds \right)
+ \int_{\mathbb{R}^d} \left| \Sigma(\eta(s), x) - \Sigma(x, \nu_{\eta(s)}) \right|^2 \mu_{\eta(s)}^L(dx) ds \).
\]

\( \square \)
3 Iteration of the MLMC algorithm

3.1 Interacting kernels

In order to perform MLMC iteration, we fix a Picard step $m \geq 1$ and correspond each particle $Z^{i,\ell}$ in the abstract framework to the particle $Y^{i,m,\ell}$ in the framework of MLMC, since the abstract framework corresponds to one Picard iteration. $\mathcal{F}_{m-1}^V$ corresponds to the sigma-algebra $\mathcal{F}_{m-1}$ generated by all the particles $Y^{i,m-1,\ell}$ in the $(m-1)$th Picard step, $0 \leq \ell \leq L, 1 \leq i \leq N_{m-1,\ell}$. Therefore, we set $V_i = \mathcal{V}_i^{m-1}$ and $\kappa = 2L + 1$, where

$$
\mathcal{V}_i^k := \left( \mu_t^{Y_{k,0},N_{k,0},0}, \mu_t^{Y_{k,1},N_{k,1}}, \mu_t^{Y_{k,1},N_{k,1}}, \mu_t^{Y_{k,1},N_{k,1}}, \ldots, \mu_t^{Y_{k,L-1},N_{k,L}}, \mu_t^{Y_{k,L-1},N_{k,L}} \right),
$$

(3.1)

for any Picard step $k$. We also define

$$
\bar{b}(x, \mathcal{V}_i) := \int_{\mathbb{R}^d} b(x, y) \mathcal{V}_i^{(1)}(dy) + \sum_{k=1}^L \left( \int_{\mathbb{R}^d} b(x, y) \mathcal{V}_i^{(2k+1)}(dy) - \int_{\mathbb{R}^d} b(x, y) \mathcal{V}_i^{(2k)}(dy) \right)
$$

and

$$
\bar{\sigma}(x, \mathcal{V}_i) := \int_{\mathbb{R}^d} \sigma(x, y) \mathcal{V}_i^{(1)}(dy) + \sum_{k=1}^L \left( \int_{\mathbb{R}^d} \sigma(x, y) \mathcal{V}_i^{(2k+1)}(dy) - \int_{\mathbb{R}^d} \sigma(x, y) \mathcal{V}_i^{(2k)}(dy) \right),
$$

for any $x \in \mathbb{R}^d$. Then, $\bar{b}$ and $\bar{\sigma}$ are functions corresponding respectively to the MLMC operator $\mathcal{M}_i^{(m-1)}$ applied to $b$ and $\sigma$, i.e.

$$
\bar{b}(x, \mathcal{V}_i) = \mathcal{M}_i^{(m-1)}(b(x, \cdot)) \quad \text{and} \quad \bar{\sigma}(x, \mathcal{V}_i) = \mathcal{M}_i^{(m-1)}(\sigma(x, \cdot)),
$$

for any $x \in \mathbb{R}^d$. The measure $\mathcal{V}_i$ satisfies the criterion $(V \perp (W, Z_0))$, since $\{Y^{m-1}\} \perp (W^m, Z_0^m)$. The criteria $(V\text{-bound})$, $(V\text{-Reg})$ and $(V\text{-Lip})$ are verified below.

Lemma 3.1 (Verification of $(V\text{-Lip})$). Assume $(\text{Lip})$ and $(\mu_0\cdot L_p)$ . Let $\mathcal{V}_i$ be defined as in (3.1). Then, for each $t \in [0, T]$, there exists a constant $c$ such that

$$
|\bar{b}(x_1, \mathcal{V}_i) - \bar{b}(x_2, \mathcal{V}_i)| + \|\bar{\sigma}(x_1, \mathcal{V}_i) - \bar{\sigma}(x_2, \mathcal{V}_i)\| \leq c|x_1 - x_2|, \quad \forall x_1, x_2 \in \mathbb{R}^d,
$$

$$
|\bar{b}(x, \mathcal{V}_{\eta(t)})| + \|\bar{\sigma}(x, \mathcal{V}_i)\| \leq c \left( 1 + |x| + \left| \int_{\mathbb{R}^d} |y| \mathcal{V}_i(dy) \right| \right), \quad \forall x \in \mathbb{R}^d.
$$

Proof. For any $t \in [0, T]$ and $x_1, x_2 \in \mathbb{R}^d$, we apply the definition of $\mathcal{M}_i^{(m-1)}$ and compute
that

\[
|\tilde{B}(x_1, V_l) - \tilde{B}(x_2, V_l)| = |\mathcal{M}^{(m-1)}_t(b(x_1, \cdot)) - \mathcal{M}^{(m-1)}_t(b(x_2, \cdot))|
\]

\[
= \sum_{l=1}^L \frac{1}{N_{m-1,\ell}} \sum_{i=1}^{N_{m-1,\ell}} \left[ \left( \frac{t - \eta_{\ell}(t)}{h_{\ell}} \right) \cdot \left( b(x_1, \gamma^{i,m-1,\ell,1}_{\eta_{\ell}(t)+h_{\ell}}) - b(x_2, \gamma^{i,m-1,\ell,1}_{\eta_{\ell}(t)+h_{\ell}}) \right) \\
+ \left( 1 - \frac{t - \eta_{\ell}(t)}{h_{\ell}} \right) \cdot \left( b(x_1, \gamma^{i,m-1,\ell,0}_{\eta_{\ell}(t)+h_{\ell}}) - b(x_2, \gamma^{i,m-1,\ell,0}_{\eta_{\ell}(t)+h_{\ell}}) \right) \\
- \left( \frac{t - \eta_{\ell-1}(t)}{h_{\ell-1}} \right) \cdot \left( b(x_1, \gamma^{i,m-1,\ell-1}_{\eta_{\ell-1}(t)+h_{\ell-1}}) - b(x_2, \gamma^{i,m-1,\ell-1}_{\eta_{\ell-1}(t)+h_{\ell-1}}) \right) \\
- \left( 1 - \frac{t - \eta_{\ell-1}(t)}{h_{\ell-1}} \right) \cdot \left( b(x_1, \gamma^{i,m-1,\ell-1}_{\eta_{\ell-1}(t)}) - b(x_2, \gamma^{i,m-1,\ell-1}_{\eta_{\ell-1}(t)}) \right) \right].
\]

Assumption (Lip) implies that

\[
|\tilde{B}(x_1, V_{\eta(t)}) - \tilde{B}(x_2, V_{\eta(t)})| \leq c|x_1 - x_2| \quad \text{and} \quad |\tilde{B}(x_1, V_{\eta(t)})| \leq c\left(1 + |x_1| + \left| \int_{\mathbb{R}^d} |y| V_{\eta(t)}(dy) \right| \right).
\]

We can estimate \(\|\Sigma(x_1, V_{\eta(t)}) - \Sigma(x_2, V_{\eta(t)})\|\) and \(\|\Sigma(x_1, V_{\eta(t)})\|\) in a similar way, which concludes the proof this lemma. \(\Box\)

**Lemma 3.2** (Verification of \((V\text{-}bound})\). Assume (Lip) and \((\mu_0, L_p)\). Then for any \(p \geq 2\), there exists a constant \(c\) such that

\[
\sup_{k \in [\mathbb{N}\cup\{0\}]} \sup_{t \in [0, T]} \mathbb{E} \left| \int_{\mathbb{R}^d} |x|^p V_t^k(dx) \right| \leq c.
\]

**Proof.** We note that the integral can be rewritten as

\[
\left| \int_{\mathbb{R}^d} |x|^p V_t^k(dx) \right| = \left| \frac{1}{N_{k,0}} \sum_{i=1}^{N_{k,0}} P_t^{i,0} \right| + \sum_{l=1}^L \left[ \left| \frac{1}{N_{k,\ell}} \sum_{i=1}^{N_{k,\ell}} P_t^{i,\ell} \right| + \left| \frac{1}{N_{k,\ell}} \sum_{i=1}^{N_{k,\ell}} P_t^{i,\ell-1} \right| \right],
\]

