Unbiased Sampling of Multidimensional Partial Differential Equations with Random Coefficients

Jose Blanchet, Fengpei Li, and Xiaou Li

September 8, 2018

Abstract

We construct an unbiased estimator for function value evaluated at the solution of a partial differential equation with random coefficients. We show that the variance and expected computational cost of our estimator are finite and our estimator is insensitive to the problem dimension so that it can be applied in problems across various disciplines. For the error analysis, we connect the parabolic partial differential equations with the expectations of stochastic differential equations and perform rough path estimation.

1 Introduction

1.1 Motivation and background

Consider the solution $u(x, t): \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}$ of the following random parabolic partial differential equation (PDE):

$$\begin{cases} \partial_t u(x, t) = \sum_{i=1}^{d} \mu_i(x) D_x u(x, t) + \frac{1}{2} tr \left( \sigma(x) \sigma^T(x) D_{xx} u(x, t) \right) \\ u(x, 0) = f(x) \end{cases} \tag{1.1}$$

where $\{\mu(x) : x \in \mathbb{R}^d\} \subset \mathbb{R}^d$ is a random field living in some probability space $(\Omega, \mathcal{F}, P)$ with each realization denoted by $\mu(\cdot, \omega): \mathbb{R}^d \to \mathbb{R}^d$ and $\{\sigma(x)\sigma^T(x) : x \in \mathbb{R}^d\} \subset \mathbb{R}^{d \times d}$ is constructed from some deterministic (non-random) function $\sigma(\cdot)$. In general, both functions satisfy certain regularity conditions to be specified later. On the other hand, $D_x u(x, t) \in \mathbb{R}^d$ and $D_{xx} u(x, t) \in \mathbb{R}^{d \times d}$ in (1.1) denote the partial derivatives while $tr(\cdot)$ is the matrix trace operator. The function $\{f(x) : x \in \mathbb{R}^d\}$ describes the initial condition.

The heat equation (1.1) above is a classic PDE that has applications in various disciplines. Moreover, for different cases, the interpretations for the coefficients in (1.1) and the solutions
$u(x,t)$ are also different. For example, in the theory of thermal conductivity, the heat equation (1.1) follows from Fourier’s law [2]. In this case, the solution $u(x,t)$ represents the temperature of the material at the location $x$ and time $t$, while the coefficients $\mu$ and $\sigma$ characterize the thermal conductivity of the material. On the other hand, in the theory of flow dynamics [40], the heat equation (1.1) follows from Darcy’s law when one tries to describe the flow of fluids through a porous medium. Here, the solution $u(x,t)$ represents the fluid pressure at location $x$ and time $t$, while the coefficients $\mu$ and $\sigma$ characterize the medium permeability. In both cases, the coefficients at location $x$ reflect medium property, which are usually modeled as random fields due to the heterogeneity of the media in practical applications. A partial literature on the modeling and analysis for the heterogeneous/random medium includes [10, 11, 33, 37, 41].

Moreover, for applications in finance, we note that Equation (1.1) represents the price of a European contract with payoff given by $f(x)$ at maturity [12]. In this case, the introduction of a random $\mu$ and deterministic $\sigma$ is justified by the fact that the diffusion coefficient $\mu$ can often be estimated with reasonable accuracy in the setting of financial applications while the drift $\sigma$ is typically difficult to calibrate [35, 21].

In these applications, because $\mu$ is a random vector field, the solution $u(x,t): \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}$ to Equation (1.1) is also random. To avoid confusion, we will henceforth use $u(x,t)$ to denote the solution of (1.1) when the field $\mu$ is considered random and use $u(x,t)$ to denote the solution of (1.1) when the field $\mu$ (or $\mu(\cdot,\omega)$ in particular) is considered fixed. As it turned out, in the context of random PDEs, it is quite common to evaluate expectations of the form

$$\nu = \mathbb{E} [G(u(x_1,t_1),...,u(x_k,t_k))],$$

(1.2)

for certain values of $x_i,t_i, 1 \leq i \leq k$ and some given function $G: \mathbb{R}^d \to \mathbb{R}$ (e.g., compute function values with input as the payoff of European contract). As we shall see, this task presents a substantial analytic challenge and it is natural for one to resort to Monte Carlo method. In this paper, we introduce a methodology that provides an unbiased estimator for $\nu$ in (1.2) and could be easily implemented by parallel computing architectures.

### 1.2 Main contribution

Under reasonable regularity conditions to be specified in Theorem 2.2.1, we construct a random variable $W$ satisfying:

1. unbiasedness: $\mathbb{E}(W) = \nu$;
2. finite variance: $\text{Var}(W) < \infty$; and
3. finite expected cost: The computational cost required to simulate $W$ has finite expectation.
Consequently, based on (1), (2), and (3), we can generate \( n \) independent copies of \( W \) in parallel servers and combine them to provide estimation as well as confidence intervals for \( \nu \) in (1.2) with \( O(n^{-1/2}) \) rate of convergence dictated by the central limit theorem (CLT). Specifically, if the parallel computing cores are relatively cheap to obtain and wall-clock time is a relatively hard constraint, then the estimator \( W \) we propose in this paper is precisely the type of solution to one wants to use in order to estimate \( \nu \) in (1.2).

As far as we know, our paper is the first to produce unbiased estimators of \( \nu \) with square-root convergence rate for arbitrary dimension \( d \) in the PDE (1.1). For example, the unbiased estimator proposed in [27] is related to the solution of elliptic equations with random input and Dirichlet boundary conditions. However, even the sampling strategy in [27] achieves square-root convergence rate, the estimator has finite variance only if \( d \leq 3 \). The rise of the curse of dimensionality in [27] lies in the procedure. In particular, the procedure in [27] is to numerically solve the PDE using a finite elements method whose related analysis on the rate of convergence depends on the underlying dimension \( d \). In fact, there has been a substantial amount of recent literature combining the multilevel Monte Carlo technique with the numerical methods for PDE, all of which suffer from the curse of dimensionality as the rate of convergence deteriorates with the increase of problem dimensions [27, 34, 7, 32]. On the other hand, other available methods in the literature such as [39, 18, 8] produce biased estimators. In contrast, our method allows for a full Monte Carlo procedure with a traditional square-root convergence rate for any dimension \( d \). Thus, our proposed method preserves the well-known characteristic of the Monte Carlo method in effectively combating the curse of dimensionality. Although the constants in the convergence rate analysis of Monte Carlo depend on the dimension \( d \), the convergence rate as a function of the total number of random variables generated (the level of simulation accuracy) is of the same order for any \( d \) and conditions (1), (2), (3) always hold. Thus, if one tries to estimate \( \nu \) in (1.2) to a desired level of accuracy with a relatively hard wall-clock time constraint, these advantages of our estimator will become highly valuable.

1.3 Technical contribution

In this paper, we exploit the connection between the parabolic PDE and stochastic differential equations (SDEs) in order to construct \( W \). In particular, given a sample path of the random field \( \mu(\cdot, \omega) \) (i.e., fixed \( \omega \), the realization of random field \( \mu(\cdot) \)), it follows from the celebrated Feynman-Kac formula [23] that we can represent the solution of Equation (1.1) \( u(x, t) \) as the expectation of the solution to certain SDEs. Following this approach, conditioning on \( \mu(\cdot, \omega) \), we then use a multilevel Monte Carlo construction introduced in [17] that efficiently discretizes the underlying SDE to reduce variance and combine the method with a randomization step introduced in [36] to construct an estimator without bias. Finally, we
introduce an additional randomization technique similar to that in [5] to account for the randomness of $\mu$.

On the other hand, in terms of technical contribution, the error analysis of the additional randomization step requires a non-standard technical development. In particular, conditioning on $\mu(\cdot, \omega)$, the realization of the random field $\mu$, the standard error analysis in [17, 26] would yield a term with infinite expectation, which prevents us from showing the finite variance property of our estimator. This is due to the presence of the famous Gronwall’s inequality [22] as a common tool in these stochastic analyses (see, e.g., [26] and the remark following Lemma 3.2.4).

In order to overcome this issue and obtain a bound on the variance of our estimator, we use the theory of rough paths to create certain path-by-path estimates. The theory of rough paths [31, 9, 14, 13] has received substantial attention in the literature due to its connection to the theory of regularity structures and its implications in the analysis of nonlinear stochastic PDEs [20, 19]. On the other hand, a significant amount of literature has also been devoted to exploring the connection between the theory of rough paths and stochastic numerical analysis in the setting of cubature methods [30] or SDEs [3, 4]. In this light, our paper is the first to connect the rough path estimates with the numerical analysis of random PDEs and thus adds to the growing literature combining the theory of rough paths with numerical stochastic analysis.

1.4 Organization

The rest of the paper is organized as follows. In Section 2 we explain the model settings and our construction of $W$ in detail. We also address the properties (1),(2) and (3) of our estimator $W$ followed by the simulation algorithm for it. In Section 4 we provide several numerical examples to demonstrate the effectiveness of our estimator $W$. Finally, we present the technical proofs in Section 5 and in the appendix.

2 Main results

In order to state our results, we now introduce our assumptions and several technical conditions. Assumption [A1] is similar to the Karhunen-Loève expansion [25, 28] which ensures the smoothness of the random field $\mu$. Assumption [A2] includes smoothness conditions on coefficients $\sigma$ and initial condition $f$, which are related to the existence of probabilistic representation of PDE solution. Assumption [A3] is a smoothness condition on function $G(\cdot)$ related to the finite variance property of $W$. 
2.1 Assumptions and technical conditions

A1. (Smoothness of Gaussian Random Field) The random field \( \mathbf{\mu}(\cdot) : \mathbb{R}^d \to \mathbb{R}^d \) has the following expansion

\[
\mathbf{\mu}(\cdot) = \sum_{i=1}^{\infty} \frac{\lambda_i}{q} \cdot \mathbf{V}_i \cdot \psi_i(\cdot)
\]

(2.1)

where \( q > 4 \) is a fixed constant, \( \{\lambda_i : i = 1, 2, \ldots\} \) is a uniformly bounded sequence and \( \{\mathbf{V}_i : i = 1, 2, \ldots\} \) is a sequence of independent and identically distributed (I.I.D.) random variables \( \mathcal{N}(0, I_d) \), \( d \)-dimensional multivariate Gaussian. On the other hand, \( \psi_i(\cdot) : \mathbb{R}^d \to \mathbb{R} \) \( (i = 1, 2, \ldots) \) is a sequence of deterministic functions with a positive constant \( 1 < L < \infty \) such that for \( 0 \leq k, l \leq d, i \geq 1, \)

\[
\|\psi_i\|_\infty < L, \quad \|\frac{\partial \psi_i}{\partial x_k}\|_\infty < iL, \quad \text{and} \quad \|\frac{\partial^2 \psi_i}{\partial x_k \partial x_l}\|_\infty < i^2L.
\]

We will show that the expansion on the right of (2.1) is convergent almost surely and Assumption A1 essentially provides an almost surely smoothness condition on the random field \( \mathbf{\mu}(x) \). Specifically, we have the following definition and lemma.

Definition 2.1.1. We denote \( \mathcal{L}_1 \) to be the space of bounded, Lipschitz continuous and twice continuously differentiable field with each element \( \mathbf{\mu}(\cdot) : \mathbb{R}^d \to \mathbb{R}^d \) satisfying, for \( 1 \leq i, k, l \leq d, \)

\[
\|\mathbf{\mu}\|_\infty < L_1, \quad \|\frac{\partial \mathbf{\mu}_i}{\partial x_k}\|_\infty < L_1 \quad \text{and} \quad \|\frac{\partial^2 \mathbf{\mu}_i}{\partial x_k \partial x_l}\|_\infty < L_1
\]

for some positive \( L_1 < \infty \) depending on \( \mathbf{\mu}(\cdot) \) and we shall called \( L_1 \) a bounding number of \( \mathbf{\mu}(\cdot) \).

Lemma 2.1.2. Under Assumption A1, \( \mathbf{\mu}(\cdot) \in \mathcal{L}_1 \) almost surely. Also, for any \( n \geq 0 \), the partial sum \( \mathbf{S}_n = \sum_{i=1}^{n} \frac{\lambda_i}{q} \cdot \mathbf{V}_i \cdot \psi_i(\cdot) \in \mathcal{L}_1 \) almost surely. Furthermore, there exists a random variable \( L_1 > 1 \) with \( \mathbb{E}(e^{tL_1}) < \infty \) for \( t \in \mathbb{R} \) (i.e, well-defined moment-generating function) such that it is a bounding number for \( \mathbf{\mu} \) and every \( \mathbf{S}_n \) almost surely.

Proof. See Section 5 \( \square \)

After imposing smoothness condition on \( \mathbf{\mu} \), we now state our assumptions on the smoothness of our deterministic functions \( \sigma, f \) in (1.1) and \( G \) in (1.2).

A2. (Smoothness of Diffusion Coefficient \( \sigma \) and Initial Condition \( f \)) The function \( \sigma(\cdot) : \mathbb{R}^d \to \mathbb{R}^{d \times d} \) is uniformly bounded and has uniformly bounded first and second order derivatives. The function \( f(\cdot) : \mathbb{R}^d \to \mathbb{R} \) is twice differentiable with bounded first and second
order derivatives In particular, there exists a positive constant \( 1 < L < \infty \) such that for \( 1 \leq i', j' \leq d' \), \( 1 \leq i, j, l \leq d \), we have

\[
\| \sigma_{i'} \|_{\infty} < L, \quad \| \partial \sigma_{i'j'} \|_{\infty} < L \quad \text{and} \quad \| \partial^2 \sigma_{i'j'} \|_{\infty} < L, \\
\| \partial f \|_{\infty} < L \quad \text{and} \quad \| \partial^2 f \|_{\infty} < L.
\]

A3. (Smoothness of Function \( G \)) The function \( G(\cdot) : \mathbb{R}^k \to \mathbb{R} \) is twice differentiable with bounded first and second order derivatives. In particular, there exists a positive constant \( 1 < L < \infty \) such that for \( 1 \leq i, j \leq k \),

\[
\| \frac{\partial^2 G}{\partial x_i \partial x_j} \|_{\infty} < L \quad \text{and} \quad \| \frac{\partial G}{\partial x_i} \|_{\infty} < L.
\]

We note that we have used the same notation \( L \) for the bounding constants in Assumptions A1-A3 for convenience. The separate notation \( L_1 \) for the bounding term in Lemma 2.1.2 is due to its randomness. Now we present the main theorem of our paper.

### 2.2 Main theorem

**Theorem 2.2.1.** Under Assumptions A1-A3, we can construct random variable \( W \), which is an unbiased estimator for \( \nu \) defined in (1.2). Moreover, \( W \) has a finite variance and the computational cost for simulating one copy of \( W \) has finite expectation.

**Proof.** The proof of Theorem 2.2.1 is presented in Section 3.4.

Next, we explain in several steps the construction of our estimator \( W \). In the meantime, we present and prove the three properties of \( W \), namely (1) unbiasedness, (2) finite variance and (3) finite expected computational cost. Finally, we present Algorithm 2 (on page 16) to provide explicit simulation strategy for generating \( W \). We start with preliminaries, obtaining a probabilistic representation of solution \( u(x,t) \) in (1.1).

### 3 Construction of estimator \( W \)

#### 3.1 Preliminaries: Probabilistic representation of solution \( u(x,t) \)

We start by noting that fixing \( \mu \in \mathcal{L}_1 \), the solution \( u(x,t) \) to the PDE in (1.1) and certain \( d \)-dimensional diffusion process can be connected through the Feynman-Kac formula [23, 24].
Theorem 3.1.1 (Feynman-Kac formula). Fix $x \in \mathbb{R}^d$, $t \in \mathbb{R}^+$ and functions $\sigma(\cdot), f(\cdot)$ satisfying Assumption A2. For any $\mu(\cdot) \in \mathcal{L}_1$, we define $g(x,t;\mu): \mathcal{L}_1 \to \mathbb{R}$ to be

$$g(x,t;\mu) = \mathbb{E}[f(X_t)],$$

where the expectation is taken with respect to the process $X_s$, the $d$-dimensional diffusion starting at $X_0 = x$ satisfying (i.e., a unique strong solution) the following SDE

$$dX_s = \mu(X_s)dt + \sigma(X_s)dB_s \quad \text{for} \quad s > 0,$$

(3.1)

with $B_s$ being a $d'$-dimensional Brownian motion. Then, the solution $u(x,t)$ of (1.1) satisfies

$$u(x,t) = g(x,t;\mu).$$

(3.2)

Proof. For any $\mu \in \mathcal{L}_1$, the existence and uniqueness of strong solution $X_s, s \geq 0$ is guaranteed by Assumption A2 on the Lipschitz continuity of $\sigma(\cdot)$. The existence of $\mathbb{E}[f(X_t)]$ follows from Assumption A2 on the linear growth of $f(\cdot)$. Thus, $g(x,t;\mu)$ is well-defined for any $\mu \in \mathcal{L}_1$ and the Equation (3.2) follows from the Feynman-Kac formula. See, for example, Chapter 4.4 in [24] for details of the proof.

Based on Theorem 3.1.1 fixing any $\mu(\cdot) \in \mathcal{L}_1$, if we can build, for $1 \leq i \leq k$, estimators $Z_i(\mu)$ satisfying

$$\mathbb{E}[Z_i(\mu)] = g(x_i,t_i;\mu),$$

and estimator $W(\mu)$ satisfying

$$\mathbb{E}[W(\mu)] = G(\mathbb{E}[Z_1(\mu)],\ldots,\mathbb{E}[Z_k(\mu)]),$$

then we will have

$$\mathbb{E}[W(\mu)] = G(u(x_1,t_1),\ldots,u(x_k,t_k)),$$

where the expectation is taken with $\mu(\cdot) \in \mathcal{L}_1$ fixed. To account for the randomness of $\mu(\cdot)$, we can first sample $\mu(\cdot,\omega) \sim \mu$ to obtain a realization of the field $\mu(\cdot,\omega) \in \mathcal{L}_1$ (according to Lemma 2.1.2 $\mu(\cdot,\omega) \in \mathcal{L}_1$ almost surely for $\omega \in \Omega$) and then construct $W = W(\mu)$ based on this realization $\mu(\cdot,\omega)$. Thus, formally, we would have

$$\mathbb{E}[W] = \mathbb{E}_{\mu \sim \mu}[\mathbb{E}[W(\mu)]]$$

$$= \mathbb{E}_{\mu \sim \mu}[G(u(x_1,t_1),\ldots,u(x_k,t_k))]$$

$$= \mathbb{E}[G(u(x_1,t_1),\ldots,u(x_k,t_k))]$$

$$= \nu,$$

(3.3)

which is the desired property of $W$. To make the derivations in (3.3) precise, we add a technical lemma on measurability of $G(u(x_1,t_1),\ldots,u(x_k,t_k))$(i.e, the well-definedness of $\nu$ in (1.2)). The proof is rather technical and not essential for our method so we will omit it.
Lemma 3.1.2. Under Assumptions $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3$ there exists a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $G(u(x_1, t_1), \ldots, u(x_k, t_k))$ is measurable.

In the following sections, we give details on the construction of $Z$ and $W$ along with their useful properties to complete the derivation in (3.3).

3.2 Step 1: Unbiased estimator $Z_i(\mu)$ for $g(x_i, t_i; \mu)$

3.2.1 Variance reduction and bias removal

Following the above discussion, given $x, t$ and $\mu \in \mathcal{L}_1$, we first want to construct estimator $Z(\mu)$ with $\mathbb{E}[Z(\mu)] = g(x, t; \mu) = \mathbb{E}[f(X_t)]$ where $X_s$ is unique strong solution to the following SDE

\[
\begin{align*}
    dX_s &= \mu(X_s)dt + \sigma(X_s)dB_s \quad \text{for} \quad s > 0 \\
    X_0 &= x
\end{align*}
\] (3.4)

To estimate value $\mathbb{E}[f(X_t)]$, one way is to solve for the SDE (3.4) numerically using various discretization schemes (e.g., Euler scheme, Milstein scheme, see [26]) and then plug it into function $f(\cdot)$ as one copy of the “plug-in” estimator. However, such estimator usually has bias and suboptimal computational cost versus variance ratio.