(3.2)

where

\[
P_t^{i,\ell} = \left( \frac{t - \eta_{\ell}(t)}{h_{\ell}} \right) |\gamma^{i,k,\ell}_{\eta_{\ell}(t)+h_{\ell}}|^p + \left( 1 - \frac{t - \eta_{\ell}(t)}{h_{\ell}} \right) |\gamma^{i,k,\ell}_{\eta_{\ell}(t)}|^p.
\]

We fix \(\ell \in \{0, \ldots, L\}\). By exchangeability, there exists a constant \(D\) (independent of the Picard step \(k\)) such that

\[
\mathbb{E}|P_t^{i,\ell}| \leq D(\mathbb{E}|Y_t^{1,k,\ell,p}|^p + \mathbb{E}|Y_t^{1,k,\ell+1,p}|^p).
\]

27
By the triangle inequality,

\[ \mathbb{E} \left| \frac{1}{N_{k,\ell}} \sum_{i=1}^{N_{k,\ell}} P_{t}^{i,\ell} \right| \leq \sum_{i=1}^{N_{k,\ell}} \mathbb{E} |P_{t}^{i,\ell}| \leq D(\mathbb{E}|Y_{\eta(t)}^{1,k,\ell}|^p + \mathbb{E}|Y_{\eta(t)+h_{\ell}}^{1,k,\ell}|^p). \]

By (3.2), we can see that

\[ \mathbb{E} \left| \int_{\mathbb{R}^d} |x|^p \varphi_{t}^k (dx) \right| \leq 2D \sum_{t=0}^{L} \left( \mathbb{E}|Y_{\eta(t)}^{1,k,\ell}|^p + \mathbb{E}|Y_{\eta(t)+h_{\ell}}^{1,k,\ell}|^p \right). \]

By Lemma 2.1 and the proof of Lemma 3.1, there exists a constant \( R \) (which does not depend on the particular Picard step) such that

\[ \sup_{0 \leq t \leq T} \mathbb{E} \left| \int_{\mathbb{R}^d} |x|^p \varphi_{t}^k (dx) \right| \leq R \left( 1 + \int_{0}^{T} \sup_{0 \leq u \leq s} \mathbb{E} \left| \int_{\mathbb{R}^d} |x|^p \varphi_{u}^{k-1} (dx) \right| ds \right). \] (3.3)

By iteration, we conclude that

\[ \sup_{0 \leq t \leq T} \mathbb{E} \left| \int_{\mathbb{R}^d} |x|^p \varphi_{t}^k (dx) \right| \leq (RT)^r \frac{k!}{r!} + \sup_{0 \leq t \leq T} \mathbb{E} \left| \int_{\mathbb{R}^d} |x|^p \varphi_{t}^0 (dx) \right| \frac{(RT)^k}{k!} \]

\[ \leq e^{RT} \left( 1 + \sup_{0 \leq t \leq T} \mathbb{E} \left| \int_{\mathbb{R}^d} |x|^p \varphi_{t}^0 (dx) \right| \right) < +\infty. \]

\[ \square \]

**Lemma 3.3 (Verification of (\( V\)-Reg) ).** Assume (Lip) and \((\mu_0-L_p)\). Given any globally Lipschitz continuous function \( C^2_p \supseteq G : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R} \) and \( n \in \mathbb{N} \cup \{0\} \), there exists a constant \( c \) such that

\[ \mathbb{E} \left| \mathcal{M}_t^{(n)}(G(x,\cdot)) - \mathcal{M}_s^{(n)}(G(x,\cdot)) \right|^2 \leq c(t-s), \] (3.4)

for any \( x \in \mathbb{R}^d \) and \( 0 \leq s \leq t \leq T \).

**Proof.** Pick any \( \ell^* \in \{0, 1, 2, \ldots, L\} \). For simplicity of notation, we rewrite \( \mathcal{M}_t^{(n)}(G(x,\cdot)) \) as

\[ \mathcal{M}_t^{(n)}(G(x,\cdot)) := \frac{1}{N_{n,0}} \sum_{i=1}^{N_{n,0}} P_{t}^{i,0} + \sum_{\ell=1}^{L} \frac{1}{N_{n,\ell}} \sum_{i=1}^{N_{n,\ell}} \left( P_{t}^{i,\ell} - P_{t}^{i,\ell-1} \right), \] (3.5)

where

\[ P_{t}^{i,\ell} = \left( \frac{t - \eta(t)}{h_{\ell}} \right) G(x, Y_{\eta(t)+h_{\ell}}^{x,i,k,\ell}) + \left( 1 - \frac{t - \eta(t)}{h_{\ell}} \right) G(x, Y_{\eta(t)}^{x,i,k,\ell}). \]
Given any $k \in \{0, 1, \ldots, 2^L - 1\}$, we compute

$$\mathcal{M}_{t_{k+1}}^{(n)}(G(x, \cdot)) - \mathcal{M}_{t_k}^{(n)}(G(x, \cdot)) = \frac{1}{N_{n,0}} \sum_{i=1}^{N_{n,0}} (P_{t_{k+1}}^{i,0} - P_{t_k}^{i,0})$$

$$+ \sum_{\ell=1}^{L} \frac{1}{N_{n,\ell}} \sum_{i=1}^{N_{n,\ell}} \left( (P_{t_{k+1}}^{i,\ell} - P_{t_k}^{i,\ell}) + (P_{t_{k+1}}^{i,\ell-1} - P_{t_k}^{i,\ell-1}) \right).$$

Thus, we only need to consider $P_{t_{k+1}}^{i,\ell} - P_{t_k}^{i,\ell}$, for each $\ell \in \{0, 1, \ldots, L\}$. There are two cases depending on the value of $\ell$:

- For levels $\ell < \ell^*$, at least one of $P_{t_{k+1}}^{i,\ell}$ and $P_{t_k}^{i,\ell}$ is an interpolated value. Then there exist a unique $s \in \{0, 1, \ldots, 2^\ell - 1\}$ (chosen such that $\eta_\ell(t_{k+1}^s) = t_k^s$) and constants $\lambda \in (0, 1 - \frac{h_{\ell^*}}{h_\ell})$ and $\tilde{\lambda}$, given by

$$\lambda = \frac{t_k^s - t_\ell}{h_\ell} \quad \text{and} \quad \tilde{\lambda} = \frac{t_{k+1}^s - t_\ell}{h_\ell},$$

such that

$$P_{t_{k+1}^s}^{i,\ell} = (1 - \lambda)G(x, Y_{t_{k+1}^s}^{i,n,\ell}) + \lambda G(x, Y_{t_k^s}^{i,n,\ell}) \quad \text{and} \quad P_{t_k^s}^{i,\ell} = (1 - \tilde{\lambda})G(x, Y_{t_k^s}^{i,n,\ell}) + \tilde{\lambda} G(x, Y_{t_{k+1}^s}^{i,n,\ell}).$$

Note that $\tilde{\lambda} - \lambda = \frac{h_{\ell^*}}{h_\ell}$. By taking the difference between $P_{t_{k+1}}^{i,\ell}$ and $P_{t_k}^{i,\ell}$, we compute that

$$P_{t_{k+1}}^{i,\ell} - P_{t_k}^{i,\ell} = \frac{h_{\ell^*}}{h_\ell} (P_{t_{k+1}}^{i,\ell} - P_{t_k}^{i,\ell}) = \frac{h_{\ell^*}}{h_\ell} (G(x, Y_{t_{k+1}^s}^{i,n,\ell}) - G(x, Y_{t_k^s}^{i,n,\ell})). \tag{3.6}$$

- For levels $\ell \geq \ell^*$, both of them are not interpolated. This gives

$$P_{t_{k+1}}^{i,\ell} - P_{t_k}^{i,\ell} = G(x, Y_{t_{k+1}}^{i,n,\ell^*}) - G(x, Y_{t_k}^{i,n,\ell^*}). \tag{3.7}$$

By Lemmas 3.2 and 3.1, the hypotheses of Lemma 2.2 are satisfied. By applying Lemma 2.2 to (3.6) and (3.7) along with the global Lipschitz property of $G$, we have

$$E|P_{t_{k+1}}^{i,\ell} - P_{t_k}^{i,\ell}|^2 \leq ch_\ell^* \forall \ell \in \{0, 1, \ldots, L\}.$$

This shows that

$$E\left| \mathcal{M}_{t_{k+1}}^{(n)}(G(x, \cdot)) - \mathcal{M}_{t_k}^{(n)}(G(x, \cdot)) \right|^2$$

$$\leq \frac{1}{N_{n,0}} \sum_{i=1}^{N_{n,0}} E|P_{t_{k+1}}^{i,0} - P_{t_k}^{i,0}|^2 + \sum_{\ell=1}^{L} \frac{2}{N_{n,\ell}} \sum_{i=1}^{N_{n,\ell}} \left( E|P_{t_{k+1}}^{i,\ell} - P_{t_k}^{i,\ell}|^2 + E|P_{t_{k+1}}^{i,\ell-1} - P_{t_k}^{i,\ell-1}|^2 \right)$$

$$\leq ch_\ell^*.$$
The proof is complete by replacing \( s \) and \( t \) by \( \eta_L(s) \) and \( \eta_L(t) \) respectively if any of them (or both) does not belong to \( \Pi^L \).

Lemma 3.4 below gives a decomposition of MSE (mean-square-error) for MLMC along one iteration of the particle system (1.9).

**Lemma 3.4.** Assume (Ker-Reg) and \((\mu_0, L_p)\). Let \( G \in C^2_p(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R}) \) be a globally Lipschitz continuous function. We define the multilevel estimator \( \mathcal{M}_L^{(m)} \) as in (1.9). Let

\[
MSE^{(m)}_t(G(x, \cdot)) := \mathbb{E} \left[ \left( \mathbb{E}[G(x, X_t)] - \mathcal{M}_L^{(m)}(G(x, \cdot)) \right)^2 \right], \quad t \in [0, T].
\]

Then, for every \( t \in [0, T] \),

\[
\int_{\mathbb{R}^d} MSE^{(m)}_{\eta_L(t)}(G(x, \cdot)) \mu_{\eta_L(t)}^Z(dx) \leq c \left( h_T^2 + \int_0^t \left[ \sup_{x \in \mathbb{R}^d} \mathbb{E} \left| \mathcal{M}_{\eta_L(s)}^{(m-1)}(b(x, \cdot)) - \mathbb{E}[b(x, X_{\eta_L(s)})] \right|^2 \mu_{\eta_L(s)}^Z(dx) \right] ds + \sum_{\ell=0}^L \frac{h_T}{N_{m, \ell}} \right).
\]

Furthermore, if we assume that the functions \( b \) and \( \sigma \) are both bounded, then

\[
\sup_{x \in \mathbb{R}^d} MSE^{(m)}_{\eta_L(t)}(G(x, \cdot)) \leq c \left( h_T^2 + \int_0^t \left[ \sup_{x \in \mathbb{R}^d} \mathbb{E} \left| \mathcal{M}_{\eta_L(s)}^{(m-1)}(b(x, \cdot)) - \mathbb{E}[b(x, X_{\eta_L(s)})] \right|^2 \mu_{\eta_L(s)}^Z(dx) \right] ds + \sum_{\ell=0}^L \frac{h_T}{N_{m, \ell}} \right).
\]

**Proof.** For \( x \in \mathbb{R}^d \) and \( t \in [0, T] \), we consider

\[
\mathbb{E} \left[ \left( \mathbb{E}[G(x, X_{\eta_L(t)})] - \mathcal{M}_L^{(m)}(G(x, \cdot)) \right)^2 \right] = \mathbb{E} \left[ \left( \mathbb{E}[G(x, X_{\eta_L(t)})] - \mathbb{E}[\mathcal{M}_L^{(m)}(G(x, \cdot))|\mathcal{F}^{m-1}] \right)^2 \right. + \left. \mathbb{E}[\mathcal{M}_L^{(m)}(G(x, \cdot))|\mathcal{F}^{m-1}] - \mathcal{M}_L^{(m)}(G(x, \cdot)) \right)^2.
\]

Observe that

\[
MSE^{(m)}_{\eta_L(t)}(G(x, \cdot)) = \mathbb{E} \left[ \left( \mathbb{E}[G(x, X_{\eta_L(t)})] - \mathbb{E}[G(x, Y_{\eta_L(t)}^{1,m,L})|\mathcal{F}^{m-1}] \right)^2 \right. + \left. \mathbb{E}[\mathcal{M}_L^{(m)}(G(x, \cdot))|\mathcal{F}^{m-1}] - \mathcal{M}_L^{(m)}(G(x, \cdot)) \right)^2, (3.8)
\]

30
Combining (3.8), (3.9) and (3.10) yields the result.

Remark 3.5. The total variance is bounded by a constant (not sufficient for establishing a satisfactory upper bound of the mean-square-error in complexity analysis) due to interdependence of the processes, whereas the conditional variance is in the classical form of standard MLMC variance, i.e. $c \sum_{\ell=0}^{L} \frac{h_{\ell}}{N_{m,\ell}}$.

The complete algorithm consists of a sequence of nested MLMC estimators $\{M_{\ell}(G(x, \cdot))\}_{m=1, \ldots, M}$ and its error analysis is presented in Theorem 3.6. Note that we iterate the algorithm by replacing $G$ by the component real-valued functions $\{b_{1}\}_{1 \leq i \leq d}$ and $\{\sigma_{i,j}\}_{1 \leq i \leq d, 1 \leq j \leq r}$.

Theorem 3.6. Assume (Ker-Reg) and $(\mu \circ L_{\mu})$. Fix $M > 0$ and take $C_{\mu}^{2} \equiv P: \mathbb{R}^{d} \to \mathbb{R}$ to be a globally Lipschitz continuous function. As before, we define the mean-square error as

$$MSE_{t}^{(M)}(P) := \mathbb{E} \left[ (M_{t}^{(M)}(P) - \mathbb{E}[P(X_{t})])^{2} \right].$$

Then for every $t \in [0, T]$

$$MSE_{\eta_{L}(t)}^{(M)}(P) \leq c \left\{ h_{L}^{2} + \sum_{m=1}^{M} c^{M-m} \frac{1}{(M-m)!} \sum_{\ell=0}^{L} \frac{h_{\ell}}{N_{m,\ell}} + \frac{c^{M-1}}{M!} \right\}.$$

Proof. First, the assumption that $Y_{t}^{1,0,\ell} = X_{0}$ gives

$$\sup_{0 \leq t \leq T} \int_{\mathbb{R}^{d}} \mathbb{E} \left[ \left| b(x, X_{\eta_{L}(t)}) - M_{\eta_{L}(t)}^{(0)}(b(x, \cdot)) \right|^{2} + \left| \sigma(x, X_{\eta_{L}(t)}) - M_{\eta_{L}(t)}^{(0)}(\sigma(x, \cdot)) \right|^{2} \right] \mu_{\eta_{L}(s)}(dx) \leq c.$$  

(3.11)
Fixing $M > 0$ and a globally Lipschitz continuous function $C_p^2 \ni P : \mathbb{R}^d \to \mathbb{R}$, we set

$$
a_t^{(m)} := \begin{cases}
\mathbb{E} \left( (\mathcal{M}_{\eta_L(t)}^{(m)}(P) - \mathbb{E}[P(X_{\eta_L(t)})])^2 \right), & m = M, \\
\int_{\mathbb{R}^d} \mathbb{E} \left[ \left| \mathcal{M}_{\eta_L(t)}^{(m-1)}(b(x, \cdot)) - \mathbb{E}[b(x, X_{\eta_L(t)})] \right|^2 \\
+ \left\| \mathcal{M}_{\eta_L(t)}^{(m-1)}(\sigma(x, \cdot)) - \mathbb{E}[\sigma(x, X_{\eta_L(t)})] \right\|^2 \right] \mu_{\eta_L(t)}^2(dx), & m \leq M - 1.
\end{cases}
$$

(3.12)

From Lemma 3.4, we observe that

$$a_t^{(m)} \leq c \left( b^{(m)} + \int_0^1 a_s^{(m-1)} ds \right), \quad \forall m \in \{1, 2, \ldots, M\},
$$

(3.13)

where $b^{(m)} = h_L^2 + \sum_{\ell=0}^L \frac{h_{\ell}}{N_{m, \ell}}$. Then one can easily show that

$$\sup_{0 \leq t \leq T} a_t^M \leq \sum_{m=0}^{M-1} b^{(M-m)} \frac{(cT)^m}{m!} + \left( \sup_{0 \leq s \leq T} a_s^{(0)} \right) \cdot \frac{(cT)^M}{M!}.
$$

(3.14)

Inequalities (3.11) and (3.14) conclude the proof.