In order to reduce the variance of our estimator, we review the multilevel Monte Carlo method [27, 3, 39, 18]. Without loss of generality, from now on, we assume $t = 1$ and we are interested in estimating $\mathbb{E}[f(X_1)]$. Let $X_n(1)$ be the numerical solution of the SDE at $t = 1$ based on a level $n$ discretization scheme, and a random variable $\Delta_n$ should be constructed to satisfy

\[\mathbb{E}\Delta_n = \mathbb{E}f(X_{n+1}(1)) - \mathbb{E}f(X_n(1)).\] (3.5)

Then, a multilevel Monte Carlo(MLMC) estimator $Z_{\text{MLMC}}$ for $\mathbb{E}[f(X_1)]$ has the form

\[
Z_{\text{MLMC}} = \sum_{n=0}^{N} \frac{1}{M_n} \sum_{i=1}^{M_n} \Delta_n^{(i)} + \frac{1}{M_0} \sum_{i=1}^{M_0} f(X_0^{(i)}(1)),
\]

where $\{\Delta_n^{(i)}\}_{1 \leq i \leq M_n}$ are I.I.D. copies of $\Delta_n$, $\{X_0^{(i)}(1)\}_{1 \leq i \leq M_0}$ are I.I.D. copies of base level estimator and $N$ is a fixed and positive integer large enough so that the bias

\[|\mathbb{E}Z_{\text{MLMC}} - \mathbb{E}f(X_1)| = |\sum_{n=0}^{N} \mathbb{E}f(X_{n+1}(1)) - f(X_n(1)) + \mathbb{E}f(X_0(1)) - \mathbb{E}f(X_1)|
\]

\[= |\mathbb{E}f(X_{N+1}(1)) - \mathbb{E}f(X_1)|
\]

is small enough. The optimal choice of $M_n$ depends on the variance and computational cost of $\Delta_n$ and the construction of $\Delta_n$’s satisfying (3.5) usually involves variance reduction, an
important technique in MLMC (see [15, 16]). For example, in [15], \( \Delta_n \) is constructed by
taking \( f(X_{n+1}(1)) - f(X_n(1)) \), the difference of two numerical solutions produced by Euler
scheme under the same Brownian sample path. Through this coupling technique, it is shown
that \( \mathbb{E}\Delta_n^2 \) can be reduced to order \( O(\Delta t_n^2) \), as opposed to \( O(\Delta t_n^{1/2}) \) in the standard Euler
scheme, where \( \Delta t_n \) is the step size of discretization scheme. For our construction, in the
next subsection, we introduce the antithetic construction of \( \Delta_n \), which further reduces \( \mathbb{E}\Delta_n^2 \)
to order \( O(\Delta t_n^2) \).

### 3.2.2 Antithetic multilevel Monte Carlo and its properties

The original antithetic multilevel Monte Carlo construction of \( \Delta_n \) is proposed in [17] for
multidimensional SDEs. In our case, we make modifications to the numerical methods in
[17] to approximate the random field \( \mu \) since we cannot sample \( \mu \) exactly in practice.
Specifically, given \( V_1, V_2, \ldots \) and \( \mu(\cdot) \) in the form of (2.1) in Assumption A1, for some
\( \gamma > 0 \) (to be specified later in the proof section), we let
\[
\mu^{(n)}(\cdot) \triangleq \sum_{i=1}^{2^n \gamma} \frac{i^2}{2^n} \cdot V_i \cdot \psi_i(\cdot)
\]
be the level \( n \) approximation of \( \mu \). It follows from Lemma 2.1.2 that \( \mu^{(n)} \) belongs in \( L_1 \) with
bounding number \( L_1 \). We now describe our methods based on [17] in order to construct \( \Delta_n \).

**Definition 3.2.1.** Let \( \Delta t_n \triangleq 2^{-n} \) and \( t^n_k \triangleq k\Delta t_n \). Let \( \Delta B^n_k \triangleq B(t^n_{k+1}) - B(t^n_k) \) be the
Brownian increments from the \( B_t \) in SDE (3.4). We denote and \( \Delta B^n_{j,k} \triangleq B_j(t^n_{k+1}) - B_j(t^n_k) \)
to be the \( j \)th component of \( \Delta B^n_k \) where \( B_j(t) \) is the \( j \)th component of \( B_t \). Finally, we define, for \( 1 \leq i, j \leq d \'
and \( i \neq j \), the approximation of the process \( \int_s^t (B_i(r) - B_j(s)) dB_j(r) \) as
\[
\bar{A}_{i,i}(s,t) \triangleq \frac{(B_i(t) - B_i(s))^2 - (t - s)}{2} \quad \text{and} \quad \bar{A}_{i,j}(s,t) \triangleq \frac{(B_i(t) - B_i(s))(B_j(t) - B_j(s))}{2}.
\]

Then \( X_{i,n}(\cdot) \), the \( i \)th dimension of our level \( n \) numerical solution \( X_n(\cdot) \), is defined by recursion:
\[
X_{i,n}(t^n_{k+1}) = X_{i,n}(t^n_k) + \mu^{(n)}_i(X_n(t^n_k)) \Delta t_n + \sum_{j=1}^{d'} \sigma_{ij}(X_n(t^n_k)) \Delta B^n_{j,k} + \\
\sum_{j=1}^{d'} \sum_{l=1}^{d} \sum_{m=1}^{d'} \frac{\partial \sigma_{ij}}{\partial x_l}(X_n(t^n_k)) \sigma_{lm}(X_n(t^n_k)) \bar{A}_{mj}(t^n_k, t^n_{k+1}),
\]
for \( 1 \leq i \leq d \) and \( 1 \leq k \leq 2^n \).
This recursion is a modification of the Euler scheme that would significantly reduce the variance of $\Delta_n$ which is defined as

$$
\Delta_n \triangleq \frac{1}{2} \left( f(X_{n+1}^f(1)) + f(X_{n+1}^a(1)) - f(X_n(1)) \right).
$$

(3.7)

We use the notation $X_{n+1}^f$ and $X_{n+1}^a$ consistent with [17] to represent the “fine” and “antithetic” approximation on level $n+1$ as opposed to the “coarse” level $n$ approximation $X_n$. In particular, $X_{n+1}^f(\cdot)$ is just $X_{n+1}(\cdot)$ and it uses the same underlying Brownian increments as $X_n(\cdot)$. The antithetic $X_{n+1}^a(\cdot)$ is defined similarly as in [17] with the “swapping” of Brownian increments.

**Definition 3.2.2.** For $0 \leq k \leq 2^{n+1} - 1$, define the sequence of antithetic Brownian increments \( \{\Delta B_{2m+1}^{n+1,a}\}_{0 \leq k \leq 2^{n+1} - 1} \) by

$$
\Delta B_{2m+1}^{n+1,a} \triangleq \Delta B_{2m+1}^{n+1}, \quad \text{and} \quad \Delta B_{2m+1}^{n+1,a} \triangleq \Delta B_{2m+1}^{n+1}
$$

for $0 \leq m \leq 2^n - 1$ (3.8)

along with modification for \( \{\tilde{A}_{i,j}(t_{k+1}^{n+1}, t_{k+1}^{n+1})\}_{0 \leq k \leq 2^{n+1} - 1} \) as

$$
\tilde{A}_{i,j}(t_{k+1}^{n+1}, t_{k+1}^{n+1}) \triangleq \frac{(\Delta B_{k}^{n+1,a})^2 - \Delta t_{n+1}}{2} \quad \text{and} \quad \tilde{A}_{i,j}(t_{k+1}^{n+1}, t_{k+1}^{n+1}) \triangleq \frac{\Delta B_{i,k+1}^{n+1,a} \Delta B_{j,k+1}^{n+1,a}}{2}.
$$

for $1 \leq i, j \leq d'$ and $i \neq j$. Notice that we also have the relations that

$$
B_{n+1,a}(t_{k+1}^{n}) - B_{n+1,a}(t_{k}^{n}) = B(t_{k+1}^{n}) - B(t_{k}^{n})
$$

(3.9)

by summing the equations in (3.8). Then, the antithetic approximation $X_{n+1}^a(\cdot)$ follows the recursion defined in (3.6):

$$
X_{i,n+1}^{a}(t_{k+1}^{n+1}) = X_{i,n+1}^{a}(t_{k+1}^{n}) + \mu_{i}^{(n+1)}(X_{n+1}^{a}(t_{k+1}^{n+1})) \Delta t_{n+1} + \sum_{j=1}^{d'} \sigma_{ij}(X_{n+1}(t_{k+1}^{n+1})) \Delta B_{j,k}^{a,(n+1)}
$$

$$
+ \sum_{j=1}^{d'} \sum_{l=1}^{d} \sum_{m=1}^{d'} \frac{\partial \sigma_{ij}}{\partial x_{l}}(X_{n+1}(t_{k+1}^{n+1})) \sigma_{lm}(X_{n+1}(t_{k+1}^{n+1})) \tilde{A}_{mj}(t_{k+1}^{n+1}, t_{k+1}^{n+1}).
$$

Now, we discuss several important properties of our antithetic numerical approximation. First, the expectation of $\Delta_n$ satisfies (3.5).

**Lemma 3.2.3.** Fixing $V_1, V_2, \ldots$ such that $\mu \in L_1$ and $\{\mu^{(n)}\}_{n \geq 0} \subset L_1$, we have

$$
\mathbb{E} \Delta_n = \mathbb{E} f(X_{n+1}(1)) - \mathbb{E} f(X_n(1))
$$

as in (3.5).
Proof. Since the Brownian increments are I.I.D., solutions $X_{n+1}^f(1)$ and $X_{n+1}^a(1)$ produced by two recursions under the swapping of Brownian increments still follow the same distribution. Thus, we have

$$\mathbb{E}f(X_{n+1}^a(1)) = \mathbb{E}f(X_{n+1}^f),$$

which implies (3.5) by the construction of $\Delta_n$ in (3.7).

Second, we have the following bound on $\mathbb{E}\|X_n(t) - X_t\|_\infty^4$.

**Lemma 3.2.4.** Fixing $\{V_i\}_{i \geq 0}$ and $L_1$ such that $\mu \in L_1$, $\{\mu^{(n)}\}_{n \geq 0} \subset L_1$ and $L_1 > 1$ is a common bounding number for $\{\mu^{(n)}\}_{n \geq 0}$ and $\mu$, let $X(\cdot)$ be the solution of the SDE in (3.1) and $X_n(\cdot)$ be the numerical approximation in (3.6), both defined for $0 \leq t \leq 1$. Then, under Assumptions $A1 \& A2$, there exist constants $0 < \epsilon < 2$, $\gamma > 0$ and $C > 1$ such that

$$\mathbb{E}\|X_n(t) - X_t\|_\infty^4 \leq e^{CL_1} \cdot \Delta t_n^{2-\epsilon}.$$  \hspace{1cm} (3.10)

**Proof.** The proof is in Section 5.

**Corollary 3.2.5.** Under the assumptions of Lemma 3.2.4 as well as Assumption $A2$, we have

$$\lim_{n \to \infty} \mathbb{E}|f(X_n(1)) - f(X_1)|^2 = 0,$$

and consequently,

$$\lim_{n \to \infty} \mathbb{E}f(X_n(1)) = \mathbb{E}f(X_1).$$

**Proof.** It follows from Lemma 3.2.4, Assumption $A2$ and Cauchy-Schwarz inequality, we have

$$\mathbb{E}|f(X_n(1)) - f(X_1)|^2 \leq L^2 \mathbb{E}\|X_n(1) - X_1\|_\infty^2 \leq L^2 \sqrt{\mathbb{E}\|X_n(1) - X_1\|_\infty^4}$$

which tends to zero as $n$ goes to infinity by Lemma 3.2.4.

**Remark.** For $\mu \in L_1$ with a bounding number $L_1$, the results in [17] typically would show that $\mathbb{E}\|X_n(t) - X_t\|_\infty^4 = O(\Delta t_n^2)$, a standard error bound for numerical approximation of SDE deduced from Gronwall’s inequality. In particular, the bounding constant has the form

$$\mathbb{E}\|X_n(t) - X_t\|_\infty^4 \leq e^{CL_1^p} \Delta t_n^2,$$  \hspace{1cm} (3.11)

for some constant $C$. However, in our problem $\mu$ is random and $e^{L_1^p}$ is not guaranteed to have a finite expectation for $p > 1$. Thus, instead of using Gronwall’s inequality, we explore the rough path technique in [4] to develop the bound (3.10), where we substitute the $e^{CL_1^p}$ term by the term $e^{CL_1}$ by giving up $\epsilon$ order from $\Delta t_n^2$ in (3.11).
3.2.3 Construction and properties of $Z(\mu)$

After the discussion on variance reduction, we discuss techniques on bias removal. As we have pointed out, $Z_{\text{MLMC}}$ is a biased estimator because it is based on the numerical solutions of SDE. Thus, as the level of discretization gets finer, the bias can be reduced but never eliminated. In order to remove the bias, we construct $Z(\mu)$ based on $Z_{\text{MLMC}}$ using the following technique.

**Definition 3.2.6 (Definition of $Z(\mu)$).** Fixing $V_1, V_2, \ldots$ such that $\mu \in L_1$ and $\{\mu^{(n)}\}_{n \geq 0} \subset L_1$, let $N \sim \text{Geom}(1 - 2^{-\theta}), N \geq 0$ be an independent geometric random variable with success rate $1 - 2^{-\theta}$ for some $\theta > 0$. We defined $Z(\mu)$ to be:

$$Z(\mu) \triangleq f(X_{n_0}(1)) + \frac{\Delta N + n_0}{p_N},$$  \hspace{1cm} (3.12)

where $n_0 \geq 0$ is the base level estimator, $X_{n_0}(\cdot), \Delta_n$ are defined in (3.7) using the antithetic numerical scheme and $p_n \triangleq \mathbb{P}(N = n) = (1 - 2^{-\theta})(2^{-\theta n})$ is the probability mass function of $N$.

The choice of $\theta > 0$ will be specified later in the proof section. Now, we give the simulation scheme of $Z$ in Algorithm 1 below. In particular, we assume random variables $V_1, V_2, \ldots$ are given (i.e, $\mu^{(n)}$ can be constructed for all $n$), even though only a finite number of them are needed.

**Algorithm 1:** Generate $Z(\mu)$

1. **Input** parameters $x \in \mathbb{R}^d, t > 0, n_0 \geq 0, \theta > 0$ and $\gamma > 0$
2. Generate geometric random variable $N \sim \text{Geom}(1 - 2^{-\theta})$.
3. Construct $\mu^{(N+n_0+1)}(\cdot) \triangleq \sum_{i=1}^{(N+n_0+1)\gamma} \frac{\Delta_i}{\gamma} \cdot V_i \cdot \psi_i(\cdot)$ and similarly $\mu^{(N+n_0)}, \mu^{(n_0)}$. Then, use them to compute $X^{f}_{N+n_0+1}(t), X^{a}_{N+n_0+1}(t), X^{f}_{N+n_0}(t)$ and $X^{a}_{n_0}(t)$ through recursion defined in Definition 3.2.1 and Definition 3.2.2 with initial point $x$.
4. Set $\Delta N + n_0 = 0.5(f(X^{f}_{N+n_0+1}(t)) + f(X^{a}_{N+n_0+1}(t))) - f(X^{f}_{N+n_0}(t))$
5. **Output** $Z(x, t; \mu) = \frac{\Delta N + n_0}{p_N} + f(X_{n_0}(t))$

In practice, a larger value of $n_0$ gives lower variance of $Z$ at the cost of a higher computational budget. Also, during implementation, we can use the same Brownian path to generate $X_{n_0}(1)$ and $\Delta N + n_0$ in (3.12) to reduce the computational cost. However, we will not go into the details of implementation. Instead, we present several important properties of $Z$. We start with the unbiasedness.

**Lemma 3.2.7.** Under the assumptions of Lemma 3.2.4 as well as Assumption A2, we have

$$\mathbb{E}[Z(\mu)] = g(x, 1; \mu),$$

where $g(x, 1; \mu)$ is defined in Theorem 3.1.1.
Proof. It follows from the definition of $Z$ that we have

$$
\mathbb{E}[Z(\mu)] = \mathbb{E}f(X_{n_0}(1)) + \mathbb{E}\frac{\Delta_{N+n_0}}{p_N}
$$

$$
= \mathbb{E}f(X_{n_0}(1)) + \mathbb{E}_N[\frac{\Delta_{N+n_0}}{p_N} | N]
$$

$$
= \mathbb{E}f(X_{n_0}(1)) + \sum_{n=0}^{\infty} \mathbb{E}\frac{\Delta_{n+n_0}}{p_n} \cdot p_n
$$

$$
= \mathbb{E}f(X_{n_0}(1)) + \sum_{n=0}^{\infty} \mathbb{E}\Delta_{n+n_0}
$$

$$
= \mathbb{E}f(X_{n_0}(1)) + \sum_{n=0}^{\infty} \mathbb{E}f(X_{n+n_0+1}(1)) - \mathbb{E}f(X_{n+n_0}(1))
$$

$$
= \lim_{n \to \infty} \mathbb{E}f(X_{n}(1))
$$

$$
= \mathbb{E}f(X_1).
$$

The third equality follows from the independence of $N$ with our numerical solution, the fifth equality from Equation (3.5) and the last equality from Corollary 3.2.5. The proof now follows from Theorem 3.1.1 on Feynman-Kac formula and definition of $g(x, t; \mu)$. □

This shows $Z(\mu)$ is an unbiased estimator for $g(x, 1; \mu)$ as claimed. Next, instead of directly showing a bound on the variance of $Z(\mu)$, we provide a stronger bound on the fourth moment of $Z(\mu)$ which becomes useful for proving the finite variance property of $W(\mu)$ later on.

**Lemma 3.2.8.** Fixing $\{V_i\}_{i \geq 0}$ and $L_1 > 1$ such that $\mu \in \mathcal{L}_1$, $\{\mu^{(n)}\}_{n \geq 0} \subset \mathcal{L}_1$ and $L_1$ is a common bounding number for $\{\mu^{(n)}\}_{n \geq 0}$ and $\mu$, there exist $0 < \delta < 4$ and $C > 1$ such that

$$
\mathbb{E}\Delta_n^4 \leq e^{CL_1} \Delta_n^{4-\delta},
$$

(3.13)

and

$$
\mathbb{E}|f(X_{n_0}(1))|^4 \leq \mathcal{P}(L_1)
$$

(3.14)

for some polynomial function $\mathcal{P}(x) > 1$ for $x > 1$.

**Proof.** The proof is in Section 5. □

**Lemma 3.2.9.** Under the assumptions of Lemma 3.2.8, choosing $\theta$ so that $3\theta < 4 - \delta$ for $\delta$ specified in Lemma 3.2.8, then the estimator $Z$ defined in (3.12) satisfies

$$
\mathbb{E}[Z^4(\mu)] \leq e^{CL_1},
$$

for some constant $C > 1$.
Proof. The elementary inequality
\[
\sum_{n=1}^{N} |a_n|^p \leq N^{p-1} \sum_{n=1}^{N} |a_n|^p,
\]
gives
\[
\mathbb{E} Z^4(\mu) \leq 8 \mathbb{E} \frac{\Delta_4^{N+n_0}}{p_N} + 8 \mathbb{E} |f(X_{n_0}(1))|^4
\]
\[
= 8 \mathbb{E} \left[ \frac{\Delta_4^{N+n_0}}{p_N} \cdot p_n + 8 \mathbb{E} |f(X_{n_0}(1))|^4 \right]
\]
\[
\leq 8 \left( \frac{e^{CL_1}}{(1-2^{-\theta})^3} \sum_{n=0}^{\infty} \frac{\Delta_4^{4-\delta}}{\Delta_4^m} + \mathcal{P}(L_1) \right)
\]
\[
\leq e^{C' L_1}.
\]
for some constant $C' > 1$ appropriately chosen. In particular, the last inequality follows from $4 - \delta > 3\theta$, $L_1 > 1$ and the fact that $\mathcal{P}(|x|) < e^{a|x|+b}$ for some appropriately chosen $a, b$ and $e^{ax+b} < e^{cx}$ for some some appropriately chosen $c$ if $x > 1$. The second inequality follows from Lemma 3.2.8.