We are now in a position to present the complexity theorem for iterated MLMC estimators of $\{\mathbb{E}[P(X_{\eta_L(t)})]\}_{t \in [0, T]}$.

**Theorem 3.7.** Assume (Ker-Reg) and ($\mu_{\ol{C}}L_p$) : Fix $M > 0$ and let $C_p^2 \ni P : \mathbb{R}^d \to \mathbb{R}$ be a globally Lipschitz continuous function. Then for any $\epsilon < e^{-1}$, there exist $M$, $\{L_m\}_{1 \leq m \leq M}$ and $\{N_{m, \ell}\}_{1 \leq m \leq M, 0 \leq \ell \leq L_m}$ such that for every $t \in [0, T]$,

$$MSE_{\eta_L(t)}^{(M)}(P) := \mathbb{E} \left[ (\mathcal{M}_{\eta_L(t)}^{(M)}(P) - \mathbb{E}[P(X_{\eta_L(t)})])^2 \right] < \epsilon^2,
$$

and computational complexity $C$ is of order $\epsilon^{-4} |\log \epsilon|^3$.

**Proof.** To achieve the desired computational complexity, we will show that the mean-square-error of MLMC estimator should vary with $m$. We will also argue that there would be no gain (asymptotically) by letting levels $L$ vary with $m$ as well. To show this, we assume that indeed we have $L_1 \geq \ldots \geq L_m \geq \ldots \geq L_M$ and that an extension of Theorem 3.6 holds with

$$\sup_{0 \leq t \leq T} \mathbb{E} \left[ (\mathcal{M}_{\eta_L(t)}^{(M)}(P) - \mathbb{E}[P(X_{\eta_L(t)})])^2 \right] \leq c \left\{ \sum_{m=1}^{M} \frac{c^{M-m}}{(M-m)!} \cdot (h_L^2 + \sum_{\ell=0}^{L_m} \frac{h_{\ell}}{N_{m, \ell}} + \frac{c^{M-1}}{M!}) \right\}.
$$

(3.15)
In the end, we will show that in fact this extension is not needed, though this can be easily proved. The cost of obtaining $M_{\nu_{\text{ext}}}(t)(P)$ involves doing $M$ Picard iterations in total, such that for each Picard step, we perform the standard MLMC algorithm, but with an additional cost of $\sum_{l'=0}^{L_{m-1}} N_{m-1,l'}$ computations for each MLMC mean-field approximation after the first Picard iteration.

This induces the following optimisation problem \(^2\)

\[
\min_{M,\{L_m\}_{m=1}^{M},\{N_{m,l}\}_{0 \leq l \leq L_m}} C\left(M, \{L_m\}_{m=1}^{M}, \{N_{m,l}\}_{0 \leq l \leq L_m}\right) = \sum_{\ell=0}^{L_1} h_{\ell}^{-1} N_{1,\ell} + \sum_{m=2}^{M} \sum_{\ell=0}^{L_{m-1}} h_{\ell}^{-1} N_{m,\ell} \sum_{\ell'=0}^{L_{m-1}} N_{m-1,\ell'}
\]

such that
\[
\sum_{m=1}^{M} \frac{e^{M-m}}{(M-m)!} \cdot (h_{L_m}^2 + \sum_{\ell=0}^{L_m} h_{\ell} \frac{e^{M-1}}{N_{m,\ell}}) + \frac{e^{M-1}}{M!} \lesssim \epsilon^2.
\]

By the Stirling’s approximation, given by $\sqrt{2\pi n + \frac{1}{2}} e^{-n} \sim n!$, we obtain that
\[
\frac{e^{M}}{M!} \lesssim \frac{e^{M}}{\sqrt{2\pi (M+\frac{1}{2})} \epsilon^2} \implies M^* \lesssim \log(\epsilon^{-1}).
\]

We consider a sequence $\{w_m\}_{m=1}^{M^*}$ and define $\epsilon_m^2 := w_m \cdot \epsilon^2$ (for concrete examples see remark 3.9) such that
\[
\epsilon^2 \sim \sum_{m=1}^{M^*} \frac{e^{M^*-m}}{(M^*-m)!} \epsilon_m^2.
\]

To simplify the analysis we impose the following conditions:

(I) minimum condition: For each $m$, $w_m \geq w_{M^*} = 1$;

(II) weight condition: $\sum_{m=1}^{M^*} \frac{e^{M^*-m}}{(M^*-m)!} w_m \sim 1$;

(III) cost condition: $\sum_{m=2}^{M^*-1} w_m^{-1} \cdot w_{m-1}^{-1} \sim 1$.

Representation (3.19) shows that to solve (4.3)-(4.4), it is enough to consider $M^*$ optimisation problems as follows: for $m = 1$,

\[
\min_{L_1,\{N_{1,l}\}_{0 \leq l \leq L_1}} C_1\left(L_1, \{N_{1,l}\}_{0 \leq l \leq L_1}\right) = \sum_{\ell=0}^{L_1} h_{\ell}^{-1} N_{1,\ell}
\]

such that
\[
h_{L_1}^2 + \sum_{\ell=0}^{L_1} h_{\ell} N_{1,\ell} \lesssim \epsilon_1^2.
\]

\(^2\)For convenience, we use the notation $x \lesssim y$ to denote that there exists a constant $c$ such that $x \leq c y.$
and for $2 \leq m \leq M^*$,
\[
\min_{L_m, \{N_m, \ell\}_{0 \leq \ell \leq L_m}} C_m \left( L_m, \{N_m, \ell\}_{0 \leq \ell \leq L_m} \right) = \sum_{\ell=0}^{L_m} \frac{h_{\ell}^{-1} N_{m, \ell}}{N_{m-1, \ell'}} \sum_{\ell'=0}^{L_{m-1}} N_{m-1, \ell'} 
\]
(3.22)
such that
\[
h_{L_m}^2 + \sum_{\ell=0}^{L_m} \frac{h_{\ell}^2}{N_{m, \ell}} \lesssim \epsilon_m^2. 
\]
(3.23)

From (3.21) and (3.23), along with the fact $h_{L_m} = T2^{-L_m}$, we have
\[
h_{L_m}^2 \lesssim \epsilon_m^2 \implies L_m^* \sim \log(\epsilon_m^{-1}), \quad m = 1, \ldots, M. 
\]
(3.24)

Next, we introduce a Lagrange multiplier $\lambda_1$ for (3.20) and treat all $N_1, \ell$ as continuous variables in the minimisation problem.
\[
\min_{\{N_1, \ell\}_{0 \leq \ell \leq L_1^*}} \phi_1(\{N_1, \ell\}_{0 \leq \ell \leq L_1^*}) = \sum_{\ell=0}^{L_1^*} h_{\ell}^{-1} N_{1, \ell} + \lambda_1 \left( \sum_{\ell=0}^{L_1^*} \frac{h_{\ell}^2}{N_{1, \ell}} - \epsilon_1^2 \right). 
\]
(3.25)

First order conditions show that
\[
N_{1, \ell}^* = \lambda_1^* h_{\ell}, \quad \ell = 0, \ldots, L_1^*. 
\]
(3.26)

Since we want to satisfy the constraint $\sum_{\ell=0}^{L_1^*} \frac{h_{\ell}^2}{N_{1, \ell}} \lesssim \epsilon_1^2$ (from (3.21)), we set
\[
\lambda_1^* \sim \epsilon_1^{-2}(L_1^* + 1), 
\]
(3.27)

which implies the optimal choice to be
\[
N_{1, \ell}^* \sim \epsilon_1^{-2}(L_1^* + 1) h_{\ell}, \quad \ell = 0, \ldots, L_1^*. 
\]
(3.28)

For $2 \leq m \leq M$, given samples $\{N_{m-1, \ell}\}_{0 \leq \ell \leq L_{m-1}^*}$, we introduce a Lagrange multiplier $\lambda_m$ for problem (3.22) as follows.
\[
\min_{\{N_m, \ell\}_{0 \leq \ell \leq L_m^*}} \phi_m(\{N_m, \ell\}_{0 \leq \ell \leq L_m^*}) = \sum_{\ell=0}^{L_m} h_{\ell}^{-1} N_{m, \ell} \sum_{\ell'=0}^{L_{m-1}} N_{m-1, \ell'} + \lambda_m \left( \sum_{\ell=0}^{L_m} \frac{h_{\ell}^2}{N_{m, \ell}} - \epsilon_m^2 \right). 
\]
(3.29)

First order conditions show that
\[
N_{m, \ell}^* = \left( \lambda_m^* \right)^{\frac{1}{2}} h_{\ell} \left( \sum_{\ell'=0}^{L_{m-1}} N_{m-1, \ell'}^* \right)^{-\frac{1}{2}}, \quad \ell = 0, \ldots, L_m^*. 
\]
(3.30)
By the same reasoning above, we set
\[
(\lambda_m^n)^{\frac{1}{2}} \sim \epsilon_m^{-2}(L_m^* + 1)\left( \sum_{\ell' = 0}^{L_{m-1}^*} N_{m-1,\ell'}^* \right)^{\frac{1}{2}},
\]  
which implies the optimal choice to be
\[
N_{m,\ell}^* \sim \epsilon_m^{-2}(L_m^* + 1)h_{\ell}, \quad \ell = 0, \ldots, L_m^*.
\]  
Thus, for \(1 \leq m \leq M^*\) and \(0 \leq \ell \leq L_m^*\),
\[
N_{m,\ell}^* \sim \epsilon_m^{-2}(L_m^* + 1)h_{\ell}.
\]  
Next, we calculate its asymptotic order of the cost at final step. Property (II) of \(\{w_m\}_{m=1}^{M^*}\) implies that \(w_{M^*} \sim w_{M^*-1} \sim 1\). Together with (3.24), we obtain that
\[
C_{M^*}(L_{M^*}, \{N_{m,\ell}^*\}_{0 \leq \ell \leq L_m^*}) = \epsilon_m^{-2} \epsilon_{M^*-1}^{-2} (L_{M^*}^* + 1)^2 (L_{M^*-1}^* + 1) \sim \epsilon^{-4} |\log(\epsilon)|^3. \tag{3.34}
\]  
(3.34) implies that one should not expect the cost being less than \(\epsilon^{-4} |\log(\epsilon)|^3\). Therefore, the proof is complete if we can show that
\[
C\left(M^*, \{L_m^*\}_{m=1}^{M^*}, \{N_{m,\ell}^*\}_{1 \leq m \leq M^* \atop 0 \leq \ell \leq L_m^*} \right) \lesssim \epsilon^{-4} |\log(\epsilon)|^3. \tag{3.35}
\]  
Now we return to our total cost function \(C\) (see (4.3)), which consists of a cost of \(C_m\) for each Picard step, and compute its asymptotic order as follows.
\[
C\left(M^*, \{L_m^*\}_{m=1}^{M^*}, \{N_{m,\ell}^*\}_{1 \leq m \leq M^* \atop 0 \leq \ell \leq L_m^*} \right) \\
= \sum_{m=1}^{M^*} C_m(L_m^*, \{N_{m,\ell}^*\}_{0 \leq \ell \leq L_m^*}) \\
= \sum_{\ell=0}^{L_1^*} h_{\ell}^{-1} N_{1,\ell}^* + \sum_{m=2}^{M^*} \sum_{\ell=0}^{L_m^*} h_{\ell}^{-1} N_{m,\ell}^* \sum_{\ell' = 0}^{L_{m-1}^*} N_{m-1,\ell'}^* \\
= \epsilon_1^{-2}(L_1^* + 1)^2 + \sum_{m=2}^{M^*-1} \epsilon_m^{-2} \epsilon_{m-1}^{-2} (L_m^* + 1)^2 (L_{m-1}^* + 1) \\
+ \epsilon_{M^*}^{-2} \epsilon_{M^*-1}^{-2} (L_{M^*}^* + 1)^2 (L_{M^*-1}^* + 1). \tag{3.37}
\]  
Fixing \(L = \max_m \{L_m^*\}\) across all levels, (3.24) gives
\[
L_m^* \lesssim \log(\epsilon^{-1}), \quad m = 1, \ldots, M^*. \tag{3.39}
\]
By property (I) of \( \{ w_m \}_{m=1}^{M^*} \), together with (3.38) and (3.39), we obtain that
\[
C \lesssim |\log(\epsilon)|^2 \left[ \epsilon^{-2} + |\log(\epsilon)| \left( \sum_{m=2}^{M^*-1} \epsilon_m^{-2} \epsilon_{m-1}^{-2} + \epsilon^{-4} \right) \right]
\] (3.40)
\[
\lesssim \epsilon^{-4} |\log(\epsilon)|^3 \left( \sum_{m=2}^{M^*-1} w_m^{-1} w_{m-1}^{-1} + 1 \right).
\] (3.41)

From property (III) of \( \{ w_m \}_{m=1}^{M^*} \),
\[
\sum_{m=2}^{M^*-1} w_m^{-1} w_{m-1}^{-1} \sim 1,
\] (3.42)
which concludes the result by combining (3.41) and (3.42).

**Remark 3.8.** For MVSDs with additive noise, the variance can be controlled by \( h_\ell^2 \) instead of \( h_\ell \). Consequently, we are able to optimally set, for \( 1 \leq m \leq M^* \) and \( 0 \leq \ell \leq L^* \),
\[
N_{m,\ell}^* \sim \epsilon_m^{-2} h_\ell^\frac{2}{3}.
\] (3.43)
By the same computation from (3.36) to (3.42), the complexity is shown to be of the order \( \epsilon^{-4} \).

**Remark 3.9.** Here we show two possible choices of the sequence \( \{ w_m \}_{m=1}^{M^*} \) satisfying properties (I) to (III).

- \( w_m^{(1)} := \max \left( \left( \frac{(M^*-m)!}{c_{M^*-m}} \right)^{1-\alpha}, 1 \right) \), where \( \alpha \in (0, 1) \) or
- \( w_m^{(2)} := \max \left( \left( \frac{(M^*-m-2)!}{c_{M^*-m-2}} \right)^{1}, 1 \right) \) if \( m \leq M^* - 2 \) and \( w_m := 1 \) if \( M^* - 1 \leq m \leq M^* \).

In particular, for \( w_m^{(1)} \), if \( \alpha \) is set to 1, this equal weighting leads to a worse complexity with an extra \( \log \) term, i.e. \( \epsilon^{-4} |\log(\epsilon)|^4 \). We verify that \( \{ w_m^{(2)} \} \) satisfies properties (I) to (III) in Lemma A.3.

### 3.2 Non-interacting kernels

Here we remark how the theory developed in this work would simplify, if we only treated MVSDs with non-interacting kernels given by
\[
dX_t^m = b \left( X_t^m, \int_{\mathbb{R}^d} f(y) \mu_t^{X_{m-1}}(dy) \right) \, dt + \sigma \left( X_t^m, \int_{\mathbb{R}^d} g(y) \mu_t^{X_{m-1}}(dy) \right) \, dW_t^m, \quad \mu_0^X = \mu_0^X,
\] (3.44)
for some continuous functions \( b : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}^d \) and \( \sigma : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}^{d \otimes r} \). We assume (Ker-Reg), \( (\mu_0^X, L_p) \) and
Each component function of \( f \) and \( g \) belongs to the set \( C^2_b(\mathbb{R}^d, \mathbb{R}^q) \).