The third property of $Z(\mu)$ is its finite expected computational cost. Due to the additional randomization, the computational cost for generating $Z(\mu)$ is a random variable. Specifically, if we use $\text{cost}^Z$ to denote the computational cost for generating $Z$ and $\text{cost}_n$ for generating $X_n(1)$, respectively, then we have
\[
\text{cost}^Z = \text{cost}_{n_0} + \text{cost}_{N+n_0} + 2 \text{cost}_{N+n_0+1},
\]
due to the computation of $X_{n_0}(1), X_{N+n_0}(1), X_{N+n_0+1}^f$ and $X_{N+n_0+1}^g$ in the construction of $Z$.

Lemma 3.2.10. Let $\theta$ and $\gamma$ be choosen so that $\theta > 1 + \gamma$, then the computational cost for generating $Z$ has finite expectation. That is,
\[
\mathbb{E}(\text{cost}^Z) < \infty.
\]

Proof. Consider the $\text{cost}_n$ for generating $X_n(1)$. For each fixed $n$, we need to generate $2^n$ number of Brownian increments and $2^m$ number of Gaussian random variables for the construction of $\mu^{(n)}$. Then, to compute $X_n(1)$, we need to carry out $2^n$ number of iterations to iterate from $X_n(0)$ to $X_n(1)$ and during iteration $k$, $O(2^m)$ steps of computation are
performed to evaluate $\mu^{(n)}(X_{n}(t_{k}^{n}))$ and consequently $X_{n}(t_{k+1}^{n})$. Thus the computational cost for evaluating $X_{n}(1)$ satisfies

$$\text{cost}_{n} = O((1+\gamma)^{n}) \leq C(1+\gamma)^{n},$$

for some constant $C$. Therefore, from (3.16) and the fact that $P(N = n) = (1 - 2^{-\theta}) \cdot 2^{-n}$, we have

$$\mathbb{E}(\text{cost}_{Z}) = \mathbb{E}(\text{cost}_{n_0}) + \mathbb{E}(\text{cost}_{N+n_0}) + 2\mathbb{E}(\text{cost}_{N+n_0+1})$$

$$= \text{cost}_{n_0} + \sum_{n=0}^{\infty} \mathbb{P}(N = n) \cdot \text{cost}_{n_0+n} + 2 \sum_{n=0}^{\infty} \mathbb{P}(N = n) \cdot \text{cost}_{n_0+n+1}.$$ 

$$\leq C(1+\gamma)^{n_0} + C(1+\gamma)^{n_0}(1 - 2^{-\theta}) \sum_{n=0}^{\infty} 2^{(1+\gamma)\theta n} + 2 \cdot C(1+\gamma)(n_0+1)(1 - 2^{-\theta}) \sum_{n=0}^{\infty} 2^{(1+\gamma-\theta)n}$$

$$< \infty,$$

where the series converge due to the assumption $\theta > 1 + \gamma$. \qed

### 3.3 Step 2: Unbiased estimator $W(\mu)$ for $G(\mathbb{E}[Z_1(\mu)], \ldots, \mathbb{E}[Z_k(\mu)])$

#### 3.3.1 Construction of $W(\mu)$

After the construction of $Z(\mu)$ in Step 1, in Step 2, we construct a random variable $W(\mu)$ such that

$$\mathbb{E}W(\mu) = G(\mathbb{E}Z_1(\mu), \ldots, \mathbb{E}Z_k(\mu))$$

with $Z_1(\mu), \ldots, Z_k(\mu)$ constructed from Step 1 satisfying $\mathbb{E}Z_i(\mu) = g(x_i, t; \mu)$. Our strategy is to use a method recently developed in [5]. For the ease of presentation, we construct and illustrate the properties of an unbiased estimator $W(\mu)$ for $G(\mathbb{E}(Z(\mu))$ given $\mu \in \mathcal{L}_1$ where $G(\cdot): \mathbb{R} \rightarrow \mathbb{R}$. Our method for the case where $G(\cdot): \mathbb{R}^k \rightarrow \mathbb{R}$ can be easily generalized and we will leave the details for implementation in Algorithm 2.

**Definition 3.3.1 (Definition of $W(\mu)$).** Fixing $V_1, V_2, \ldots$ such that $\mu \in \mathcal{L}_1$ and $\{\mu^{(n)}\}_{n \geq 0} \subset \mathcal{L}_1$, for an integer $n > 0$, let $\{Z_j(\mu)\}_{1 \leq j \leq 2^{n+1}}$ be I.I.D. copies of random variables $Z(\mu)$. Define random variable $\tilde{\Delta}_n$ to be

$$\tilde{\Delta}_n \triangleq G\left( \sum_{j=1}^{2^{n+1}} Z_j(\mu) \right) - \frac{1}{2} \left( G\left( \sum_{j=1}^{2^{n}} Z_j(\mu) \right) + G\left( \sum_{j=2^{n+1}}^{2^{n+1}} Z_j(\mu) \right) \right).$$

(3.17)

Then, fix any integer $n_1 > 1$ and construct $W(\mu)$ from I.I.D. copies $\{Z_j(\mu)\}_{1 \leq j \leq 2^{n_1+n+1}}$ by letting

$$W(\mu) = \frac{\tilde{\Delta}_{n_1+n+1}}{\bar{p}_n} + G\left( \sum_{j=1}^{2^{n_1}} Z_j(\mu) \right).$$

(3.18)
where $\tilde{N} \sim Geom(1 - 2^{-1.5})$ is an independent geometric random variable with success rate $1 - 2^{-1.5}$ and $\tilde{p}_n \triangleq \mathbb{P}(\tilde{N} = n) = 2^{-1.5n}(1 - 2^{-1.5})$.

We now describe Algorithm 2 for generating $W(\mu)$ assuming random variables $V_1, V_2, \ldots$ are given (i.e, $\mu^{(n)}$ can be constructed for all $n$).

**Algorithm 2**: Generate $W(\mu)$

1. **Input** parameters $x_i \in \mathbb{R}^d, t_i > 0$ for $1 \leq i \leq k$, $n_1 > 1, n_0 \geq 0, \theta > 0$ and $\gamma > 0$
2. Generate $\tilde{N} \sim Geom(1 - 2^{-1.5})$
3. for $i$ in $1 : k$
   4. for $j$ in $1 : 2^{\tilde{N} + n_1 + 1}$
      5. Generate I.I.D. $Z_{ij}(x_i, t_i; \mu)$ from Algorithm 1 with input $x_i, t_i, n_0, \theta$ and $\gamma$
   7. end
8. Set $\rho(a, b) \triangleq G\left(\frac{1}{b-a+1}\sum_{j=a}^{b} Z_{ij}, \ldots, \frac{1}{b-a+1}\sum_{j=a}^{b} Z_{kj}\right)$
9. Set $\tilde{\Delta}_{\tilde{N}+n_1+1} = \rho(1, 2^{\tilde{N}+n_1+1}) - \frac{1}{2}(\rho(1, 2^{\tilde{N}+n_1}) + \rho(2^{\tilde{N}+n_1} + 1, 2^{\tilde{N}+n_1+1}))$
10. Output $W(\mu) = \frac{\tilde{\Delta}_{\tilde{N}+n_1+1}}{p_{\tilde{N}}} + \rho(1, 2^{n_1})$

**Remark.** Notice that if we denote $N_{ij}$ to be the geometric random variable generated during the construction of $Z_{ij}$ in Algorithm 1 and let

$$m = \max\{N_{ij}, 1 \leq i \leq k, 1 \leq j \leq 2^{\tilde{N}+n_1+1}\} \quad \text{and} \quad M = [2^{(m+n_0+1)\gamma}],$$

then we only need to generate $V_1, \ldots, V_M$ in the construction of $\mu(\cdot)$ because they are sufficient for carrying out the recursion in Algorithm 1.

**3.3.2 Properties of $W(\mu)$**

We will show several properties of $W(\mu)$. We start with the unbiasedness.

**Lemma 3.3.2.** Under the assumptions of Lemma 3.2.4 as well as Assumptions A2-A3, we have

$$\mathbb{E}W(\mu) = G(\mathbb{E}Z(\mu)).$$

**Proof.** According to Lemma 3.2.9 and strong law of large numbers, we have

$$\lim_{n \to \infty} \mathbb{E}\left| \sum_{j=1}^{n} \frac{Z_j(\mu)}{2^n} - \mathbb{E}Z(\mu) \right| = 0,$$

which implies

$$\lim_{n \to \infty} \mathbb{E}\left| G\left( \sum_{j=1}^{n} \frac{Z_j(\mu)}{2^n} \right) - G(\mathbb{E}Z(\mu)) \right| = 0$$
by Assumption [A3] on the bound of $\| \frac{\partial G}{\partial x_1} \|_\infty$. Therefore, we have

$$\lim_{n \to \infty} \mathbb{E}G\left( \frac{\sum_{j=1}^{2^n} Z_j(\mu)}{2^n} \right) = G(\mathbb{E}Z(\mu))$$

as $n \to \infty$. Now, since

$$\mathbb{E} \tilde{\Delta}_n = \mathbb{E}G\left( \frac{\sum_{j=1}^{2^n+1} Z_j(\mu)}{2^{n+1}} \right) - \mathbb{E}G\left( \frac{\sum_{j=1}^{2^n} Z_j(\mu)}{2^n} \right),$$

the rest of the proof follows similarly as in the proof of Lemma 3.2.7.

Next, we proceed to show that $W(\mu)$ also has finite variance.

**Lemma 3.3.3.** Under the assumptions of Lemma 3.2.4 as well as Assumptions [A2], [A3], the $\Delta_n$ defined in (3.17) satisfies

$$\mathbb{E}(\tilde{\Delta}_n)^2 \leq e^{CL_1} \Delta t_n^2$$

where $C > 1$ is some fixed constant.

**Proof.** Fixing $n > 0$, the second order Taylor expansion of $G(\cdot)$ around $\mathbb{E}Z(\mu)$ in (3.17), we have,

$$\tilde{\Delta}_n = G\left( \frac{\sum_{j=1}^{2^{n+1}} Z_j(\mu)}{2^{n+1}} \right) - \frac{1}{2}\left( G\left( \frac{\sum_{j=1}^{2^n} Z_j(\mu)}{2^n} \right) + G\left( \frac{\sum_{j=1}^{2^n+1} Z_j(\mu)}{2^{n+1}} \right) \right)$$

$$= G(\mathbb{E}Z(\mu)) + G'(\mathbb{E}Z(\mu))\left( \sum_{j=1}^{2^n} Z_j - \mathbb{E}Z(\mu) \right) + \frac{1}{2} G''(\xi_1)\left( \sum_{j=1}^{2^n} Z_j - \mathbb{E}Z(\mu) \right)^2$$

$$- \frac{1}{2} G(\mathbb{E}Z(\mu)) - \frac{1}{2} G'(\mathbb{E}Z(\mu))\left( \sum_{j=1}^{2^n} Z_j - \mathbb{E}Z(\mu) \right) - \frac{1}{4} G''(\xi_2)\left( \sum_{j=1}^{2^n} Z_j - \mathbb{E}Z(\mu) \right)^2$$

$$- \frac{1}{2} G(\mathbb{E}Z(\mu)) - \frac{1}{2} G'(\mathbb{E}Z(\mu))\left( \sum_{j=1}^{2^n+1} Z_j - \mathbb{E}Z(\mu) \right) - \frac{1}{4} G''(\xi_3)\left( \sum_{j=1}^{2^n+1} Z_j - \mathbb{E}Z(\mu) \right)^2$$

$$= \frac{G''(\xi_1)}{2}\left( \sum_{j=1}^{2^n} Z_j(\mu) \right) - \mathbb{E}Z(\mu)^2 - \frac{G''(\xi_2)}{4}\left( \sum_{j=1}^{2^n} Z_j(\mu) \right)^2$$

$$- \frac{G''(\xi_3)}{4}\left( \sum_{j=2^n+1}^{2^n+1} Z_j(\mu) \right) - \mathbb{E}Z(\mu)^2,$$

where all the first order terms cancel out. As in the mean value theorem, $\xi_1$ is a random variable between $Z$ and $\sum_{j=1}^{2^n+1} Z_j$, $\xi_2$ between $Z$ and $\sum_{j=1}^{2^n} Z_j$, and $\xi_3$ between $Z$ and $\sum_{j=2^n+1}^{2^n+1} Z_j$.

Thus, it follows from (3.15) and Assumption [A3] we have

$$|\tilde{\Delta}_n|^2 \leq \frac{3L^2}{4}\left( \left( \sum_{j=1}^{2^n+1} \left( Z_j(\mu) - \mathbb{E}Z(\mu) \right) \right)^4 + \frac{1}{4} \left( \sum_{j=1}^{2^n} \left( Z_j(\mu) - \mathbb{E}Z(\mu) \right) \right)^4 + \frac{1}{4} \left( \sum_{j=2^n+1}^{2^n+1} \left( Z_j(\mu) - \mathbb{E}Z(\mu) \right) \right)^4 \right).$$

(3.19)
However, the \((Z_j(\mu) - \mathbb{E}Z(\mu))\) are I.I.D. random variables with mean 0. Thus, when we write out the expansion in (3.19) and take expectation, the terms with odd power will vanish. In particular,

\[
\mathbb{E}[(Z_i(\mu) - \mathbb{E}Z(\mu)^2(Z_j(\mu) - \mathbb{E}Z(\mu))(Z_k(\mu) - \mathbb{E}Z(\mu))] = 0 \\
\mathbb{E}[(Z_i(\mu) - \mathbb{E}Z(\mu)^3(Z_j(\mu) - \mathbb{E}Z(\mu))] = 0 \\
\mathbb{E}[(Z_i(\mu) - \mathbb{E}Z(\mu))(Z_j(\mu) - \mathbb{E}Z(\mu))(Z_k(\mu) - \mathbb{E}Z(\mu)))] = 0.
\]

Thus, taking expectation in (3.19) gives

\[
\mathbb{E} \tilde{\Delta}_n^2 \leq \frac{3L^2}{2^{m+1}} \mathbb{E} \left[ \frac{1}{4} \left( \sum_{j=1}^{2^{n+1}} Z_j(\mu) - \mathbb{E}Z(\mu) \right)^4 + \left( \sum_{j=1}^{2^n} Z_j(\mu) - \mathbb{E}Z(\mu) \right)^4 + \left( \sum_{j=2^{n+1}}^{2^{n+1}} Z_j(\mu) - \mathbb{E}Z(\mu) \right)^4 \right]
\]

\[
\leq \frac{3L^2}{2^{m+1}} \mathbb{E} \left[ \frac{5}{4} \sum_{j=1}^{2^{n+1}} (Z_j(\mu) - \mathbb{E}Z(\mu))^4 + \frac{6}{4} \sum_{1 \leq i < j \leq 2^{n+1}} (Z_i(\mu) - \mathbb{E}Z(\mu))^2 (Z_j(\mu) - \mathbb{E}Z(\mu))^2 \\
+ 6 \sum_{1 \leq i < j \leq 2^n} (Z_i(\mu) - \mathbb{E}Z(\mu))^2 (Z_j(\mu) - \mathbb{E}Z(\mu))^2 + 6 \sum_{2^{n+1} \leq i < j \leq 2^{n+1}} (Z_i(\mu) - \mathbb{E}Z(\mu))^2 (Z_j(\mu) - \mathbb{E}Z(\mu))^2 \right]
\]

\[
\leq C \left( \begin{array}{c} 2^{n+1} \\ 2 \end{array} \right) 2^{-4n} \cdot \mathbb{E}(Z(\mu) - \mathbb{E}Z(\mu))^4
\]

for some constants \(C > 1\). The last line follows from Cauchy-Schwarz inequality.

However, it is straightforward to show that

\[
\frac{n}{2} = O(n^2).
\]

Thus, we have

\[
\left( \begin{array}{c} 2^{n+1} \\ 2 \end{array} \right) 2^{-4n} \leq C \Delta t_n^2
\]

for some constant \(C > 1\). Furthermore, Hölder’s inequality and Jensen’s inequality provide methods to bound \(\mathbb{E}(Z(\mu) - \mathbb{E}Z(\mu))^4\) using \(\mathbb{E}Z^4(\mu)\). Consequently, we can use the bound from Lemma 3.2.9 on \(\mathbb{E}Z^4(\mu)\) to obtain

\[
\mathbb{E}(Z(\mu) - \mathbb{E}Z(\mu))^4 \leq e^{CL_1}
\]

for some constant \(C > 1\), where we have used our assumption that \(e^{ax+b} < e^{cx}\) for appropriately chosen \(c\) if \(x > 1\) and the fact that \(L_1 > 1\). Combining these observations together, we can find a constant \(C > 1\) such that

\[
\mathbb{E}(\tilde{\Delta}_n)^2 \leq e^{CL_1} \Delta t_n^2.
\]

\[\square\]
Lemma 3.3.4. Under the assumptions of Lemma 3.2.4 as well as Assumptions A2-A3, the $W$ defined in (3.18) satisfies
\[ \mathbb{E} W^2(\mu) \leq e^{CL_1} \]
for some constant $C > 1$.

Proof. Using bounds on the fourth moment of $Z$ in Lemma 3.2.9, the linear growth condition of $G(\cdot)$ in Assumption A3 and the Cauchy-Schwarz inequality, we have
\[
\mathbb{E} |G(\frac{\sum_{j=1}^{2n_1} Z_j(\mu)}{2^{n_1}})|^2 \leq \mathbb{E}(|G(0)| + L|\sum_{j=1}^{2n_1} Z_j(\mu)|)^2
\leq |G(0)|^2 + 2|G(0)|L\mathbb{E}|Z(\mu)| + L^2\mathbb{E}Z^2(\mu)
\leq C + Ce^{CL_1}
\tag{3.20}
\]
for some constant $C > 1$. Now, using (3.15), (3.20), Lemma 3.3.3 and the fact that $\bar{N} \sim \text{Geom}(1 - 2^{-1.5})$, we have
\[
\mathbb{E} W^2(\mu) \leq 2\mathbb{E}\left(\frac{\bar{N}_{N+n_1}^2}{p_{n_1}^2} + |G(\frac{\sum_{j=1}^{2n_1} Z_j}{2^{n_1}})|^2\right)
= 2\mathbb{E}_N \mathbb{E}\left[\frac{\bar{N}_{N+n_1}^2}{p_{n_1}^2} |N| + 2\mathbb{E}|G(\frac{\sum_{j=1}^{2n_1} Z_j}{2^{n_1}})|^2\right]
= 2\sum_{n=0}^{\infty} \frac{\mathbb{E}_{N+n_1} \bar{N}_{n_1+1}^2}{p_n^2} + 2\mathbb{E}|G(\frac{\sum_{j=1}^{2n_1} Z_j}{2^{n_1}})|^2
\leq \frac{2e^{CL_1}}{(1 - 2^{-1.5})} \sum_{n=0}^{\infty} \frac{2^{-2n}}{2^{-1.5n}} + C + Ce^{CL_1}
\leq e^{c' L_1}
\]
for some appropriately chosen $C' > 1$. The first inequality follows from Lemma 3.3.3 and (3.20). The last line follows from the fact that for any $a, b$ and $c$, we can find $d$ such that $a + ce^{bx} < e^{dx}$ for $x > 1$.

\[\square\]

Finally, we discuss the computational cost for generating $W(\mu)$. Denote the cost of generating $W(\mu)$ by $\text{cost}_W$. Since we can use the first $2^{n_1}$ samples of $Z_j$ in the construction of $\bar{N}_{N+n_1}$ to construct $G(\sum_{j=1}^{2n_1} Z_j/2^{n_1})$, we only consider the cost induced by term $\bar{N}_{N+n_1}$,
\[
\text{cost}_W = \sum_{j=1}^{2N_{N+n_1}^2} \text{cost}_{Z_j}.
\]

Lemma 3.3.5. The total expected computational cost satisfies
\[ \mathbb{E}(\text{cost}_W) < \infty. \]
Proof. Using Wald’s identity, we have

\[ \mathbb{E}(\text{cost}_W) = 2^{n+1} \mathbb{E}(\text{cost}_{\tilde{Z}}) = 2^{n+1} \left( \sum_{n=0}^{\infty} 2^n \cdot 2^{-1.5n}(1 - 2^{-1.5}) \right) \mathbb{E}(\text{cost}_{\tilde{Z}}) < \infty, \]

where \( \mathbb{E}(\text{cost}_{\tilde{Z}}) < \infty \) follows from Lemma 3.2.10.