The difference compared to the more general case is that the interaction kernels no longer have any dependence on the state variable \( x \in \mathbb{R}^d \) (they depend only on the measure). Hence, the simulation cost at every step of the corresponding Euler particle system is \( N \) (not \( N^2 \)). This leads to an overall computational complexity of \( O(\epsilon^{-3}) \). Using the above notation, the algorithm becomes

\[
dY^{i,m}_t = b(Y^{i,m}_t, \mathcal{M}^{(m-1)}_{\eta_t}(f)) dt + \sigma(Y^{i,m}_t, \mathcal{M}^{(m-1)}_{\eta_t}(g)) dW^{i,m}_t.
\]

As an example of MVSDEs with non-interacting kernels, one may consider the simplified Kuramoto model with synchronised oscillators [Acebrón et al., 2005], which has the form

\[
dx_t = \int_{\mathbb{R}} \sin(x_t - y) \mu_t(dy) dt + \sigma dW_t.
\]

Its corresponding particle approximation can be rewritten as

\[
dY^{i,N}_t = \left( \sin(Y^{i,N}_t) \frac{1}{N} \sum_{j=1}^{N} \cos(Y^{j,N}_t) - \cos(Y^{i,N}_t) \frac{1}{N} \sum_{j=1}^{N} \sin(Y^{j,N}_t) \right) dt + \sigma dW_t^i,
\]

where \( \{X^i_0\}_{i=1,...,N} \) are i.i.d with law \( \mu_0 \) and \( \{W^i_t\}_{i=1,...,N} \) are independent Brownian motions.

To fit the abstract framework into the non-interacting case, we set \( \mathcal{V}_t \) as in (3.1), \( \kappa = 2L + 1 \), and define

\[
B(x, \mathcal{V}_t) := b(x, \int_{\mathbb{R}^d} f(y) \mathcal{V}_t^{(1)}(dy) + \sum_{k=1}^{L} \left( \int_{\mathbb{R}^d} f(y) \mathcal{V}_t^{(2k+1)}(dy) - \int_{\mathbb{R}^d} g(y) \mathcal{V}_t^{(2k)}(dy) \right))
\]

and

\[
\Sigma(x, \mathcal{V}_t) := \sigma(x, \int_{\mathbb{R}^d} g(y) \mathcal{V}_t^{(1)}(dy) + \sum_{k=1}^{L} \left( \int_{\mathbb{R}^d} g(y) \mathcal{V}_t^{(2k+1)}(dy) - \int_{\mathbb{R}^d} g(y) \mathcal{V}_t^{(2k)}(dy) \right)),
\]

for any \( x \in \mathbb{R}^d \). Though the analysis in this paper is done for the interacting case, all the results remain true for the non-interacting case. Nonetheless, many of the theorems and proofs in Section 2 will be much simpler under the non-interacting case. For example, we no longer need to use the weighted measures \( \mu_{\eta_t}(s) \) and \( \bar{\mu}_{\eta_t}(s) \) in Theorem 2.7 and Lemma 2.8 respectively, since there is no longer a need to consider functions of the form \( G(x, \cdot) \).

Finally, computational complexity is reduced to the order of \( \epsilon^{-2} |\log \epsilon|^2 \), due to the fact that the sum \( \sum_{\ell'=0}^{L} N_{\ell'} \) in (4.3) becomes 1.
4 Iteration of the MC algorithm

In this section, we consider iterated particle system (1.8). Using the notation of Section 2, we set

\[ V_{m-1}(t) = \frac{1}{N_{m-1}} \sum_{i=1}^{N_{m-1}} \delta_{Y_{i,m-1,L}}(t), \]

\[ \bar{B}(x, V_{m-1}) := \frac{1}{N_{m-1}} \sum_{i=1}^{N_{m-1}} b(x, Y_{i,m-1,L}) \] (4.1)

and

\[ \bar{\Sigma}(x, V_{m-1}) := \frac{1}{N_{m-1}} \sum_{i=1}^{N_{m-1}} \sigma(x, Y_{i,m-1,L}), \] (4.2)

for any \( x \in \mathbb{R}^d \). By the definition of equation (1.8), the conditions (\( V \perp (W, Z_0) \)), \( (V\text{-bound}) \), \( (V\text{-Reg}) \) and \( (V\text{-Lip}) \) hold.

Accordingly, Lemma 4.1 gives a decomposition of MSE (mean-square-error) for MC along one iteration of the particle system (1.8).

**Lemma 4.1.** Assume \((\text{Ker-Reg})\) and \((\mu_0 - L_p)\). Let \( G \in C^2_p(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R}) \) be a globally Lipschitz continuous function. We define \( \bar{B} \) and \( \bar{\Sigma} \) as in (4.1) and (4.2). Let

\[ MSE_t^{(m)}(G(x, \cdot)) := \mathbb{E} \left[ \left( \mathbb{E}[G(x, X_t)] - \frac{1}{N_m} \sum_{i=1}^{N_m} G(x, Y_{i,m,L}^t) \right)^2 \right], \quad t \in [0, T]. \]

Then, for every \( t \in [0, T] \)

\[
\int_{\mathbb{R}^d} MSE^{(m)}_{\eta_L(t)}(G(x, \cdot)) \mu^{2,\eta_L}_{\eta_L(t)}(dx) \leq c \left( h^2 + \int_0^T \left[ \mathbb{E} \left| \bar{B}(x, V_{m,L}) - \mathbb{E}[b(x, X_{m,L})] \right| \right]^2 \mu^{2,\eta_L}_{\eta_L(s)}(dx) ds \right) 
+ \int_0^T \left[ \mathbb{E} \left| \bar{\Sigma}(x, V_{m,L}) - \mathbb{E}[\sigma(x, X_{m,L})] \right| \right]^2 \mu^{2,\eta_L}_{\eta_L(s)}(dx) ds + \frac{1}{N_m}. \]

Subsequently, its error analysis is presented in Theorem 4.2.

**Theorem 4.2.** Assume \((\text{Ker-Reg})\) and \((\mu_0 - L_p)\). Fix \( M > 0 \) and take \( C^2_p \supseteq P : \mathbb{R}^d \to \mathbb{R} \) to be a globally Lipschitz continuous function. As before, we define the mean-square error as

\[ MSE_t^{(M)}(P) := \mathbb{E} \left[ \left( \frac{1}{N_M} \sum_{i=1}^{N_M} P(Y_{i,M,L}^t) - \mathbb{E}[P(X_t)] \right)^2 \right]. \]
Then for every $t \in [0, T]$,

$$MSE_{\eta_1(t)}^{(M)}(P) \leq c\left\{ h_1^2 L + \sum_{m=1}^{M} \frac{c^{M-m}}{(M-m)!} \frac{1}{N_m} + \frac{c^{M-1}}{M!} \right\}.$$ 

Finally, we are now in a position to present the complexity theorem for iterated MC estimators of $\{E[P(X_{\eta_1(t)})]\}_{t \in [0,T]}$.

**Theorem 4.3.** Assume (Ker-Reg) and $(\mu_0,L_p)$. Fix $M > 0$ and let $C_p^2 \ni P : \mathbb{R}^d \to \mathbb{R}$ be a globally Lipschitz continuous function. Then for any $\epsilon < e^{-1}$, there exist $M$, $L$ and $\{N_{m,L}\}_{0 \leq m \leq M}$ such that for every $t \in [0, T]$,

$$MSE_{\eta_1(t)}^{(M)}(P) := \mathbb{E}\left( \frac{1}{N_M} \sum_{i=1}^{N_M} P(Y_{i,M,L}^{\eta_1(t)}) - \mathbb{E}[P(X_{\eta_1(t)})]\right)^2 < \epsilon^2,$$

and computational complexity $C$ is of order $\epsilon^{-5}$.

**Proof.** The optimisation problem simplifies to

$$\min_{M,L_{m=1}^{M}, N_{m=1}^{M}} C\left( M, \{L_{m}\}_{m=1}^{M}, \{N_{m,L}\}_{1 \leq m \leq M} \right) = h_{1,L_1,N_1}^{-1} + \sum_{m=2}^{M} h_{L_m,N_m,L_m}^{-1} N_{m,L_m,N_{m-1,L_{m-1}}}$$

such that

$$\sum_{m=1}^{M} \frac{c^{M-m}}{(M-m)!} \left( h_{L_m}^{-2} + \frac{1}{N_{m,L_m}} \right) + \frac{c^{M-1}}{M!} \lesssim \epsilon^2.$$ 

We obtain the result by following the same procedures shown in the proof of Theorem 3.7 and setting $N_{m,L_m}^* \sim \epsilon_m^{-2}$ and $L_m^* \sim \log(\epsilon^{-1})$.

**Remark 4.4.** If we set the accuracy level to be $\epsilon^2$ for all Picard steps, we see that its complexity has an extra log term, i.e. $\epsilon^{-5} \log(\epsilon)$. However, by setting a sequence with accuracy levels $\epsilon_m^2$ to the corresponding levels, we can remove the log term and reach the same (asymptotic) complexity as the standard particle system.

### 5 Numerical results

In this section, we provide a numerical example of a one-dimensional stochastic differential equation derived from the Kuramoto model:

$$dX_t = \int_{\mathbb{R}} \sin(X_t - y) \mu_X^t(dy) \, dt + dW_t, \quad t \in [0,1], \quad X_0 = 0,$$

with payoff function $P(x) = \sqrt{1 + x^2}$. We initialise 100 Gaussian sample paths $Y_t^{i,0,\ell} \sim N(0,t)$. We compare the following methods:
• the classical particle system method outlined in Section 1.1,
• an iterated MC method (MC Picard I) by using a fixed number of particles \( N \) for all Picard steps,
• an iterated MC method (MC Picard II) by using an increasing sequence of particles \( \{N_m\}_{m=1,...,M} \) (briefly discussed in Remark 5.1),
• the iterated MLMC method from Algorithm 1.

As a measurement of the true computational cost, one needs to consider the number of random numbers and the evolutions of drift and diffusion coefficients.

**Remark 5.1.** The Euler scheme for MC Picard II is given by

\[
Y_{i,m}^{k+1} = Y_{i,m}^k + \left[ \sin(Y_{i,m}^k) \frac{1}{N_{m-1}} \sum_{j=1}^{N_{m-1}} \cos(Y_{j,m}^{j-1}) + \cos(Y_{i,m}^k) \frac{1}{N_{m-1}} \sum_{j=1}^{N_{m-1}} \sin(Y_{j,m}^{j-1}) \right] h + \Delta W_{i,m}^{k+1}.
\]

A comparison of computational complexity is presented in Figure 5.1. Figure 5.1a shows that both MC Picard I and MC Picard II are more expensive than the classical particle system. In Figure 5.1b, the iterated MLMC method is less expensive than the classical particle system method concerning the computational cost, as the mean-square-error bound \( \epsilon^2 \) decreases. Figure 5.1c illustrates that the error incurred from the iterated MLMC estimator is within \( 2\epsilon \) of that of the classical particle system and that it decreases as number of particles increases.

Moreover, Figure 5.2 confirms numerically that the iterated MLMC method can be improved by taking a decreasing sequence \( \{\epsilon_m\}_{m=1,...,M} \), as opposed to taking a constant sequence.

Finally, Figure 5.3 verifies, in the logarithm sense, the variance of \( P_{T,m,\ell}^{i,m} \) for each Picard step. Also, it presents the logarithmic variance of \( P_{T,m,\ell}^{i,m} - P_{T,m,\ell-1}^{i,m} \). All lines of \( \log_2 \text{Var}[P_{T,m,\ell}^{i,m} - P_{T,m,\ell-1}^{i,m}] \) against \( \ell \) have a slope of approximately \(-2\) and this is higher than the rate given in Lemma 2.4, since it is well-known that for constant diffusion coefficients, the strong error of Euler scheme (2.4) is of order \( O(h^{-2}) \).
Figure 5.1: Comparison of computational complexity against $\epsilon$

Figure 5.2: Constant against a decreasing sequence
A Useful lemmas

Lemma A.1. Let \( \{Q_t\}_{t\in[0,T]} \) be a cadlag square-integrable process adapted to the filtration \( \{\mathcal{F}_t\}_{t\in[0,T]} \). Suppose that the Brownian motion \( \{W_t\}_{t\in[0,T]} \) is adapted to \( \{\mathcal{F}_t\}_{t\in[0,T]} \). Let \( \mathcal{G} \) be a \( \sigma \)-algebra such that \( \mathcal{G} \subseteq \mathcal{F}_0 \). Then the following equalities hold for any \( t \in [0,T] \).

(a) \( \mathbb{E} \left[ \int_0^t Q_s dW_s \mid \mathcal{G} \right] = 0 \),
(b) \( \mathbb{E} \left[ \left( \int_0^t Q_s dW_s \right)^2 \mid \mathcal{G} \right] = \mathbb{E} \left[ \int_0^t Q_s^2 ds \mid \mathcal{G} \right] \).

Proof.
(a) By the tower property of conditional expectation,
\[
\mathbb{E} \left[ \int_0^t Q_s dW_s \mid \mathcal{G} \right] = \mathbb{E} \left[ \mathbb{E} \left[ \int_0^t Q_s dW_s \mid \mathcal{F}_0 \right] \mid \mathcal{G} \right] = 0.
\]

(b) Similarly, by Itô’s isometry and the tower property of conditional expectation,
\[
\mathbb{E} \left[ \left( \int_0^t Q_s dW_s \right)^2 \mid \mathcal{G} \right] = \mathbb{E} \left[ \mathbb{E} \left[ \left( \int_0^t Q_s dW_s \right)^2 \mid \mathcal{F}_0 \right] \mid \mathcal{G} \right]
= \mathbb{E} \left[ \int_0^t Q_s^2 ds \mid \mathcal{F}_0 \right] \mathbb{E} \left[ \left( \int_0^t Q_s dW_s \right)^2 \mid \mathcal{G} \right]
= \mathbb{E} \left[ \int_0^t Q_s^2 ds \mid \mathcal{G} \right].
\]
The following result concerning regular conditional probability distributions is needed in the proofs of Theorem 2.7 and Lemma 2.8. (For example, see Th. 7.1 in [Parthasarathy, 1967] for details.)

**Lemma A.2.** Let \((M_1, B_1)\) and \((M_2, B_2)\) be complete separable metric spaces. Let \(X\) and \(Y\) be random variables in the same probability space \((\Omega, \mathcal{F}, \mathbb{P})\), with \(X\) taking values in \((M_1, B_1)\) and \(Y\) taking values in \((M_2, B_2)\). Let \(G \subseteq \mathcal{F}\) be a sub-\(\sigma\)-algebra. Suppose that \(Y\) is \(G\)-measurable. Then, for every Borel function \(h : M_1 \times M_2 \to \mathbb{R}\) such that \(\mathbb{E}[h(X, Y)] < +\infty\), there exists (a.s.) a probability measure \(\mu_{X|G} : B_1 \times \Omega \to [0, 1]\) (called the regular conditional probability measure of \(X\) given \(G\)) such that

\[
(\mathbb{E}[h(X, Y)|G])(\omega) = \int_{M_1} h(x, Y(\omega)) \mu_{X|G}(dx, \omega).
\]

Suppose that, in addition, \(X\) is independent of \(G\). Then, \(\mu_{X|G}(.|\omega) = \text{Law}(X)\) for \(\omega\)-almost surely.

**Lemma A.3.** The sequence \(\{w_m\}_{m=1}^{M^*}\) defined as

\[
w_m := \begin{cases} \max\left(\frac{M^*-m-2}{e^{M^*-m-2}}, 1\right), & m \leq M^*-2 \\ 1, & M^*-1 \leq m \leq M^* \end{cases}
\]

satisfies properties (I) to (III).