3.4 Proof of Main Theorem 2.2.1

Proof. We construct our estimator \( W \) as follows. We first sample \( \mu(\cdot) \) by sampling I.I.D. \( \{V_i\}_{i \geq 1} \) as in Assumption A1 and constructing as in (2.1) the field \( \mu(\cdot) \) (during implementation, we only need to sample a finite number of \( V_i \), see Remark 10). Next, using the constructed \( \mu(\cdot) \), we simulate \( W(\mu) \) according to Algorithm 1 since \( \mu(\cdot) \in L_1 \) almost surely by Lemma 2.1.2. It follows from Lemma 3.3.2 that \( W(\mu) \) satisfies

\[ \mathbb{E}[W(\mu)] = G(\mathbb{E}[Z_1(\mu)], ..., \mathbb{E}[Z_k(\mu)]), \]

and consequently

\[ \mathbb{E}[W(\mu)] = G(u(x_1, t_1), ..., u(x_k, t_k)), \]

according to Lemma 3.2.7 and Theorem 3.1.1, where the expectation is taken with \( \mu(\cdot) \in L_1 \) fixed. It then follows from Lemma 3.1.2 that

\[ \mathbb{E}[W] = \mathbb{E}_{\mu \sim \mu}[\mathbb{E}[W(\mu)]] = \mathbb{E}_{\mu \sim \mu}[G(u(x_1, t_1), ..., u(x_k, t_k))] = \mathbb{E}[G(u(x_1, t_1), ..., u(x_k, t_k))] = \nu, \]

which proves the unbiasedness of \( W \). To show the finite variance property of \( W \), we use Lemma 2.1.2 and Lemma 3.3.4 to obtain that

\[ \mathbb{E}W^2 = \mathbb{E}_{\mu \sim \mu}[\mathbb{E}[W^2(\mu)]] \leq \mathbb{E}_{\mu \sim \mu}[e^{CL_1}] < \infty. \]

Finally, the finite expected computational cost follows directly from Lemma 3.3.5 and we have proven Theorem 2.2.1.

Remark. We can still provide an unbiased estimator for \( \nu \) if \( \sigma \) is not bounded. In this case, we can localize \( \sigma \) by constructing \( \sigma^N \):

\[ \begin{align*}
\sigma^N(x) &= \sigma(x) \quad \text{when} \quad \|x\| \leq N \\
\sigma^N(x) &= 0 \quad \text{when} \quad \|x\| > N + 1.
\end{align*} \]
We denote our estimator $W$ by $W^N$ when it is generated under $\sigma^N(\cdot)$. Then, we pick some positive integer $n_2$, and add another layer in the estimator by defining

$$\tilde{W} \triangleq W^{n_2} + \frac{W^{N'+1+n_2} - W^{N'+n_2}}{p_{N'}},$$

where $N'$ is a geometric random variable. This estimator can be proven unbiased following similar techniques in the previous discussion. However, we will not go into the details here.

4 Simulation

In this section, we present two numerical examples using our estimator $W$. In Example 1, we checked the unbiasedness of our estimator. In Example 2, we compare our proposed unbiased estimator with a standard biased Monte Carlo estimator.

Example 1  We first introduce an example to check the unbiasedness of our estimator. For simplicity, we use the following one-dimensional SDE known as the Ornstein-Uhlenbeck Process:

$$\begin{cases}
  dX_t = -\alpha X_t dt + dB_t & \text{for } t \geq 0, \\
  X_0 = 0
\end{cases}, \quad (4.1)$$

where $\alpha \in \mathbb{R}$ is random coefficient in the SDE. In this one-dimensional model, we can solve the SDE (4.1) exactly. In particular, for any realization of $\alpha(\omega)$ the solution has the form of an Itô integral:

$$X_1 = e^{-\alpha t} \int_0^1 e^{\alpha s} dB_s.$$

Consequently, for any realization of $\alpha(\omega)$, using Itô’s isometry (see, for example, [38]), it can be shown that $X_1$ is normally distributed with mean 0 and variance $(2\alpha(\omega))^{-1}(1 - e^{-2\alpha(\omega)})$. For illustration, we let $\alpha(\omega)$ to be normally distributed with mean 1 and variance 0.05$^2$ along with $f(x) = x^2$, $G(x) = e^{-x^2}$ so that

$$\mathbb{E}[f(X_1)|\alpha] = \frac{1 - e^{-2\alpha}}{2\alpha}.$$

It follows from routine calculation that

$$\mathbb{E}[G(\mathbb{E}[f(X_1)|\alpha])] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \cdot 0.05^2} \cdot e^{-\frac{(x+1)^2}{2\cdot 0.05^2}} \cdot e^{-\frac{(1-e^{-2x})^2}{2x}} dx \approx 0.8291.$$

Furthermore, if we also want to examine the unbiasedness properties of $Z$, we can consider the particular case $\alpha = 1$ and obtain

$$\mathbb{E}[f(X_1)|\alpha = 1] \approx 0.4323.$$
Now, we can evaluate the unbiasedness property of the estimators $Z$ and $W$. Specifically, picking $n_0 = n_1 = 5$ as the baseline level of approximation for both estimators $Z$ and $W$, we generate 10,000 copies of $Z$ given that $\alpha = 1$ and obtain a sample mean of 0.4303 to compare with its true mean 0.4323, as in Figure 1a. Then, we also generate 10,000 copies of $W$ and obtain a sample mean of 0.8323 to compare with its true mean 0.8291, as in Figure 1b. In both cases, the sample size is 10,000 and the difference between sample mean and true mean is well within $\frac{1}{\sqrt{10000}} = 0.01$. This finding is consistent with our theoretical results that $Z$ and $W$ are unbiased.

Notice that our algorithm would simplify quite a bit here because the random field $\mu$ is replaced by a single random variable $\alpha$ and we no longer need to sample $V_1, V_2, \ldots$ to approximate it.

**Example 2** In this example, we consider the more complicated SDE

\[
\begin{aligned}
    dX_t &= -\mu(X_t)dt + \cos(X_t)dB_t & \text{for } t \geq 0 \\
    X_0 &= 0,
\end{aligned}
\]

where $\mu(x) = \sum_{n=1}^{\infty} n^{-4} \sin(nx)V_n$ as in Assumption (A1). In this example, we compare the proposed method with the standard Monte Carlo method without debiasing.

We take $\gamma = \frac{1}{3}$ and $\theta = \frac{1}{3}$ for simplicity (we leave the detail discussion in Section 5). Similar to the previous example, we take $n_0 = n_1 = 5$. We generate 10,000 copies of our estimator and compare it with 10,000 copies of a standard Monte Carlo estimator where we remove the debiasing part $\frac{\Delta N}{p_N}$ in both estimator $Z$ and $W$. To be specific, we replace $Z$ with only $f(X_{n_0}(1))$ in (3.12) and replace $W$ with only $G(\frac{\sum_{j=1}^{n_1} Z_j}{2n_1})$ in (3.18). As a result, using
CLT, we compute the 95% confidence interval to obtain $[0.4610, 0.4656]$ for our estimator while we obtain $[0.5189, 0.5255]$ for the standard Monte Carlo estimator. As we can see, these two intervals are not overlapping, suggesting that the standard Monte Carlo estimator has shown a significant bias.

5 Proofs

In this section, we present the proofs for Lemma 3.2.4, Lemma 3.2.8 and Lemma 2.1.2. The proof for all the supporting lemmas are provided in the Appendix.

5.1 Definitions and supporting lemmas

To prove Lemma 3.2.4, we introduce several definitions and supporting lemmas.

Definition 5.1.1. Let $\epsilon$ to be a positive constant small enough to satisfy

$$
\epsilon < \frac{1}{144} \quad \text{and} \quad \epsilon < \frac{1}{36} \left(\frac{1}{6} - 12\epsilon\right)(q - 4),
$$

where $q > 4$ is from Assumption A1 so that we can define positive quantities

$$
\alpha \triangleq \frac{1}{2} - \epsilon, \quad \beta \triangleq \frac{1}{2} + 2\epsilon, \quad \gamma \triangleq \frac{1}{3} - 12\epsilon, \quad \theta \triangleq \frac{4}{3} - \frac{23}{2} \epsilon \quad \text{and} \quad \delta \triangleq 33\epsilon.
$$

It is easy to check that the following important inequalities are satisfied:

$$
\gamma \geq \frac{1}{4}, \quad (3 + \frac{q - 4}{2})\gamma > 1, \quad 8(2\alpha - \beta) > 4 - \delta > 0,
$$

$$
4 - \delta > 3\theta > 0 \quad \text{(as in Lemma 3.2.9)} \quad \text{and} \quad \theta > 1 + \gamma > 0 \quad \text{(as in Lemma 3.2.10)}.
$$

Definition 5.1.2. For a standard one-dimensional Brownian motion $B(t)$ on $[0, 1]$, let $\alpha$ and $\beta$ be defined as in Definition 5.1.1. Then, define

$$
\|B\|_\alpha \triangleq \sup_{0 \leq s < t \leq 1} \frac{\|B(t) - B(s)\|_\infty}{|t - s|^\alpha} \quad \text{and} \quad \|A\|_{2\alpha} \triangleq \sup_{0 \leq s < t \leq 1} \max_{1 \leq i, j \leq d'} \frac{|A_{i,j}(s, t)|}{|t - s|^{2\alpha}}.
$$

$$
\|\bar{A}\|_{2\alpha} \triangleq \sup_{0 \leq s \leq t \leq 1} \max_{1 \leq i, j \leq d'} \frac{|\bar{A}_{i,j}(s, t)|}{|t - s|^{2\alpha}} \quad \text{and} \quad \Gamma_{\bar{R}} \triangleq \sup_{0 \leq s \leq t \leq 1} \max_{1 \leq i, j \leq d'} \frac{|\bar{R}_{i,j}(s, t)|}{|t - s|^{\beta(\Delta t_n)^{2\alpha - \beta}}}.
$$

where $D_n$ is the dyadic rationals that are multiples of $\frac{1}{2^n}$ in $[0, 1]$, and for $1 \leq i, j \leq d', i \neq j$,

$$
A_{i,j}(s, t) \triangleq \int_s^t (B_i(u) - B_i(s))dB_j(u) \quad \text{and} \quad \bar{A}_{i,i}(s, t) \triangleq A_{i,i}(s, t) = \frac{(B_i(t) - B_i(s))^2 - (t - s)}{2},
$$

$$
\bar{R}_{i,j}(s, t) \triangleq \max_{1 \leq i, j \leq d'} \frac{|\bar{R}_{i,j}(s, t)|}{|t - s|^{\beta(\Delta t_n)^{2\alpha - \beta}}},
$$

$$
\|\bar{R}\|_\beta \triangleq \sup_{0 \leq s \leq t \leq 1} \max_{1 \leq i, j \leq d'} \frac{|\bar{R}_{i,j}(s, t)|}{|t - s|^{\beta(\Delta t_n)^{2\alpha - \beta}}}.
$$
\[ \tilde{A}_{i,j}(s,t) \triangleq \frac{(B_i(t) - B_i(s))(B_j(t) - B_j(s))}{2} \quad \text{and} \quad \tilde{R}_{i,j}(t^n_i, t^n_m) \triangleq \sum_{k=l+1}^{m} \{A_{i,j}(t^n_{k-1}, t^n_k) - \tilde{A}_{i,j}(t^n_{k-1}, t^n_k)\}, \]

some of which we have already defined in Definition 3.2.1.

**Definition 5.1.3 (Notation).** Throughout the proof section, we will use \(C\) to represent any constant greater than 1 (i.e, \(C > 1\)) and use \(P(\cdot)\) to represent any polynomial function from \(\mathbb{R}^n \to \mathbb{R}\) where \(n \geq 1\) such that \(P(x) > 1\) for any \(x > 0\) for \(2 \leq i \leq n\). We will simply write this as \(P(x) > 1\) for \(x > 1\) and it will not affect our analysis.

It is straightforward to verify that for any \(P_1(\cdot), P_2(\cdot)\) and \(n \geq 0\), we can find some \(P_3(x)\) such that

\[
(P_1(x))^n < P_3(x) \quad P_1(x) + P_2(x) < P_3(x) \quad P_1(x) \cdot P_2(x) < P_3(x) \quad P_2(P_1(x)) < P_3(x).
\]

**Lemma 5.1.4 (Supporting Lemma).** The quantities \(\|B\|_{\alpha}, \|A\|_{2\alpha}, \|\tilde{A}\|_{2\alpha}\) and \(\Gamma_R\) defined in Definition 5.1.2 have moments of arbitrary order.

**Lemma 5.1.5 (Supporting Lemma).** Let \(X_n(\cdot)\) be the discretization scheme in Definition 3.6 generated under \(\mu^n(\cdot) \in \mathcal{L}_1\) with bounding number \(L_1 > 1\) and Brownian motion \(B(t), 0 \leq t \leq 1\). Then, we can find some fixed polynomial function \(P(x) > 1\) for \(x > 1\) such that

\[
\|X_n(t) - X_n(r)\|_\infty \leq P(L_1, \|B\|_{\alpha}, \|\tilde{A}\|_{2\alpha})|t - r|^{\alpha}
\]

for \(0 \leq r \leq t \leq 1\) and for all \(n \geq 0\).

**Lemma 5.1.6 (Supporting Lemma).** Let \(X_n^\mu(\cdot)\) be the discretization scheme modified from Definition 3.6 generated under \(\mu(\cdot) \in \mathcal{L}_1\) with bounding number \(L_1 > 1\) instead of \(\mu^n(\cdot) \in \mathcal{L}_1\) and Brownian motion \(B(t), 0 \leq t \leq 1\). Also, let \(X_t, 0 \leq t \leq 1\) be the solution of SDE in (3.1). Then, we can find some fixed polynomial \(P(x) > 1\) for \(x > 1\) such that

\[
\|X_n^\mu(t) - X_t\|_\infty \leq P(L_1, \|B\|_{\alpha}, \|A\|_{2\alpha}, \Gamma_R)\Delta t_n^{2\alpha - \beta}
\]

for all \(n \geq 0\) and \(0 \leq t \leq 1\).

**Lemma 5.1.7 (Supporting Lemma).** Let \(X_{n+1}(1)\) and \(X_{n+1}^a(1)\) be defined as in Definition 3.2.1 and 3.2.2 generated under fixed \(\mu^{(n+1)}(\cdot) \in \mathcal{L}_1\) with bounding number \(L_1 > 1\), we can find some fixed polynomial \(P(x) > 1\) for \(x > 1\) such that

\[
\mathbb{E}[\|X_{n+1}(1) - X_{n+1}^a(1)\|_8] \leq P(L_1)\Delta t_n^{8(2\alpha - \beta)}.
\]
5.2 Proof of Lemma 3.2.4

We now begin our proof of Lemma 3.2.4.

Proof of Lemma 3.2.4 Let $X_t, 0 \leq t \leq 1$ denote the solution of the SDE under $\mu(\cdot) \in \mathcal{L}_1$ with bounding number $L_1 > 1$ and $X_n(t)$ be the discretization scheme in Definition 3.6 generated under $\mu^n(\cdot) \in \mathcal{L}_1$ with bounding number $L_1 > 1$. Additionally, let $X_n^\mu(\cdot)$ be the discretization scheme modified from Definition 3.6 generated under $\mu^n(\cdot) \in \mathcal{L}_1$ with bounding number $L_1 > 1$ instead of $\mu^n(\cdot)$. Then, for $0 \leq t \leq 1$, we have the following bound $\|X_n(t) - X_t\|_\infty$.

\[
\|X_n(t) - X_t\|_\infty \leq \|X_n(t) - X_n^\mu(t)\|_\infty + \|X_n^\mu(t) - X_t\|_\infty. \tag{5.1}
\]

In order to prove Lemma 3.2.4, we provide bounds for both $\|X_n(t) - X_n^\mu(t)\|_\infty$ and $\|X_n^\mu(t) - X_t\|_\infty$.