**Proof.** First, property (I) follows easily from the definition of it, since \(w_m \geq 1 = w_{M^*}\) for all values of \(m\). For property (II), i.e. \(\sum_{m=1}^{M^*} \frac{e^{M^*-m}}{(M^*-m)!} w_m \sim 1\), we confirm that \(\sum_{m=1}^{M^*} \frac{e^{M^*-m}}{(M^*-m)!} w_m \geq 1\) by observing that

\[
\sum_{m=1}^{M^*} \frac{e^{M^*-m}}{(M^*-m)!} w_m = \sum_{m=1}^{M^*-2} \frac{e^{M^*-m}}{(M^*-m)!} \max\left(\frac{M^*-m-2}{e^{M^*-m-2}}, 1\right) + \sum_{m=M^*-1}^{M^*} \frac{e^{M^*-m}}{(M^*-m)!}.
\]

\[
\geq \sum_{m=1}^{M^*-2} \frac{e^{M^*-m}}{(M^*-m)!} + \sum_{m=M^*-1}^{M^*} \frac{e^{M^*-m}}{(M^*-m)!} = \sum_{m=1}^{M^*} \frac{e^{M^*-m}}{(M^*-m)!} \sim 1.
\]

Subsequently, we verify that \(\sum_{m=1}^{M^*} \frac{e^{M^*-m}}{(M^*-m)!} w_m \lesssim 1\), since

\[
\sum_{m=1}^{M^*} \frac{e^{M^*-m}}{(M^*-m)!} w_m = \sum_{m=1}^{M^*-2} \frac{e^{M^*-m}}{(M^*-m)!} \max\left(\frac{M^*-m-2}{e^{M^*-m-2}}, 1\right) + \sum_{m=M^*-1}^{M^*} \frac{e^{M^*-m}}{(M^*-m)!}.
\]
\[
\sum_{m=1}^{M^*-2} \frac{c^{M^*-m}}{(M^*-m)!} \left( \frac{(M^*-m-2)!}{c^{M^*-m-2}} + 1 \right) + \sum_{m=M^*-1}^{M^*} \frac{c^{M^*-m}}{(M^*-m)!} \\
= \sum_{m=1}^{M^*} \frac{c^{M^*-m}}{(M^*-m)!} + \sum_{m=1}^{M^*-2} \frac{c^2}{(M^*-m)(M^*-m-1)} \\
= \sum_{m=1}^{M^*} \frac{c^{M^*-m}}{(M^*-m)!} + c^2 \sum_{m=1}^{M^*-2} \frac{1}{(M^*-m)(M^*-m-1)} \\
= \sum_{m=1}^{M^*} \frac{c^{M^*-m}}{(M^*-m)!} + c^2 \sum_{m=2}^{M^*-1} \frac{1}{m(m-1)} \\
= \sum_{m=1}^{M^*} \frac{c^{M^*-m}}{(M^*-m)!} + c^2 \left( 1 - \frac{1}{M^*-1} \right) \sim 1.
\]

Lastly, we show this sequence has property (III). Since
\[
0 < \sum_{m=2}^{M^*-3} w_m^{-1} w_{m-1}^{-1} = \sum_{m=2}^{M^*-3} \frac{c^{M^*-m-2}}{(M^*-m-2)!} \cdot \frac{c^{M^*-m-3}}{(M^*-m-3)!} \leq c^{2c},
\]
the proof is complete. \qed

**Acknowledgements**

We are grateful to Mireille Bossy, Mike Giles and David Šiška for helpful comments.

**References**

[Acebrón et al., 2005] Acebrón, J. A., Bonilla, L. L., Vicente, C. J. P., Ritort, F., and Spigler, R. (2005). The kuramoto model: A simple paradigm for synchronization phenomena. *Reviews of modern physics*, 77(1):137.

[Antonelli et al., 2002] Antonelli, F., Kohatsu-Higa, A., et al. (2002). Rate of convergence of a particle method to the solution of the McKean–Vlasov equation. *The Annals of Applied Probability*, 12(2):423–476.

[Bosy, 2004] Bosy, M. (2004). Optimal rate of convergence of a stochastic particle method to solutions of 1d viscous scalar conservation laws. *Mathematics of computation*, 73(246):777–812.

[Bossy et al., 1997] Bossy, M., Fezou, L., Piperno, S., et al. (1997). Comparison of a stochastic particle method and a finite volume deterministic method applied to burgers equation. *Monte Carlo Methods and Applications*, 3:113–140.
[Bossy et al., 2002] Bossy, M., Jourdain, B., et al. (2002). Rate of convergence of a particle method for the solution of a 1d viscous scalar conservation law in a bounded interval. *The Annals of Probability*, 30(4):1797–1832.

[Bossy and Talay, 1997] Bossy, M. and Talay, D. (1997). A stochastic particle method for the McKean-Vlasov and the Burgers equation. *Mathematics of Computation of the American Mathematical Society*, 66(217):157–192.

[Bossy et al., 1996] Bossy, M., Talay, D., et al. (1996). Convergence rate for the approximation of the limit law of weakly interacting particles: application to the Burgers equation. *The Annals of Applied Probability*, 6(3):818–861.

[Bujok et al., 2013] Bujok, K., Hambly, B., and Reisinger, C. (2013). Multilevel simulation of functionals of Bernoulli random variables with application to basket credit derivatives. *Methodology and Computing in Applied Probability*, pages 1–26.

[Carmona et al., 2013] Carmona, R., Delarue, F., and Lachapelle, A. (2013). Control of McKean–Vlasov dynamics versus mean field games. *Mathematics and Financial Economics*, 7(2):131–166.

[Chassagneux et al., 2014] Chassagneux, J.-F., Crisan, D., and Delarue, F. A probabilistic approach to classical solutions of the master equation for large population equilibria.

[Delarue et al., 2015] Delarue, F., Inglis, J., Rubenthaler, S., and Tanré, E. (2015). Particle systems with a singular mean-field self-excitation. Application to neuronal networks. *Stochastic Processes and their Applications*, 125(6):2451–2492.

[Friedman, 2006] Friedman, A. (2006). *Stochastic differential equations and applications*. Courier Corporation.

[Giles, 2008] Giles, M. B. (2008). Multilevel monte carlo path simulation. *Operations Research*, 56(3):607–617.

[Giles et al., 2015] Giles, M. B., Nagapetyan, T., and Ritter, K. (2015). Multilevel monte carlo approximation of distribution functions and densities. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):267–295.

[Haji-Ali and Tempone, 2016] Haji-Ali, A.-L. and Tempone, R. (2016). Multilevel and multi-index monte carlo methods for mckean-vlasov equations. *arXiv preprint arXiv:1610.09934*.

[Heinrich, 2001] Heinrich, S. (2001). Multilevel monte carlo methods. In *Large-scale scientific computing*, pages 58–67. Springer.
[Hutzenthaler et al., 2016] Hutzenthaler, M., Jentzen, A., Kruse, T., et al. (2016). On full history recursive multilevel picard approximations and numerical approximations of high-dimensional nonlinear parabolic partial differential equations. *arXiv preprint arXiv:1607.03295*.

[Kebaier et al., 2005] Kebaier, A. et al. (2005). Statistical romberg extrapolation: a new variance reduction method and applications to option pricing. *The Annals of Applied Probability*, 15(4):2681–2705.

[Krylov, 1980] Krylov, N. V. (1980). *Controlled diffusion processes*, volume 14 of *Applications of Mathematics*. Springer-Verlag, New York-Berlin. Translated from the Russian by A. B. Aries.

[Krylov, 2002] Krylov, N. V. (2002). *Introduction to the theory of random processes*, volume 43. American Mathematical Soc.

[Lemaire et al., 2017] Lemaire, V., Pagès, G., et al. (2017). Multilevel richardson–romberg extrapolation. *Bernoulli*, 23(4A):2643–2692.

[Mao, 2007] Mao, X. (2007). *Stochastic differential equations and applications*. Elsevier.

[McKean Jr, 1966] McKean Jr, H. (1966). A class of Markov processes associated with nonlinear parabolic equations. *Proceedings of the National Academy of Sciences of the United States of America*, 56(6):1907.

[Méléard, 1996] Méléard, S. (1996). Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. In *Probabilistic models for nonlinear partial differential equations*, pages 42–95. Springer.

[Parthasarathy, 1967] Parthasarathy, K. R. (1967). *Probability measures on metric spaces*, volume 352. American Mathematical Soc.

[Pope, 2000] Pope, S. B. (2000). *Turbulent flows*. Cambridge university press.

[Ricketson, 2015] Ricketson, L. (2015). A multilevel monte carlo method for a class of mckean-vlasov processes. *arXiv preprint arXiv:1508.02299*.

[Sznitman, 1991] Sznitman, A.-S. (1991). *Topics in propagation of chaos*. Springer.