For $\|X_n^\mu(t) - X_t\|_\infty$, using Lemma 5.1.6, we can find polynomial $\mathcal{P}(x) > 1$ for $x > 1$ that

\[
\|X_n^\mu(t) - X_t\|_\infty \leq \mathcal{P}(L_1, \|B\|_\alpha, \|A\|_{2\alpha}, \Gamma_R)\Delta t_n^{2(\alpha - \beta)}.\]

Similarly, using Lemma 5.1.3, we can further find some polynomial $\mathcal{P}(x) > 1$ for $x > 1$ such that

\[
\|X_n^\mu(t) - X_t\|_\infty \leq \mathcal{P}(L_1, \|B\|_\alpha, \|A\|_{2\alpha}, \Gamma_R)\Delta t_n^{4(\alpha - \beta)}.\]

It follows from Lemma 5.1.4 that the quantities associated with Brownian motions $\|B\|_\alpha, \|A\|_{2\alpha}$ and $\Gamma_R$ have moments of arbitrary order. Thus, fixing $\mu(\cdot) \in \mathcal{L}_1$, we can find some polynomial $\mathcal{P}'(x) > 1$ for $x > 1$ such that

\[
\mathbb{E}\|X_n^\mu(t) - X_t\|_\infty^4 \leq \mathbb{E}[\mathcal{P}(L_1, \|B\|_\alpha, \|A\|_{2\alpha}, \Gamma_R)]\Delta t_n^{4(\alpha - \beta)} \leq \mathcal{P}'(L_1)\Delta t_n^{4(\alpha - \beta)} \leq e^{CL_1}\Delta t_n^{4(\alpha - \beta)},
\]

for some constant $C > 1$ since $L_1 > 1$. Combining this with (5.1), we have

\[
\mathbb{E}\|X_n(t) - X_t\|_\infty^4 \leq 8\mathbb{E}\|X_n(t) - X_n^\mu(t)\|_\infty^4 + 8e^{CL_1}\Delta t_n^{4(\alpha - \beta)}. \tag{5.2}
\]

Thus, we can complete the proof if we can show

\[
\mathbb{E}\|X_n(t) - X_n^\mu(t)\|_\infty^4 \leq e^{CL_1}\Delta t_n^{4\alpha}, \tag{5.3}
\]

for some $C > 1$. This is because we have, according to Lemma 5.1.1

\[4\alpha = 2 - 4\epsilon > 2 - 16\epsilon = 4(2\alpha - \beta)\]
and thus (5.3) would imply
\[ \mathbb{E}\|X_n(t) - X_n^{\mu}(t)\|_\infty^4 \leq e^{C_1 \Delta t_n^{4(2\alpha - \beta)}}, \] (5.4)

since \( \Delta t_n < 1 \). Finally, we can simply conclude the proof using (5.2) and (5.4) by adjusting the constant \( C \). To prove (5.3), we define
\[ \tilde{\mu}^{(n)}(\cdot) \equiv \mu - \mu^{(n)} = \sum_{i=\lceil 2^n \rceil + 1}^{\infty} \frac{\lambda_i}{\ell_i^2} V_i(\omega) \psi_i(\cdot) \]
to be the remaining sum when we approximate \( \mu \) by \( \mu^{(n)} \). In Section 5.4 of the proof of Lemma 2.1.2, we will show that
\[ \| \tilde{\mu}^{(n)}(\cdot) \|_\infty \leq L_1 \Delta t_n^{3 + \frac{2\alpha - \beta}{2}}. \] (5.5)

Then, we may conduct the analysis on \( \| X_n(t) - X_n^{\mu^{(n)}}(t) \|_\infty \) based on the following recursion: For \( 1 \leq i \leq d, 0 \leq k \leq 2^n - 1, \)
\[
X_{i,n}(t_{k+1}) - X_{i,n}(t_k) \]
\[
= X_{i,n}(t_k) - X_{i,n}(t_k) + \left( \mu^{(n)}_i(X_n^{\mu}(t_k)) - \mu_i^{(n)}(X_n(t_k)) \right) \Delta t_n + \tilde{\mu}^{(n)}(X_n^{\mu}(t_k)) \Delta t_n
+ \sum_{j=1}^{d} \sigma_{ij}(X_n^{\mu}(t_k)) - \sigma_{ij}(X_n(t_k)) \Delta B_{j,k}^{n}
+ \sum_{j=1}^{d'} \sum_{l=1}^{d'} \sum_{m=1}^{d'} \left( \frac{\partial \sigma_{ij}}{\partial x_l}(X_n^{\mu}(t_k)) \sigma_{lm}(X_n^{\mu}(t_k)) - \frac{\partial \sigma_{ij}}{\partial x_l}(X_n(t_k)) \sigma_{lm}(X_n(t_k)) \right) \tilde{A}_{mj}(t_k, t_{k+1}), \] (5.6)

which is obtained by modifying (3.6) and simply taking the difference. Now, let
\[ \xi_{i,k} \equiv X_{i,n}(t_k) - X_{i,n}(t_k) \quad \text{and} \quad \xi_{i,n,k} \equiv X_{i,n}(t_k) - X_{i,n}(t_k) \] (5.7)
for \( 1 \leq i \leq d \) and \( 0 \leq k \leq 2^n \) and let
\[
\eta_{i,n,k} \equiv \left( \mu_i^{(n)}(X_n^{\mu}(t_k)) - \mu_i^{(n)}(X_n(t_k)) \right) \Delta t_n + \tilde{\mu}^{(n)}(X_n^{\mu}(t_k)) \Delta t_n
+ \sum_{j=1}^{d} \left( \sigma_{ij}(X_n^{\mu}(t_k)) - \sigma_{ij}(X_n(t_k)) \right) \Delta B_{j,k}^{n}
+ \sum_{j=1}^{d'} \sum_{l=1}^{d'} \sum_{m=1}^{d'} \left( \frac{\partial \sigma_{ij}}{\partial x_l}(X_n^{\mu}(t_k)) \sigma_{lm}(X_n^{\mu}(t_k)) - \frac{\partial \sigma_{ij}}{\partial x_l}(X_n(t_k)) \sigma_{lm}(X_n(t_k)) \right) \tilde{A}_{mj}(t_k, t_{k+1}), \]
(5.8)

so that (5.6) simplifies to, for \( 1 \leq i \leq d, 0 \leq k \leq 2^n - 1, \)
\[ \xi_{i,n,k+1} = \xi_{i,n,k} + \eta_{i,n,k}. \] (5.9)
Fixing \( \mu \in \mathcal{L}_1 \) and \( \mu^{(n)} \mathcal{L}_1 \) with bounding number \( L_1 > 1 \) and taking expectation on (5.9) after raising it to the fourth power, we have

\[
E(\xi_{i,n,k+1}^4) = E(\xi_{i,n,k}^4) + \mu(\omega)\eta_{i,n,k}^4 + 3E(\xi_{i,n,k}^3\eta_{i,n,k}) + 3E(\xi_{i,n,k}\eta_{i,n,k}^3) + 6E(\xi_{i,n,k}^2\eta_{i,n,k}^2).
\]

It now follows from the definition of \( \xi_{i,n,k} \) in (5.7) that it is sufficient to show

\[
E\|X_n(t) - X_n^\mu(t)\|_\infty^4 = E\|\xi_{n,2n}\|_\infty^4 \leq e^{C_L t}\Delta n^{4\alpha},
\]

for some constant \( C > 1 \). Thus, in what follows, we focus on the proof of (5.11), which consists of proofs for the following two statements: Fixing \( \mu \in \mathcal{L}_1 \) and \( \mu^{(n)} \mathcal{L}_1 \) with bounding number \( L_1 > 1 \),

- (I) We prove that there exists a constant \( C > 1 \) and a polynomial \( P(x) > 1 \) for \( x > 1 \) such that for \( 1 \leq i \leq d \) and \( 0 \leq k \leq 2^n \), we have
  \[
  E|\xi_{i,n,k}|^4 \leq e^{C_L t}\Delta n^{4\alpha}
  \]
  if \( n \) is large enough so that \( 2^n > P(L_1) \).

- (II) We prove that there is a polynomial \( P'(x) > 1 \) for \( x > 1 \) such that for \( 1 \leq i \leq d \) and \( 0 \leq k \leq 2^n \), we have
  \[
  E|\xi_{i,n,k}|^4 \leq P'(L_1)\Delta n^{4\alpha},
  \]
  if \( n \) is not large enough and \( 2^n \leq P(L_1) \).

**Proof of statement (I)** Fixing \( \mu \in \mathcal{L}_1 \) and \( \mu^{(n)} \mathcal{L}_1 \) with bounding number \( L_1 > 1 \), we use induction on \( 0 \leq k \leq 2^n \). First of all, when \( k = 0 \), for \( 1 \leq i \leq d \), the claim holds since \( \xi_{i,n,0} = X_{i,n}(0) - X_{i,n}(0) = x - x = 0 \).

Next, for \( 0 < k \leq 2^n - 1 \), assume that the induction hypothesis holds so that whenever \( 0 \leq j \leq k \),

\[
E|\xi_{i,n,j}|^4 \leq e^{C_L t}\Delta n^{4\alpha}
\]

for \( 1 \leq i \leq d \) and some \( C > 1 \). Our goal is to show

\[
E|\xi_{i,n,k+1}^4| \leq e^{C_L t_{k+1}}\Delta n^{4\alpha}
\]

for all \( 1 \leq i \leq d \). To do this, we provide bounds for every term on the right hand side of (5.10). For \( \eta_{i,n,k} \), according to Definition 5.1.2, we have

\[
|\eta_{i,n,k}| \leq |\partial \mu^{(n)}|_\infty |\xi_{n,k}|_\infty \Delta t_n + |\bar{\mu}^{(n)}|_\infty \Delta t_n + \bar{d}L|\xi_{n,k}|_\infty ||B||_\alpha \Delta t_n^\alpha + \bar{d}^2 L|\xi_{n,k}|_\infty ||A||_\alpha \Delta t_n^{2\alpha}
\]

\[
\leq L_1|\xi_{n,k}|_\infty \Delta t_n + L_1 \Delta t_n^{4+2\alpha} + \bar{d}L|\xi_{n,k}|_\infty ||B||_\alpha \Delta t_n^\alpha + \bar{d}^2 L|\xi_{n,k}|_\infty ||A||_\alpha \Delta t_n^{2\alpha}
\]

(5.12)
where the last line follows from (5.3). Since \( \xi_{n,k} \) and the shifted Brownian motion on \( B(t) - B(t^n_k), t^n_k \leq t \leq t^n_{k+1} \) are independent of each other (i.e., independent increments of Brownian motion), we can consider quantities \( \|B\|_\alpha \) and \( \|A\|_{2\alpha} \) to be associated with the new Brownian motion \( B(t) - B(t^n_k), t^n_k \leq t \leq t^n_{k+1} \) and thus independent of \( \xi_{n,k} \). Consequently, it then follows from Lemma 5.1.4 that we can find a constant \( C' > 1 \) such that

\[
E_{\eta_{i,n,k}}[\xi_{i,n,k}] = C' \left( t^n_{k+1} - t^n_k \right)^4 \Delta t^n_k + E_{\eta_{i,n,k}}^{1/2} \Delta t^n_k + E_{\eta_{i,n,k}}^{1/2} \Delta t^n_k + E_{\eta_{i,n,k}} \Delta t^n_k + E_{\eta_{i,n,k}} \Delta t^n_k
\]

(5.13)

for some \( C'' > 1 \) where the last line follows from both the induction hypothesis and the fact that \( 8\alpha < 16 + 2(q - 4) \) in Definition 5.1.1.

For the bound on \( E_{\eta_{i,n,k}}[\xi_{i,n,k}] \) in (5.10), we observe the terms in (5.8) and use (5.5) along with the martingale property (i.e., the independence of \( \Delta B^n_k \) and \( X_n(t^n_k) \)) to obtain

\[
E(\xi_{i,n,k}|\eta_{i,n,k}) = E\left[ \left( \mu_i^{(n)}(t^n_k) - \mu_i^{(n)}(t^n_k) \right)^3 \cdot \left( \left( \mu_i^{(n)}(X_n(t^n_k)) - \mu_i^{(n)}(X_n(t^n_k)) \right) + \mu_i^{(n)}(X_n(t^n_k)) \right) \Delta t^n_k \right]
\]

\[
\leq E\left[ \left( \mu_i^{(n)}(X_n(t^n_k)) - \mu_i^{(n)}(X_n(t^n_k)) \right) \Delta t^n_k \right]
\]

\[
\leq 2L_1 \Delta t^n_k \Delta t^n_k \cdot \Delta t^n_k
\]

(5.14)

The last inequality follows from induction hypothesis, the second inequality follows from Hölder’s inequality and the fact that \( \alpha < 4 + \frac{q}{2} \) as in Definition 5.1.1 and the first inequality follows from the bound on \( \|\partial \mu_i^{(n)}\|_\infty \) in Assumption A1.

For the bound on \( E(\xi_{i,n,k}^2|\eta_{i,n,k}) \), using the bound on \( |\eta_{i,n,k}| \) in (5.12) and the fact that \( E[(B(t) - B(s))^2] = O(|t - s|) \) and \( E[(A_{ij}(s,r))^2] = O((t - s)^2) \) (see, for example, [24]), we can find some \( C' > 1 \) that

\[
E(\xi_{i,n,k}^2|\eta_{i,n,k}) \leq C' \left( E(\|\xi_{i,n,k}\|^4) L_1^2 \Delta t^n_k + E(\|\xi_{i,n,k}\|^2) L_1^2 \Delta t^n_k + \Delta t^n_k \right)
\]

(5.15)

\[
+ C' \left( E(\|\xi_{i,n,k}\|^4) \Delta t^n_k + E(\|\xi_{i,n,k}\|^2) \Delta t^n_k \right)
\]

\[
\leq C'' \left( E(\|\xi_{i,n,k}\|^4) \Delta t^n_k + E(\|\xi_{i,n,k}\|^2) \Delta t^n_k \right)
\]

\[
\leq 2C'' \left( E(\|\xi_{i,n,k}\|^4) \Delta t^n_k \right) \Delta t^n_k + 1 \Delta t^n_k + 1 \Delta t^n_k + 1
\]

for some \( C'' > 1 \). The last line follows from induction hypothesis. The second to last line follows from Hölder’s inequality and the fact that \( 2\alpha + 2 < 4 + q \) as in Definition 5.1.1.
Finally, to bound $E(\xi_{i,n,k}\eta_{i,n,k}^3)$ in (5.10), following similar techniques, we use inequality (5.13), induction hypothesis and Hölder’s inequality to obtain

$$E(\xi_{i,n,k}\eta_{i,n,k}^3) \leq (E(\xi_{i,n,k}^4))^{\frac{1}{4}}(E(\eta_{i,n,k}^4))^{\frac{3}{4}} \leq CL \cdot \Delta t_n^{4\alpha}$$

(5.16)

Now we are ready to prove the induction hypothesis. Let

$$C = 12C''d^4 + 6d^4 + 1 \quad \text{and} \quad P(x) = (C''(x^4d^4 + 3x^3d^3 + 12x^2d^4))^3.$$

(5.17)

It is easy to check that $C > 1$ and the polynomial $P(x) > 1$ for $x > 1$. Then, it follows from Definition 5.1.1 and standard calculation that if $n$ is large enough that $2^n > P(L_1)$ (i.e, $\Delta t_n < (\mathcal{P}(L_1))^{-1}$), then

$$C''L_1^4d^4 + 3C''L_1^3d^3 \Delta t_n^{4\alpha-1} + 12C''L_1^2d^4 \Delta t_n 
\leq (C''L_1^4d^4 + 3C''L_1^3d^3 + 12C''L_1^2d^4)\Delta t_n^{3\alpha-1}
= (\mathcal{P}(L_1))^{\frac{3}{4}} \Delta t_n^{3\alpha-1}
< (\mathcal{P}(L_1))^{\frac{3}{4}} - 3\alpha < 1$$

(5.18)

where the last inequality follows from the fact that $0 > \frac{4}{3} - 3\alpha$, $L_1 > 1$ and $P(x) > 1$ for $x > 1$. Thus, for $n$ such that $2^n > P(L_1)$, we use (5.10), Hölder’s inequality and the bound acquired in (5.13), (5.14), (5.15) and (5.16) to get

$$E(\xi_{i,n,k+1}^4)
= E(\xi_{i,n,k}^4) + E_{\mu(\omega)}(\eta_{i,n,k}) + 3E(\xi_{i,n,k}\eta_{i,n,k}^3) + 3E(\xi_{i,n,k}\eta_{i,n,k}^2) + 6E(\xi_{i,n,k}\eta_{i,n,k}^2) 
\leq e^{CL_1 \cdot t_n^{\alpha}} \Delta t_n^{4\alpha} \left( 1 + C''L_1^4d^4 \Delta t_n^{4\alpha} + 6L_1d^4 \Delta t_n + 3C''L_1^3d^3 \Delta t_n^{3\alpha} + 12C''d^4(L_1^2 \Delta t_n + 1)\Delta t_n \right)
\leq e^{CL_1 \cdot t_n^{\alpha}} \Delta t_n^{4\alpha} \left( 1 + (6d^3 + 12C''d^4 + 1)L_1\Delta t_n \right)
= e^{CL_1 \cdot t_n^{\alpha}} \Delta t_n^{4\alpha} \left( 1 + CL_1 \Delta t_n \right)
\leq e^{CL_1 \cdot t_n^{\alpha}} \Delta t_n^{4\alpha},$$

where the last line follows from convexity of exponential function: $e^y \geq e^x + e^x \cdot (y - x)$ for $y \geq x$. The second to last inequality follows from (5.17), (5.18) and the fact that $L_1 > 1$. This concludes the induction. However, since $t_n^k \leq 1$ for all $0 \leq k \leq 2^n$, we have actually proven that when $2^n > P(L_1)$ (i.e, $\Delta t_n < (\mathcal{P}(L_1))^{-1}$),

$$E\|X_{i,n}^\mu(t) - X_{i,n}(t)\|_4^4 \leq e^{CL_1} \cdot \Delta t_n^{4\alpha}$$

for all $1 \leq i \leq d$ and $0 \leq t \leq 1$. 

29
Proof of statement (II) Next, we extend the result to the case where $2^n \leq \mathcal{P}(L_1)$. By observing (5.6), we can find polynomial function $\mathcal{P}'(x) > 1$ for $x > 1$ so that:

$$|(X_{i,n}(t^n_{k+1}) - X_{i,n}(t^n_{k})) - (X_{i,n}(t^n_{k}) - X_{i,n}(t^n_{k}))| \leq \mathcal{P}'(L_1, \|B\|_{\alpha}, \|\tilde{A}\|_{2\alpha}) \Delta t^n_k.$$ 

Since the number of iterations in the discretization scheme $2^n$ is at most $\mathcal{P}(L_1)$, we have

$$\|X_{i,n}(\cdot) - X_{i,n}(\cdot)\|_{\infty} \leq \mathcal{P}(L_1) \mathcal{P}'(L_1, \|B\|_{\alpha}, \|\tilde{A}\|_{2\alpha}) \Delta t^n_k,$$

and consequently, from Lemma [5.1.4] that

$$\mathbb{E}\|X_{i,n}(\cdot) - X_{i,n}(\cdot)\|_{\infty}^4 \leq \mathcal{P}''(L_1) \Delta t^n_k$$

for some polynomial $\mathcal{P}''(x) > 1$ when $x > 1$.

This concludes the proof of Lemma 3.2.4. \hfill \Box

5.3 Proof of Lemma 3.2.8

We first prove the second claim (3.14) of Lemma 3.2.8.

Proof of (3.14) in Lemma 3.2.8 Assume without loss of generality that $x = 0$. Fixing $\mu \in \mathcal{L}_1$ and $\mu^{(n)} \in \mathcal{L}_1$ with bounding number $L_1 > 1$, since

$$\sup_{1 \leq k \leq 2^n} |X_{i,n}(t^n_k)|^4 \leq \left(\sum_{k=0}^{2^n} \|\mu\|_{\infty} \Delta t^n_k + \sum_j \sup_{1 \leq h \leq 2^n} \left| \sum_{k=0}^h \sigma_{ij}(X_n(t^n_k)) \Delta B^n_{j,k} \right| \right.$$ 

$$+ \sum_{j,l,m} \sup_{1 \leq h \leq 2^n} \left( \sum_{k=0}^h \frac{\partial \sigma_{ij}}{\partial x_l}(X_n(t^n_k)) \sigma_{lm}(X_n(t^n_k)) \tilde{A}_{mj}(t^n_k, t^n_{k+1}) \right),$$

it follows from (3.6) and Assumption A1 that we can find constant $C > 1$ such that

$$\sup_{1 \leq k \leq 2^n} |X_{i,n}(t^n_k)|^4 \leq C \cdot \left( L_1^4 + \sum_j \left( \sup_{1 \leq h \leq 2^n} \left| \sum_{k=0}^h \sigma_{ij}(X_n(t^n_k)) \Delta B^n_{j,k} \right| \right)^4 \right.$$ 

$$+ \sum_{j,l,m} \left( \sup_{1 \leq h \leq 2^n} \left| \sum_{k=0}^h \frac{\partial \sigma_{ij}}{\partial x_l}(X_n(t^n_k)) \sigma_{lm}(X_n(t^n_k)) \tilde{A}_{mj}(t^n_k, t^n_{k+1}) \right| \right)^4 \right).$$

Now, using the fact that $\mathbb{E}(B(t) - B(s))^4 = O(t - s)^2$ and $\mathbb{E}(\tilde{A}_{ij}(s,t))^4 = O(t - s)^4$, we recall Burkholder-Davis-Gundy inequality [6] to further find constant $C' > 1$ and $C'' > 1$ so
that

$$
\mathbb{E} \sup_{1 \leq k \leq 2^n} |X_{i,n}(t^n_k)|^4 \leq C \left( L_1^4 + \sum_j \sup_{1 \leq h \leq 2^n} \sum_{k=0}^h |\sum_{i,j} \sigma_{i,j}(X_n(t^n_k)) \Delta B_{j,k}^n|^4 \right)
$$

$$
\leq C' \left( L_1^4 + \sum_j \mathbb{E} \left( \sum_{k=0}^{2^n} \left( \frac{\partial \sigma_{i,j}}{\partial x_l}(X_n(t^n_k)) \sigma_{lm}(X_n(t^n_k)) \tilde{A}_{mj}(t^n_k, t^n_{k+1}) \right)^4 \right) \right)^{1/2}
$$

$$
\leq C'' \left( L_1^4 + \sum_{j=1}^{d'} \mathbb{E} 2^n \sum_{k=0}^{2^n} \left( \frac{\partial \sigma_{i,j}}{\partial x_l}(X_n(t^n_k)) \sigma_{lm}(X_n(t^n_k)) (\tilde{A}_{mj}(t^n_k, t^n_{k+1}))^4 \right) \right)^{1/2}
$$

$$
\leq C''\left( L_1^4 + \sum_{j=1}^{d'} \mathbb{E} 2^n \left( \sum_{k=0}^{2^n} \left( \frac{\partial \sigma_{i,j}}{\partial x_l}(X_n(t^n_k)) \sigma_{lm}(X_n(t^n_k)) (\tilde{A}_{mj}(t^n_k, t^n_{k+1}))^4 \right) \right) \right)^{1/2}
$$

$$
< \mathcal{P}(L_1)
$$

for some polynomial function $\mathcal{P}(x) > 1$ for $x > 1$. Now, the claim on $\mathbb{E}|f(X_n(1))|^4$ follows by invoking the bound on $\|\frac{\partial f}{\partial x_1}\|_{\infty}$ in Assumption A2. \hfill \square

We proceed to the proof of the first claim of Lemma 3.2.8, namely (3.13). It follows from Equation (3.15) in [17] that we have for $p \geq 2$

$$
|\Delta_n|^p \leq 2^{p-1} L^p \mathbb{E}\left[ \left| \frac{1}{2}(X_{n+1}(1) + X_{n+1}(1)) - X_{n}(1) \right|_{\infty}^p \right] + 2^{-p-1} L^p \mathbb{E}\left[ \left| X_{n+1}(1) - X_{n+1}(1) \right|_{\infty}^p \right].
$$

(5.19)

Thus, according to (5.19), we can prove Lemma 3.2.8

$$
\mathbb{E}(\Delta_n^4) \leq e^{C L_1} \Delta_t^{4-\delta} = e^{C L_1(w)} \Delta_t^{4-3\delta},
$$

where $\delta > 0$ is defined in Definition 5.1.1, it is sufficient to provide bound on

$$
\mathbb{E}\left[ \left| \frac{1}{2}(X_{n+1}(1) + X_{n+1}(1)) - X_{n}(1) \right|_{\infty}^4 \right]
$$

(5.20)

and

$$
\mathbb{E}\left[ \left| X_{n+1}(1) - X_{n+1}(1) \right|_{\infty}^8 \right].
$$

(5.21)

Note that bound on (5.21) is provided by Lemma 5.1.7 since $4 - \delta < 8(2\alpha - \beta)$ as in Definition 5.1.1 and $\mathcal{P}(L_1) < e^{C L_1}$ for appropriately chosen $C > 1$. So we just need to prove (5.20).
Proof of (3.13) in Lemma 3.2.8. First we write the recursion for $X_{n+1}(\cdot)$ over the coarse step $\Delta t_n$ instead of $\Delta t_{n+1}$. For $1 \leq i \leq d$ and $1 \leq k \leq 2^n$, adding up two steps of recursion for $X_{n+1}(\cdot)$, we have:

$$X_{i,n+1}(t^n_{k+1}) = X_{i,n+1}(t^n_k) + \mu_i^{(n+1)}(X_{n+1}(t^n_k)) + \sum_{j=1}^{d'} \sigma_{ij}(X_{n+1}(t^n_k))\Delta B^n_{j,k}$$

$$+ \sum_{j=1}^{d'} \sum_{l=1}^{d'} \sum_{m=1}^{d'} \frac{\partial \sigma_{ij}}{\partial x_l}(X_{n+1}(t^n_k))\sigma_{lm}(X_{n+1}(t^n_k))\tilde{A}_{mj}(t^n_k, t^n_{k+1})$$

$$- \sum_{j=1}^{d'} \sum_{l=1}^{d'} \sum_{m=1}^{d'} \frac{\partial \sigma_{ij}}{\partial x_l}(X_{n+1}(t^n_k))\sigma_{lm}(X_{n+1}(t^n_k))((\Delta B_{j,2k}^{n+1} \Delta B_{m,2k+1}^{n+1} - \Delta B_{m,2k}^{n+1} \Delta B_{j,2k+1}^{n+1})$$

$$+ N_{i,n,k}^f + M_{i,n,k}^f(1) + M_{i,n,k}^f(2) + M_{i,n,k}^f(3)$$

where we define

$$M_{i,n,k}^f(2) \triangleq \left(\sum_{j=1}^{d'} (\sigma_{ij}(X_{n+1}(t^n_{2k+1})) - \sigma_{ij}(X_{n+1}(t^n_k))) - \sum_{j=1}^{d'} \sum_{l=1}^{d'} \sum_{m=1}^{d'} \left(\frac{\partial \sigma_{ij}}{\partial x_l} \cdot \sigma_{lm}\right)(X_{n+1}(t^n_k))\Delta B_{m,2k}^{n+1}\right) \Delta B_{j,2k+1}^{n+1}$$

$$M_{i,n,k}^f(3) \triangleq \sum_{j=1}^{d'} \sum_{l=1}^{d'} \sum_{m=1}^{d'} \left(\frac{\partial \sigma_{ij}}{\partial x_l}(X_{n+1}(t^n_k))\sigma_{lm}(X_{n+1}(t^n_k))\tilde{A}_{mj}(t^n_{2k+1}, t^n_{k+1})ight)$$

$$M_{i,n,k}^f(1) \triangleq \left(\sum_{j=1}^{d} \frac{\partial \mu_i^{(n+1)}}{\partial x_j}(X_{n+1}(t^n_k))(X_{j,n+1}(t^n_{2k+1}) - X_{j,n+1}(t^n_k))ight)$$

$$+ \frac{1}{2} \sum_{j=1}^{d} \sum_{m=1}^{d} \frac{\partial^2 \mu_i^{(n+1)}}{\partial x_j \partial x_m}(\eta)(X_{j,n+1}(t^n_{2k+1}) - X_{j,n+1}(t^n_k))(X_{m,n+1}(t^n_{2k+1}) - X_{m,n+1}(t^n_k))\Delta t_n^{2k} - M_{i,n,k}^f(1)$$

$$= \left(\sum_{j=1}^{d} \frac{\partial \mu_i^{(n+1)}}{\partial x_j}(X_{n+1}(t^n_k))(\mu_j^{(n+1)}(X_{n+1}(t^n_k))\Delta t_n^{2k} + \sum_{m,l,\tilde{m}} \left(\frac{\partial \sigma_{jm}}{\partial x_l} \cdot \sigma_{l\tilde{m}}\right)(X_{n+1}(t^n_k))\tilde{A}_{m\tilde{m}}(t^n_k, t^n_{2k+1})ight)$$

$$+ \frac{1}{2} \sum_{j,m=1}^{d} \frac{\partial^2 \mu_i^{(n+1)}}{\partial x_j \partial x_m}(\rho)(X_{j,n+1}(t^n_{2k+1}) - X_{j,n+1}(t^n_k))(X_{m,n+1}(t^n_{2k+1}) - X_{m,n+1}(t^n_k))\Delta t_n^{2k}$$

for some $\rho$ that lies between $X_{n+1}(t^n_k)$ and $X_{n+1}(t^n_{2k+1})$.

Furthermore, we similarly define $N_{i,n,k}^a, M_{i,n,k}^a(\cdot)$ associated with $X_{n+1}^a(\cdot)$, $B^{n+1,a}(t)$ and $\tilde{A}^a(t^n_{2k+1}, t^n_{k+1})$ so we can write the recursion over the coarse step $\Delta t_n$ for $X_{n+1}^a(\cdot)$ by using
\( X^n_{i,n+1}(t_{k+1}) = X^n_{i,n+1}(t_k) + \mu_i^{(n+1)}(X^n_{i,n+1}(t_k)) \Delta B^n_{j,k} \\
+ \sum_{j=1}^{d'} \sum_{l=1}^d \sum_{m=1}^{d'} \frac{\partial \sigma_{ij}}{\partial x_l}(X^n_{i+1}(t_k)) \sigma_{lm}(X^n_{i+1}(t_k)) \tilde{A}_{m_j}(t_k, t_{k+1}) \\
- \sum_{j=1}^{d'} \sum_{l=1}^d \sum_{m=1}^{d'} \frac{\partial \sigma_{ij}}{\partial x_l}(X^n_{i+1}(t_k)) \sigma_{lm}(X^n_{i+1}(t_k))(\Delta B^n_{j,2k} \Delta B^n_{m,2k+1} - \Delta B^n_{m,2k} \Delta B^n_{j,2k+1}) \\
+ N^n_{i,n,k} + M^n_{i,n,k}^{(1)} + M^n_{i,n,k}^{(2)} + M^n_{i,n,k}^{(3)} \\
\tag{5.23}
\]

Now, combining these results, we can write the recursion for \( \bar{X}_{n+1}(\cdot) \triangleq \frac{1}{2}(X_{n+1}(\cdot) + X^n_{n+1}(\cdot)) \) over the coarse step \( \Delta t_n \):

\( \bar{X}_{n+1}(t_{k+1}) = \bar{X}_{n+1}(t_k) + \mu_i^{(n+1)}(\bar{X}_{n+1}(t_k)) \Delta t_n + \sum_{j=1}^{d'} \sigma_{ij}(\bar{X}_{n+1}(t_k)) \Delta B^n_{j,k} \\
+ \sum_{j=1}^{d'} \sum_{l=1}^d \sum_{m=1}^{d'} \frac{\partial \sigma_{ij}}{\partial x_l}(\bar{X}_{n+1}(t_k)) \sigma_{lm}(\bar{X}_{n+1}(t_k)) \tilde{A}_{m_j}(t_k, t_{k+1}) + R_{i,n,k} \\
\]

where we define

\( R_{i,n,k} \triangleq N^{(1)}_{i,n,k} + M^{(1)}_{i,n,k} + M^{(2)}_{i,n,k} + M^{(3)}_{i,n,k} \\
+ \frac{1}{2}(N^f_{i,n,k} + M^f_{i,n,k} + M^f_{i,n,k}^{(2)} + M^f_{i,n,k}^{(3)} + N^a_{i,n,k} + M^a_{i,n,k}^{(1)} + M^a_{i,n,k}^{(2)} + M^a_{i,n,k}^{(3)}) \tag{5.23} \)

where we define

\( N^{(1)}_{i,n,k} \triangleq \frac{1}{2} \left( \mu_i^{(n+1)}(X_{n+1}(t_k)) + \mu_i^{(n+1)}(X^n_{n+1}(t_k)) \right) - \mu_i^{(n+1)}(\bar{X}_{n+1}(t_k)) \),

\( M^{(1)}_{i,n,k} \triangleq \sum_{j=1}^{d'} \left( \frac{1}{2}(\sigma_{ij}(X_{n+1}(t_k)) + \sigma_{ij}(X^n_{n+1}(t_k))) - \sigma_{ij}(\bar{X}_{n+1}(t_k)) \right) \Delta B^n_{j,k} \),

\( M^{(2)}_{i,n,k} \triangleq \sum_{j,m=1}^{d'} \sum_{l=1}^d \left( \frac{1}{2}(\sigma_{ij}(X_{n+1}(t_k)) + \sigma_{ij}(X^n_{n+1}(t_k))) - \sigma_{ij}(\bar{X}_{n+1}(t_k)) \right) \Delta B^n_{j,k} \),

\( M^{(3)}_{i,n,k} \triangleq \sum_{j,m=1}^{d'} \sum_{l=1}^d \frac{1}{2} \left( \sigma_{ij}(X^n_{n+1}(t_k)) - \sigma_{ij}(X_{n+1}(t_k)) \right) \mu_i(t_k) \),

\( \Delta B^n_{j,2k} \Delta B^n_{m,2k+1} - \Delta B^n_{m,2k} \Delta B^n_{j,2k+1} \). \tag{5.24} \)
Finally, subtract the recursion in (3.6) for \( X_n(\cdot) \) from \( \tilde{X}_n(\cdot) \) to obtain

\[
\tilde{X}_{i,n+1}(t^u_{k+1}) - X_{i,n}(t^u_{k+1}) = \tilde{X}_{i,n+1}(t^u_k) - X_{i,n}(t^u_k) + (\mu^{(n)}_i(\tilde{X}_{i,n+1}(t^u_k)) - \mu^{(n)}_i(X_{i,n}(t^u_k))) \Delta t_n + (\mu^{(n+1)}_i - \mu^{(n)}_i)(\tilde{X}_{i,n+1}(t^u_k)) \Delta t_n + \sum_{j=1}^{d'} \sigma_{ij}(\tilde{X}_{i,n+1}(t^u_k)) - \sigma_{ij}(X_{i,n}(t^u_k)) \Delta B^n_{j,k} + \sum_{j=1}^{d'} \sum_{l=1}^{d} \sum_{m=1}^{d'} \left( (\frac{\partial \sigma_{ij}}{\partial x_l}) \cdot \sigma_{lm}(\tilde{X}_{i,n+1}(t^u_k)) - (\frac{\partial \sigma_{ij}}{\partial x_l}) \cdot \sigma_{lm}(X_{i,n}(t^u_k)) \right) \tilde{A}_{n,j,k} + R_{i,n,k}
\]

We are now ready to prove (5.20) by bounding \( \mathbb{E}|\tilde{X}_{i,n+1}(t^u_k) - X_{i,n}(t^u_k)|^4 \). Similarly as in the proof of Lemma 3.2.4, we simplify the notation by defining

\[
\xi_{i,n,k} \triangleq \tilde{X}_{i,n+1}(t^u_k) - X_{i,n}(t^u_k) \quad \text{and} \quad \xi_{n,k} \triangleq \tilde{X}_{n+1}(t^u_k) - X_{n}(t^u_k)
\]

with

\[
\eta_{i,n,k} \triangleq (\mu^{(n)}_i(\tilde{X}_{i,n+1}(t^u_k)) - \mu^{(n)}_i(X_{i,n}(t^u_k))) \Delta t_n + \sum_{j=1}^{d'} (\sigma_{ij}(\tilde{X}_{i,n+1}(t^u_k)) - \sigma_{ij}(X_{i,n}(t^u_k))) \Delta B^n_{j,k} + \sum_{j=1}^{d'} \sum_{l=1}^{d} \sum_{m=1}^{d'} \left( (\frac{\partial \sigma_{ij}}{\partial x_l}) \cdot \sigma_{lm}(\tilde{X}_{i,n+1}(t^u_k)) - (\frac{\partial \sigma_{ij}}{\partial x_l}) \cdot \sigma_{lm}(X_{i,n}(t^u_k)) \right) \tilde{A}_{n,j,k} + R_{i,n,k} + (\mu^{(n+1)}_i - \mu^{(n)}_i)(\tilde{X}_{i,n+1}(t^u_k)) \Delta t_n
\]

so that we have

\[
\xi_{i,n,k+1} = \xi_{i,n,k} + \eta_{i,n,k}
\]

for \( 0 \leq k \leq 2^n - 1 \). Fixing \( V_1, V_2, \ldots \) such that \( \mu \in \mathcal{L}_1 \) and \( \{\mu^{(n)}\}_{n \geq 0} \subset \mathcal{L}_1 \), we want to find constant \( C > 1 \) and polynomial \( \mathcal{P}(x) > 1 \) for \( x > 1 \) such that if \( n \) is large enough that \( 2^n > \mathcal{P}(L_1) \), then

\[
\mathbb{E}(\xi_{i,n,k})^4 \leq C L_1 \Delta t_k \Delta t^{4-\delta}_n
\]

for all \( 1 \leq i \leq d \) and \( 0 \leq k \leq 2^n \). Similarly, we prove by induction on \( 0 \leq k \leq 2^n \). We first need to analyze all the terms of \( R_{i,n,k} \) in (5.23).

We start by bounding \( N^{(1)}_{i,n,k} \) in (5.24) using Taylor expansion

\[
N^{(1)}_{i,n,k} \triangleq \frac{1}{2} \left( \mu^{(n+1)}_i(X_{i,n+1}(t^u_k)) - \mu^{(n+1)}_i(X^n_{i,n+1}(t^u_k)) \right) - \mu^{(n+1)}_i(X^n_{i,n+1}(t^u_k))
\]

\[
= \frac{1}{16} \sum_{j=1}^{d} \sum_{m=1}^{d} \left( \frac{\partial^2 \mu_i}{\partial x_j \partial x_m}(\rho_1) + \frac{\partial^2 \mu_i}{\partial x_j \partial x_m}(\rho'_1) \right) (X^{a}_{j,n+1}(t^u_k) - X^n_{j,n+1}(t^u_k)) (X^{a}_{m,n+1}(t^u_k) - X^n_{m,n+1}(t^u_k)) \Delta t_n
\]
where \( \rho_1 \) and \( \rho'_1 \) lie somewhere between \( X_n^{a+1}(t_k^n) \) and \( X_{n+1}(t_k^n) \). Now we use Lemma 5.1.7 on \( (X_{j,n+1}(t_k^n) - X_{j,n+1}(t_k^n))(X_{m,n+1}(t_k^n) - X_{m,n+1}(t_k^n)) \) and H"older’s inequality to obtain
\[
\mathbb{E}(N^{(1)}_{i,n,k})^4 < P(L_1)\Delta t^{8(2\alpha-\beta)+4} \tag{5.27}
\]
for some fixed polynomial \( P(x) > 1 \) for \( x > 1 \).

Now, for \( N_{i,n,k}^f \) in (5.22), we also use Taylor expansion to obtain
\[
N_{i,n,k}^f = \left( \sum_{j=1}^{d} \frac{\partial \mu^{(n+1)}_i}{\partial x_j} (X_{n+1}(t_k^n)) \left( \mu^{(n+1)}_j (X_{n+1}(t_k^n)) \Delta t_n \right)
+ \sum_{m=1}^{d'} \sum_{l=1}^{d} \sum_{\tilde{m}=1}^{d'} \left( \frac{\partial \sigma_{im}}{\partial x_l} \cdot \sigma_{\tilde{m}m} \right) (X_{n+1}(t_k^n)) \tilde{A}_{m\tilde{m}}(t_k^n, t_{2k+1}^{n+1}) \right)
+ \frac{1}{2} \sum_{j=1}^{d} \sum_{m=1}^{d} \frac{\partial^2 \mu^{(n+1)}_i}{\partial x_j \partial x_m} (\rho)(X_{j,n+1}(t_{2k+1}^{n+1}) - X_{j,n+1}(t_k^n))(X_{m,n+1}(t_{2k+1}^{n+1}) - X_{m,n+1}(t_k^n)) \Delta t_n \frac{1}{2}
\]
by using Lemma 5.1.3 on \( (X_{j,n+1}(t_{2k+1}^{n+1}) - X_{j,n+1}(t_k^n))(X_{m,n+1}(t_{2k+1}^{n+1}) - X_{m,n+1}(t_k^n)) \) and Lemma 5.1.4

Thus, we can also find some fixed polynomial \( P(x) > 1 \) for \( x > 1 \) such that
\[
\mathbb{E}(N_{i,n,k}^f)^4 < P(L_1)\Delta t^{8\alpha+4}. \tag{5.28}
\]
For other terms of \( R_{i,n,k} \) in (5.24), we similarly write out their Taylor expansion as follows:
\[
M_{i,n,k}^{(1)} = \sum_{j=1}^{d'} \left( \frac{1}{2} \left( \sigma_{ij}(X_{n+1}(t_k^n)) + \sigma_{ij}(X_{n+1}(t_k^n)) - \sigma_{ij}(\tilde{X}_{n+1}(t_k^n)) \right) \Delta B_{j,k}^n \right)
= \frac{1}{16} \sum_{j=1}^{d'} \sum_{m,l=1}^{d} \left( \frac{\partial^2 \sigma_{ij}}{\partial x_m \partial x_l} (\rho_2) + \frac{\partial^2 \sigma_{ij}}{\partial x_m \partial x_l} (\rho'_2) \right) (X_{m,n+1}(t_k^n) - X_{m,n+1}(t_k^n)) \cdot (X_{l,n+1}(t_k^n) - X_{l,n+1}(t_k^n)) \Delta B_{j,k}^n.
\]
and also
\[
M_{i,n,k}^{(2)} \triangleq \sum_{j,m=1}^{d} \sum_{l=1}^{d} \left( \frac{1}{2} \left( \frac{\partial \sigma_{ij}}{\partial x_l} \cdot \sigma_{lm} \right)(X_{n+1}(t_k^n)) + \left( \frac{\partial \sigma_{ij}}{\partial x_l} \cdot \sigma_{lm} \right)(X_{n+1}(t_k^n)) \right) - \left( \frac{\partial \sigma_{ij}}{\partial x_l} \cdot \sigma_{lm} \right)(\bar{X}_{n+1}(t_k^n))
\]
\[
\times \tilde{A}_{mj}(t_k^n, t_{k+1}^n)
\]
\[
= \frac{1}{4} \sum_{j,m,l}^{d'} \sum_{l'=1}^{d} \left( \left( \frac{\partial^2 \sigma_{ij}}{\partial x_l \partial x_{l'}} \sigma_{lm} + \frac{\partial \sigma_{ij}}{\partial x_l} \frac{\partial \sigma_{lm}}{\partial x_{l'}} \right)(\rho_3) - \left( \frac{\partial^2 \sigma_{ij}}{\partial x_l \partial x_{l'}} \sigma_{lm} + \frac{\partial \sigma_{ij}}{\partial x_l} \frac{\partial \sigma_{lm}}{\partial x_{l'}} \right)(\rho_4) \right)
\]
\[
\times (X_{l', n+1}(t_k^n) - X_{l', n+1}^a(t_k^n)) \tilde{A}_{mj}(t_k^n, t_{k+1}^n)
\]
\[
M_{i,n,k}^{(3)} \triangleq \sum_{j,m,l}^{d'} \sum_{l'=1}^{d} \left( \left( \frac{\partial^2 \sigma_{ij}}{\partial x_l \partial x_{l'}} \sigma_{lm} + \frac{\partial \sigma_{ij}}{\partial x_l} \frac{\partial \sigma_{lm}}{\partial x_{l'}} \right)(\rho_5) + \left( \frac{\partial^2 \sigma_{ij}}{\partial x_l \partial x_{l'}} \sigma_{lm} + \frac{\partial \sigma_{ij}}{\partial x_l} \frac{\partial \sigma_{lm}}{\partial x_{l'}} \right)(\rho_6) \right)
\]
\[
\times (X_{l', n+1}(t_k^n) - X_{l', n+1}^a(t_k^n)) (\Delta B_{j,2k}^{n+1} \Delta B_{m,2k+1}^{n+1} - \Delta B_{m,2k}^{n+1} \Delta B_{j,2k+1}^{n+1})
\]
where all the \( \rho_i \) and \( \rho'_i \) lie somewhere between \( X_{n+1}(t_k^n) \) and \( X_{n+1}^a(t_k^n) \). For the sake of completeness, the other terms in (5.24) from \( M_{i,n,k}^{f,(1)} \) can be written as
\[
M_{i,n,k}^{f,(1)} \triangleq \left( \sum_{j=1}^{d} \frac{\partial u_j^{(n+1)}}{\partial x_j} (X_{n+1}(t_k^n)) \sum_{m=1}^{d'} \sigma_{jm}(X_{n+1}(t_k^n)) \Delta B_{m,2k}^{n+1} \frac{\Delta t_m}{2} \right)
\]
\[
M_{i,n,k}^{f,(2)} \triangleq \left( \sum_{j=1}^{d'} \left( \sigma_{ij}(X_{n+1}(t_{2k+1}^{n+1})) - \sigma_{ij}(X_{n+1}(t_k^n)) \right) - \sum_{j,m=1}^{d} \sum_{l=1}^{d} \left( \frac{\partial \sigma_{ij}}{\partial x_l} \cdot \sigma_{lm} \right)(X_{n+1}(t_k^n)) \Delta B_{m,2k+1}^{n+1} \right)
\]
\[
= \sum_{j=1}^{d'} \sum_{m=1}^{d} \frac{\partial \sigma_{ij}}{\partial x_m} (X_{n+1}(t_k^n)) \left( \mu_{m+1}(X_{n+1}(t_k^n)) \right) \Delta t_m \frac{\Delta t_n}{2}
\]
\[
+ \sum_{j', r=1}^{d'} \sum_{l=1}^{d} \frac{\partial \sigma_{ij}}{\partial x_{l'}} \left( \sigma_{ir} \right)(X_{r, n+1}(t_{2k+1}^{n+1}) - X_{r, n+1}(t_k^n)) \Delta B_{j',2k+1}^{n+1}
\]
\[
+ \frac{1}{2} \sum_{j=1}^{d'} \sum_{m,l=1}^{d} \frac{\partial^2 \sigma_{ij}}{\partial x_m \partial x_l}(\rho_7)(X_{m,n+1}(t_{2k+1}^{n+1}) - X_{m,n+1}(t_k^n)) (X_{l,n+1}(t_{2k+1}^{n+1}) - X_{l,n+1}(t_k^n)) \Delta B_{j',2k+1}^{n+1}
\]
\[
M_{i,n,k}^{f,(3)} \triangleq \sum_{j,m=1}^{d'} \sum_{l=1}^{d} \left( \left( \frac{\partial \sigma_{ij}}{\partial x_l} \cdot \sigma_{lm} \right)(X_{n+1}(t_{2k+1}^{n+1})) - \left( \frac{\partial \sigma_{ij}}{\partial x_l} \cdot \sigma_{lm} \right)(X_{n+1}(t_k^n)) \right) \tilde{A}_{mj}(t_{2k+1}^{n+1}, t_k^n)
\]
\[
= \sum_{j,m=1}^{d'} \sum_{l', l=1}^{d} \left( \left( \frac{\partial \sigma_{ij}}{\partial x_l} \cdot \sigma_{lm} \right)(X_{l', n+1}(t_{2k+1}^{n+1}) - X_{l', n+1}(t_k^n)) \tilde{A}_{mj}(t_{2k+1}^{n+1}, t_k^n) \right)
\]
where all the \( \rho_i \) lie somewhere between \( X_{n+1}(t_{2k+1}^{n+1}) \) and \( X_{n+1}(t_k^n) \). Based on the expansion above, we use the fact that \( \mathbb{E}(B(t) - B(s))^4 = O(t - s)^2 \) and \( \mathbb{E}(\tilde{A}_{ij}(s,t))^4 = O((t - s)^4) \),
induction of $X_{n+1}(t^*_k)$ (and $X^n_{i+1}(t^*_k)$) with $\Delta B^n_k$, Lemma 5.1.5 and Lemma 5.1.7 to perform similar analysis on these terms like we did for $N_{i,n,k}$ and $N_{i,n,k}^{(1)}$. For convenience, we omit the detail and conclude that we can find polynomial $P(x) > 1$ for $x > 1$ such that

$$\mathbb{E} R_{i,n,k}^4 \leq \mathcal{P}(L_1) \Delta t^{8(2\alpha - \beta) + 2}. \quad (5.29)$$

Now we are ready to prove the hypothesis in (5.26) by induction on $0 \leq k \leq 2^n$. First of all, when $k = 0$, for $1 \leq i \leq d$, the claim holds since $\xi_{i,n,0} = X^n_i(0) - X_i(0) = x - x = 0$.

Now, fixing $0 \leq k \leq 2^n - 1$ and $1 \leq i \leq d$ suppose the induction hypothesis holds so that we can find $C > 1$ where

$$\mathbb{E} |\xi_{i,n,j}|^4 \leq e^{CL_1 t^*_j} \cdot \Delta t^4 \quad (5.30)$$

for all $0 \leq j \leq k$. We want to show

$$\mathbb{E} |\xi_{i,n,k+1}|^4 \leq e^{CL_1 t^*_k+1} \cdot \Delta t^4 \quad (5.31)$$

for all $1 \leq i \leq d$. To achieve this, we again use (5.10)

$$\mathbb{E}(\xi_{i,n,k+1}) = \mathbb{E}(\xi_{i,n,k}) + 3 \mathbb{E}(\xi_{i,n,0}^3) + 3 \mathbb{E}(\xi_{i,n,k}^3) + 6 \mathbb{E}(\xi_{i,n,k}^2) \quad (5.30)$$

to provide upper bounds for terms in (5.31).

We start with $\eta_{i,n,k}^4$ by observing (5.25) and using (5.5) to find constant $C > 1$ that:

$$(\eta_{i,n,k})^4 \leq C(L_1 \|\xi_{i,n,k}\|_\infty \Delta t_n^4 + L_4 \|\xi_{i,n,k}\|_\infty \Delta (B^{n}_{j,k})^4 + L_4 \|\xi_{i,n,k}\|_\infty (\tilde{A}_{ij}(t^n_k, t^n_{k+1}))^4)
+ |R_{i,n,k}|^4 + L_4 \Delta t_n^{4\left(3 + \frac{2\alpha}{2}\right)\gamma + 4}$$

where we can use the fact that $\mathbb{E}(B(t) - B(s))^4 = O(t - s)^2$ and $\mathbb{E}(\tilde{A}_{ij}(s, t))^4 = O((t - s)^4)$ and (5.29) to conclude:

$$\mathbb{E}\eta_{i,n,k}^4 \leq C(L_1^4 \mathbb{E}\|\xi_{i,n,k}\|_\infty^4 \Delta t_n^4 + L_4^4 \mathbb{E}\|\xi_{i,n,k}\|_\infty^4 \Delta t_n^4 + L_4^4 \mathbb{E}\|\xi_{i,n,k}\|_\infty^4 \Delta t_n^4 + \mathcal{P}(L_1) \Delta t^{8(2\alpha - \beta) + 2} + L_4^4 \omega^4 \Delta t_n^{4\left(3 + \frac{2\alpha}{2}\right)\gamma + 4})$$

$$\leq e^{CL_1 t^*_k} \Delta t^{5 - \delta} (\mathcal{P}(L_1) \Delta t^{1 + 8(2\alpha - \beta) - (4 - \delta) + 2L_4^4 \Delta t^3 + 2L_4^4 \Delta t_n}) \quad (5.31)$$

where the last line follows induction hypothesis, the fact that $(3 + \frac{2\alpha}{2})\gamma > 1$, $4 - \delta < 8(2\alpha - \beta)$ in Definition 5.1.1 and the results in (5.29).

For the bound on $\mathbb{E}(\xi_{i,n,k}^3)$, because of the independence of Brownian increments
with \( \mu(\cdot) \in \mathcal{L}_1 \), we can simplify the Equation (5.20) using martingale property and write
\[
\mathbb{E}(\xi_{i,n,k}^3 \eta_{i,n,k}) \\
= \mathbb{E}\left[ (\xi_{i,n,k})^3 ((\mu_i^{(n)}(X_{n+1}(t^n_k))) - (\mu_i^{(n)}(X_n(t^n_k)))) \Delta t_n + N_{i,n,k}^{(1)} + N_{i,n,k}^f + (\mu_i^{(n+1)} - \mu_i^{(n)})(\bar{X}_{i,n,k}(t^n_k)) \Delta t_n) \right] \\
\le L_1 \left( \mathbb{E}\|\xi_{i,n,k}\|^{4\infty} \Delta t_n + (\mathbb{E}\|\xi_{i,n,k}\|^{4\infty})^{\frac{4}{3}}(\mathbb{E}(N_{i,n,k}^{(1)})^4)^{\frac{1}{2}} + (\mathbb{E}\|\xi_{i,n,k}\|^{4\infty})^{\frac{4}{3}}(\mathbb{E}(N_{i,n,k}^f)^4)^{\frac{1}{2}} + (\mathbb{E}\|\xi_{i,n,k}\|^{4\infty})^{\frac{4}{3}} \Delta t_n^{(3+\frac{4\infty}{4})}\gamma + 1 \right) \\
\le e^{C_1 t_n^\delta} \Delta t_n^{5-\delta}(2L_1 + 2L_1 \Delta t_n^{\frac{1}{4} + 2(2\alpha - \beta) - 1})
\]
where the second inequality follows from Hölder’s inequality and Equation (5.3). The last inequality follows from induction hypothesis, Equations (5.27) and (5.28) and the fact that \((3+\frac{4\infty}{4})\gamma > 1\) and \(8(2\alpha - \beta) > 4 - \delta\) as in Definition 5.1.1.

Similarly, we have
\[
\mathbb{E}\xi_{i,n,k}(\eta_{i,n,k})^3 \le (\mathbb{E}(\xi_{i,n,k})^4)^{\frac{1}{4}}(\mathbb{E}(\eta_{i,n,k})^4)^{\frac{3}{4}} \\
\le e^{C_1 t_n^\delta} \Delta t_n^{5-\delta}(\mathcal{P}(L_1) + 2L_1^4 + 2L_1^4) \Delta t_n^{\frac{1}{4}} \\
\mathbb{E}(\xi_{i,n,k})^2(\eta_{i,n,k})^2 \le (\mathbb{E}(\xi_{i,n,k})^4)^{\frac{1}{2}}(\mathbb{E}(\eta_{i,n,k})^4)^{\frac{1}{2}} \\
\le e^{C_1 t_n^\delta} \Delta t_n^{5-\delta}(\mathcal{P}(L_1) \Delta t_n^{8(2\alpha - \beta) - (4 - \delta)} + 2L_1^4 \Delta t_n^2 + 2L_1^4)^{\frac{1}{2}}
\]
following from Hölder’s inequality and Equation (5.31).

Let \( C = 5 + 2L > 1 \) and find polynomial \( \mathcal{P}'(x) > 1 \) for \( x > 1 \) such that when \( 2^n > \mathcal{P}'(L_1) \), we have
\[
(2L_1 + 2L_1 \Delta t_n^{\frac{1}{4} + 2(2\alpha - \beta) - 1}) \le 3L_1 \\
(\mathcal{P}(L_1) \Delta t_n^{1 + 8(2\alpha - \beta) - (4 - \delta)} + 2L_1^4 \Delta t_n^3 + 2L_1^4 \Delta t_n) \le 1 \\
(\mathcal{P}(L_1) + 2L_1^4 + 2L_1^4) \Delta t_n^{\frac{1}{2}} \le 1 \\
(\mathcal{P}(L_1) \Delta t_n^{8(2\alpha - \beta) - (4 - \delta)} + 2L_1^4 \Delta t_n^2 + 2L_1^4)^{\frac{1}{2}} \le 2L_1^2 
\]
(5.32)

Now we are ready to prove the induction hypothesis. In particular, when \( 2^n > \mathcal{P}'(L_1) \), we use the bound in Equations (5.30) and (5.32) to obtain
\[
\mathbb{E}(\xi_{i,n,k+1})^4 \le e^{C_1 t_n^\delta} \Delta t_n^{4-\delta} + e^{C_1 t_n^\delta} \Delta t_n^{5-\delta}(3L_1 + 2 + 2L_1^2) \\
\le e^{C_1 t_n^\delta} \Delta t_n^{4-\delta} \cdot (1 + C_1 \Delta t_n) \\
\le e^{C_1 t_n^\delta} \Delta t_n^{4-\delta}
\]
where the last line follows from convexity of exponential function \( e^y \ge e^x + e^x \cdot (y - x) \) for \( y \ge x \). Now we can use the same method as in the proof of Lemma 3.2.4 to extend the induction hypothesis to the case where \( \Delta t_n \le \mathcal{P}'(L_1) \) and finish the proof of (5.20) and thus the proof of Lemma 3.2.8.

5.4 Proof of Lemma 2.1.2

We first present two supporting lemmas that are useful in the proof of Lemma 2.1.2.

Lemma 5.4.1. For fixed $t > 0$, there exists some constant $b_t > 3$ large enough such that

$$
\prod_{n=1}^{\infty} \left(1 - \frac{2t}{\sqrt{2\pi n^3} \cdot \log(b)} e^{-\frac{(\log(b))^2 n^{2\epsilon}}{2t^2}} \right) \geq \exp\left\{ -2 \sum_{n \geq 1} \frac{2t}{\sqrt{2\pi n^3} \cdot \log(b)} e^{-\frac{(\log(b))^2 n^{2\epsilon}}{2t^2}} \right\}
$$

and

$$
\frac{1}{n^2} \cdot e^{-\frac{(\log(b))^2}{2t^2}} \geq \frac{1}{n^\epsilon \cdot \log(b)} \cdot e^{-\frac{(\log(b))^2 n^{2\epsilon}}{2t^2}}
$$

$\forall b > b_t$ and $n \geq 1$.

Lemma 5.4.2. Fixing $\epsilon > 0$, define $M_\epsilon \equiv \sup_{n \geq 1} \frac{|V_n|}{n^\epsilon}$ where $\{V_n\}$ are I.I.D. standard normal variables. Then $M_\epsilon$ has finite moment-generating function (i.e., $E[e^{t M_\epsilon}] < \infty$) for all $t \geq 0$.

Proof of Lemma 2.1.2. First, we write $V_n = (V_{n,1}, V_{n,2}, ..., V_{n,d})$, where $\{V_{n,i}\}_{1 \leq i \leq d}$ are I.I.D. standard normal random variables for all $n$ and $i$. Now, using Lemma 5.4.2 and the fact that $\|V_n\|_{\infty} \leq \sum_{i=1}^{d} |V_{n, i}|$, we can deduce

$$
N_{\frac{q-4}{2}} \triangleq \sup_{n \geq 1} \frac{\|V_n\|_{\infty}}{n^{\frac{q-4}{2}}}
$$

also has finite moment-generating function for all $t \geq 0$. Thus, we have

$$
\frac{\|\mu(x) - \mu(y)\|_{\infty}}{\|x - y\|_{\infty}} \leq \sum_{n=1}^{\infty} \frac{|\lambda_n|}{n^{4+\frac{q-2}{2}} n^{\frac{q-4}{2}}} \frac{\|V_n\|_{\infty}}{x - y} |\psi_n(x) - \psi_n(y)| \leq \sum_{n=1}^{\infty} \frac{|\lambda_n|}{n^{3+\frac{q-2}{2}}} \frac{N_{\frac{q-4}{2}} L \leq C N_{\frac{q-4}{2}}},
$$

for some constant $C > 1$ according to Assumptions $A1$ and $A2$. Notice we have found a bound for $\|\frac{\partial \mu}{\partial x^l}\|_{\infty}$ that has finite moment-generating function on the real line. Using a similar method, we can find bounds with finite moment-generating function on the real line for $\mu$ and $\|\frac{\partial^2 \mu}{\partial x^l \partial x^j}\|_{\infty}$. The same bound applies for $S_n$ (thus $\mu^{(n)}$, since $\mu^{(n)} = S_{\lfloor 2n^\gamma \rfloor}$) and $\mu^{(n)}$ for all $n$, and we can thus define a uniform bound finite moment-generating function for all these quantities on the real line denoted by $L_1$. The requirement $L_1 > 1$ can be added without affecting the result.

References

[1] Robert J Adler and Jonathan E Taylor. Random fields and their geometry. Birkhäuser, 2003.

[2] Jean Baptiste Joseph Baron Fourier. The analytical theory of heat. The University Press, 1878.
Christian Bayer, Peter K Friz, Sebastian Riedel, and John Schoenmakers. From rough path estimates to multilevel Monte Carlo. *SIAM Journal on Numerical Analysis, 54*(3):1449–1483, 2016.

Jose Blanchet, Xinyun Chen, and Jing Dong. $\epsilon$-strong simulation for multidimensional stochastic differential equations via rough path analysis. *Ann. Appl. Probab.*, 27(1):275–336, 02 2017.

Jose H Blanchet and Peter W Glynn. Unbiased Monte Carlo for optimization and functions of expectations via multi-level randomization. In *Proceedings of the 2015 Winter Simulation Conference*, pages 3656–3667. IEEE Press, 2015.

D. L. Burkholder, B. J. Davis, and R. F. Gundy. Integral inequalities for convex functions of operators on martingales. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory*, pages 223–240, Berkeley, Calif., 1972. University of California Press.

K Andrew Cliffe, Mike B Giles, Robert Scheichl, and Aretha L Teckentrup. Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients. *Computing and Visualization in Science, 14*(1):3, 2011.

D Crevillén-García and H Power. Multilevel and quasi-Monte Carlo methods for uncertainty quantification in particle travel times through random heterogeneous porous media. *Royal Society open science, 4*(8):170203, 2017.

Alexander M Davie. Differential equations driven by rough paths: an approach via discrete approximation. *Applied Mathematics Research eXpress, 2008*, 2008.

Gh De Marsily, Frédéric Delay, J Goncalves, Ph Renard, Vanessa Teles, and S Violette. Dealing with spatial heterogeneity. *Hydrogeology Journal, 13*(1):161–183, 2005.

JP Delhomme. Spatial variability and uncertainty in groundwater flow parameters: a geostatistical approach. *Water Resources Research, 15*(2):269–280, 1979.

Darrell Duffie. *Dynamic asset pricing theory*. Princeton University Press, 2010.

Peter Friz and Martin Hairer. A course on rough paths. *Preprint*, 2014.

Peter K Friz and Nicolas B Victoir. *Multidimensional stochastic processes as rough paths: theory and applications*, volume 120. Cambridge University Press, 2010.

Michael B Giles. Multilevel Monte Carlo path simulation. *Operations Research, 56*(3):607–617, 2008.
[16] Michael B Giles. Multilevel Monte Carlo methods. In *Monte Carlo and Quasi-Monte Carlo Methods 2012*, pages 83–103. Springer, 2013.

[17] Michael B. Giles and Lukasz Szpruch. Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Levy area simulation. *Ann. Appl. Probab.*, 24(4):1585–1620, 08 2014.

[18] Mike Giles. Multilevel Monte Carlo method for SPDEs and PDEs with uncertain data. In *IMA Conference on Numerical Methods for Simulation*, 2015.

[19] Martin Hairer. Solving the KPZ equation. *arXiv preprint arXiv:1109.6811*, 2011.

[20] Martin Hairer. A theory of regularity structures. *Inventiones mathematicae*, 198(2):269–504, 2014.

[21] Marc Hofmann et al. Lp estimation of the diffusion coefficient. *Bernoulli*, 5(3):447–481, 1999.

[22] Ralph Howard. The Gronwall inequality. *Lecture notes*, 1998.

[23] Mark Kac. On distributions of certain Wiener functionals. *Transactions of the American Mathematical Society*, 65(1):1–13, 1949.

[24] Ioannis Karatzas and Steven Shreve. *Brownian motion and stochastic calculus*, volume 113. Springer Science & Business Media, 2012.

[25] K. Karhunen. *Ueber lineare Methoden in der Wahrscheinlichkeitsrechnung*. Annales Academiae scientiarum Fennicae. Series A. 1, Mathematica-physica. 1947.

[26] P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg, 2011.

[27] Xiaoou Li and Jingchen Liu. A multilevel approach towards unbiased sampling of random elliptic partial differential equations. *arXiv preprint arXiv:1605.06349*, 2016.

[28] M. Loeve. *Probability Theory II*. F.W.Gehring P.r.Halmos and C.c.Moore. Springer, 1978.

[29] Wei-Liem Loh et al. Estimating the smoothness of a Gaussian random field from irregularly spaced data via higher-order quadratic variations. *The Annals of Statistics*, 43(6):2766–2794, 2015.
[30] Terry Lyons and Nicolas Victoir. Cubature on Wiener space. In Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, volume 460, pages 169–198. The Royal Society, 2004.

[31] Terry J Lyons. Differential equations driven by rough signals. Revista Matemática Iberoamericana, 14(2):215–310, 1998.

[32] Siddhartha Mishra, Ch Schwab, and J Šukys. Multi-level Monte Carlo finite volume methods for nonlinear systems of conservation laws in multi-dimensions. Journal of Computational Physics, 231(8):3365–3388, 2012.

[33] Martin Ostoja-Starzewski. Microstructural randomness and scaling in mechanics of materials. CRC Press, 2007.

[34] Minho Park and Aretha Teckentrup. Improved multilevel Monte Carlo methods for finite volume discretisations of Darcy flow in randomly layered media. arXiv preprint arXiv:1506.04694, 2015.

[35] Sergio Pastorello. Diffusion coefficient estimation and asset pricing when risk premia and sensitivities are time varying: A comment. Mathematical Finance, 6(1):111–117, 1996.

[36] Chang-Han Rhee and Peter W. Glynn. Unbiased estimation with square root convergence for sde models. Operations Research, 63(5):1026–1043, 2015.

[37] Kazimierz Sobczyk and David J Kirkner. Stochastic modeling of microstructures. Springer Science & Business Media, 2012.

[38] J.M. Steele. Stochastic Calculus and Financial Applications. Applications of mathematics: stochastic modelling and applied probability. Springer, 2001.

[39] Aretha L Teckentrup, Peter Jantsch, Clayton G Webster, and Max Gunzburger. A multilevel stochastic collocation method for partial differential equations with random input data. SIAM/ASA Journal on Uncertainty Quantification, 3(1):1046–1074, 2015.

[40] Stephen Whitaker. Flow in porous media I: A theoretical derivation of Darcy’s law. Transport in porous media, 1(1):3–25, 1986.

[41] Dongxiao Zhang and Zhiming Lu. An efficient, high-order perturbation approach for flow in random porous media via Karhunen–Loeve and polynomial expansions. Journal of Computational Physics, 194(2):773–794, 2004.
A Proof of Supporting Lemmas

First, we introduce the following Levy-Ciesielski construction of the Brownian motion (see, for example [38]) for the understanding of supporting Lemma 5.1.7.

Lemma A.0.1. Let \( \{U_j^m : 1 \leq j \leq 2^{m-1}, m \geq 1\} \) along with \( U_0^0 \) be a sequence of I.I.D. standard normal random variables, and we define

\[
H(t) \triangleq I(0 \leq t < 1/2) - I(1/2 \leq t \leq 1)
\]

along with its family of functions \( \{H_j^m(t) = 2^m/2H(2^{m-1}t - j + 1) : 1 \leq j \leq 2^{m-1}, m \geq 1\} \) and constant function \( H_0^0(\cdot) = 1 \). Now, if we define \( B(t) \) for \( t \in [0,1] \) by

\[
B(t) \triangleq U_0^0 \int_0^t H_0^0(s)ds + \sum_{m \geq 1} \sum_{j=1}^{2^{m-1}} \left( U_j^m \int_0^t H_j^m(s)ds \right),
\]

then it can be shown that the right-hand side converges uniformly on \([0,1]\) almost surely and the process \( \{B(t) : t \in [0,1]\} \) is a standard Brownian motion on \([0,1]\).

Proof. See Section 2.3 of [24].

This theoretical construction provides ways to sample Brownian motion by sampling independent normal random variables. Here, for \( d' \)-dimensional Brownian motion we use \( d' \)-dimensional normal random variables. Furthermore, using Lemma A.0.1 and the fact that changing the sign of a standard normal variable does not change its distribution, we have the following corollary on \( B^{(n+1),a}(t) \) related to Definition 3.2.2.

Corollary A.0.2. Fixing \( n \geq 0 \) and the sequence of I.I.D. standard normal random variables \( \{U_j^m : 1 \leq j \leq 2^{m-1}, m \geq 1\} \) along with \( U_0^0 \), we can define

\[
B^{n+1,a}(t) \triangleq U_0^0 \int_0^t H_0^0(s)ds + \sum_{j=1}^{2^n} \left( -U_j^{n+1} \int_0^t H_j^{n+1}(s)ds \right) + \sum_{m \geq 1} \sum_{m \neq n+1} \sum_{j=1}^{2^{m-1}} \left( U_j^m \int_0^t H_j^m(s)ds \right),
\]

which is a Brownian motion on \([0,1]\).

Lemma A.0.3. Given a sequence of I.I.D. standard normal \( \{U_j^m : 1 \leq j \leq 2^{m-1}, m \geq 1\} \) along with \( U_0^0 \) and fixing \( n \geq 0 \), define \( B(t), 0 \leq t \leq 1 \) as in (A.1) and \( B^{n+1,a}(t), 0 \leq t \leq 1 \) as in (A.2). Then for \( 1 \leq k \leq 2^{n+1} \), let

\[
\Delta B^{n+1}_k = B(t_{k+1}^{n+1}) - B(t_k^{n+1})
\]

\[
\Delta B^{n+1,a}_k = B^{(n+1),a}(t_{k+1}^{n+1}) - B^{(n+1),a}(t_k^{n+1}).
\]
Then $\Delta B_k^{n+1}$ and $\Delta B_k^{n+1,a}$ satisfy equations (3.8) and thus (3.9) in Definition 3.2.2. Thus, we may regard $X_n^{a+1}(\cdot)$ to be an antithetic scheme $X_n^{a+1}(\cdot)$ generated under Brownian motion $B^{n+1,a}(\cdot)$ instead of $B(\cdot)$.

Proof of Lemma A.0.3 Following Definition A.0.1, fixing $n \geq 1$ and $0 \leq k \leq 2^{n-1} - 1$, we observe that

$$
\begin{aligned}
\int_{t_{2k}}^{t_{2k+1}} H_j^{m}(t)dt &= \int_{t_{2k}}^{t_{2k+1}} H_j^{m}(t)dt \quad \text{for all } m \neq n \text{ and } 1 \leq j \leq 2^{m-1}, \\
\int_{t_{2k}}^{t_{2k+1}} H_j^{m}(t)dt &= -\int_{t_{2k}}^{t_{2k+1}} H_j^{m}(t)dt \quad \text{for all } m = n \text{ and } 1 \leq j \leq 2^{m-1}.
\end{aligned}
$$

(A.3)

Thus, we have that, for $0 \leq k \leq 2^n - 1$,

$$
B^{n+1,a}(m+1) - B^{n+1,a}(t_{2k}) = B^{n+1,a}(t_{2k+1}) - B^{n+1,a}(t_{2k}) = \Delta B^{n+1,a}_{2k+1}
$$

by simply taking the difference in (A.2) and checking (A.3).

Proof of Lemma 3.1.4 Following Definition 5.1.2, define $R^R_{i,j}(t^n_{m}, t^n_{m}) = \sum_{k = l+1}^{m} A_{i,j}(t^n_{k-1}, t^n_{k})$ for $0 \leq l < m \leq 2^n, 1 \leq i, j \leq d'$ and $i \neq j$. Then, we can define

$$
\Gamma_R \triangleq \sup_{n \geq 1} \sup_{0 \leq s \leq t \leq 1} \max_{s \leq l < t \leq d', i \neq j} \left| \frac{R^R_{i,j}(s, t)}{t - s^{1/2}} \right| \quad \text{and} \quad \Gamma_{R-R} \triangleq \sup_{n \geq 1} \sup_{0 \leq s \leq t \leq 1} \max_{s \leq l < t \leq d', i \neq j} \left| \frac{R^R_{i,j}(s, t)}{t - s^{1/2}} \right|.
$$

Observing the definition for both the case $i = j$ and $i \neq j$, we have the following bound:

$$
\| \tilde{A} \|_{2\alpha} \leq \| A \|_{2\alpha} + \| B \|_{2\alpha} \quad \text{and} \quad \Gamma_R \leq \Gamma_R + \Gamma_{R-R}.
$$

(A.4)

Now, following Lemma 3.1 in [4], we define a family of random variables $(L^n_{i,j}(k)) : k = 0, 1, ..., 2^n-1, 1 \leq i, j \leq d', i \neq j; n \geq 1)$ satisfying:

$$
\begin{aligned}
L^n_{i,j}(0) &= 0, \\
L^n_{i,j}(k) &= L^n_{i,j}(k - 1) + (B_i(t^n_{2k-1}) - B_i(t^n_{2k-2})) (B_j(t^n_{2k}) - B_j(t^n_{2k-1})).
\end{aligned}
$$

Then, following Lemma 3.4 and its proof in [4], we define, for $1 \leq i, j \leq d'$ and $i \neq j$,

$$
N_{i,j,2} = \max\{n : |L^n_{i,j}(m) - L^n_{i,j}(l)| > (m - l)^{1/2} \Delta t^n_{2\alpha} \} \quad \text{for some } 0 \leq l < m \leq 2^{n-1},
$$

and define $N_2 = \max\{N_{i,j,2} : 1 \leq i, j \leq d', i \neq j\}$ along with

$$
\Gamma_L \triangleq \max\{1, \max_{1 \leq i, j \leq d', i \neq j} \max_{n < N_2} \max_{0 \leq l < m \leq 2^{n-1}} \left| \frac{L^n_{i,j}(m) - L^n_{i,j}(l)}{(m - l)^{1/2} \Delta t^n_{2\alpha}} \right| \}.
$$

Finally, we can use Definition 3.1.2 and apply the result of Lemma 3.5 in [4] to write:

$$
\Gamma_R \leq \frac{2^{-(2\alpha - 1)}}{1 - 2^{-(2\alpha - 1)}} \cdot \Gamma_L, \quad \text{and} \quad \| A \|_{2\alpha} \leq \Gamma_R \cdot \frac{2}{1 - 2^{2\alpha}} + \| B \|_{2\alpha} \cdot \frac{2^{1-\alpha}}{1 - 2^{2\alpha}}.
$$

(A.5)
Combining the result from (A.4) and (A.5), to conclude the proof, it suffices to show that \( \|B\|_\alpha, \Gamma_L \) and \( \Gamma_{R-\tilde{R}} \) has finite moments of every order. The fact that \( \|B\|_\alpha \) has finite moments of every order follows from Borell’s inequality for continuous Gaussian random fields (see Section 2.3 of [1]). To show \( \Gamma_L \) has finite moments of every order, we first follow the proof of Lemma 3.4 in [4] to show that

\[
\mathbb{P}(N_{i,j,2} \geq n) \leq \sum_{h=n}^{\infty} 2^{2h} \exp(-\theta' 2^{h(1-2\alpha)}) \\
\leq \sum_{h=0}^{\infty} 2^{2h} \exp(-\theta' 2^{h(1-2\alpha)}) \\
\leq C \exp(-\theta' 2^{n(1-2\alpha)})
\]

for some \( C > 1 \) and \( \theta' > 0 \). It follows that,

\[
\mathbb{P}(N_2 \geq n) \leq C(d')^2 \exp(-\theta' 2^{n(1-2\alpha)}), \quad (A.6)
\]

Therefore, we can show

\[
\mathbb{E}(\exp(\eta N_2)) \leq \sum_{n=1}^{\infty} C(d')^2 \exp(\eta n) \exp(-\theta' 2^{n(1-2\alpha)}) < \infty \quad (A.7)
\]

for every \( \eta > 0 \). On the other hand, since for \( m > l, n \leq N_2 \), we have

\[
(m - l)^{-\beta} \Delta t_n^{-2\alpha} = (m - l)^{-\beta} 2^{2\alpha n} \leq 2^{2\alpha N_2}
\]

and thus

\[
\Gamma_L \leq 1 + 2^{2\alpha N_2} \cdot \left( \max_{1 \leq i,j \leq d',i \neq j} \max_{n \leq N_2} \max_{0 \leq l \leq m < 2^{n-1}} |L_{i,j}^n (m) - L_{i,j}^n (l)| \right).
\]

Since \( N_2 \) has a finite moment-generating function on the whole real line according to (A.7), in order to establish that \( \Gamma_L \) has finite moments of every order, it suffices to show that

\[
\mathbb{E} \left[ \left( \sum_{n=1}^{N_2} \sum_{1 < l < m < 2^{n-1}} \sum_{1 \leq i,j \leq d',i \neq j} |L_{i,j}^n (m) - L_{i,j}^n (l)| \right)^k \right] < \infty
\]

for every \( k \geq 1 \). Letting \( \tilde{n} \) be the number of total elements being summed up inside the
previous expectation, it follows that \( \bar{n} \leq N_2 \cdot 2^{2N_2(d')^2} \) and therefore, by (3.15), that

\[
\mathbb{E} \left[ \left( \sum_{n=1}^{N_2} \sum_{1 \leq l < m < 2^n-1} \sum_{1 \leq i,j \leq d', i \neq j} |L_{i,j}^n(m) - L_{i,j}^n(l)| \right)^k \right] \leq \mathbb{E} \left[ \bar{n}^{k-1} \sum_{n=1}^{N_2} \sum_{1 \leq l < m < 2^n-1} \sum_{1 \leq i,j \leq d', i \neq j} |L_{i,j}^n(m) - L_{i,j}^n(l)|^k \right] \leq \sum_{n=1}^{\infty} \sum_{1 \leq l < m < 2^n-1} \sum_{1 \leq i,j \leq d', i \neq j} \mathbb{E} \left[ (N_2 \cdot 2^{2N_2(d')^2})^{k-1} |L_{i,j}^n(m) - L_{i,j}^n(l)|^k I (N_2 \geq n) \right]. \tag{A.8}
\]

To bound the term in (A.8), we first show that, fixing any \( h \geq 1 \), \( \mathbb{E}|L_{i,j}^n(m) - L_{i,j}^n(l)|^h \) is uniformly bounded for any \( n \geq 1, 1 \leq l < m \leq 2^n-1, 1 \leq i, j \leq n \) and \( i \neq j \).

Let \( \{Y_{i'}\}_{i' \geq 1} \) be I.I.D. random variables such that \( Y \overset{d}{=} Z_1 \cdot Z_2 \) where \( Z_1, Z_2 \) are independent standard normal random variables. It follows from Hölder’s inequality and Jensen’s inequality that we can find \( C_h > 0 \) such that \( \mathbb{E}[\frac{\sum_{n=1}^{N_2} Y_{i'}}{n}]^h < C_h \) for all \( n \geq 1 \). Then \( \mathbb{E}|L_{i,j}^n(m) - L_{i,j}^n(l)|^h \) follows from \( |L_{i,j}^n(m) - L_{i,j}^n(l)| \overset{d}{=} |\Delta \bar{t}_n \sum_{l'=1}^{m-l} Y_{i'}| \leq |\sum_{m-l}^{m-l} Y_{i'}| \). Specifically \( \mathbb{E}|L_{i,j}^n(m) - L_{i,j}^n(l)|^{4k} < C_{4k} \) for all \( n \geq 1 \). Now we can use Hölder’s inequality multiple times and the fact that \( N_2 \) has moment-generating function to conclude:

\[
\mathbb{E} \left[ (d')^{2(k-1)} 2^{3N_2(k-1)} |L_{i,j}^n(m) - L_{i,j}^n(l)|^k I (N_2 \geq n) \right] \leq C' f (N_2 \geq n)^{1/2}
\]

for some \( C' > 0 \) and therefore, it follows from (A.6) and the estimate in (A.8) that \( \Gamma_L \) has moments of every order.

Finally, to show that \( \Gamma_{R-R} \) has finite moments of every order, we define another family of random variables \( (\tilde{L}_{i,j}^n(k) : k = 0, 1, \ldots, 2^n, 1 \leq i, j \leq d', i \neq j, n \geq 1) \) satisfying:

\[
\begin{align*}
\tilde{L}_{i,j}^n(0) &= 0 \\
\tilde{L}_{i,j}^n(k) &= \tilde{L}_{i,j}^n(k-1) + (B_i(t_{k}) - B_i(t_{k}) (t_{k})) (B_j(t_{k}) - B_j(t_{k}))
\end{align*}
\]

and similarly define

\[
\tilde{N}_2 = \max\{n : |\tilde{L}_{i,j}^n(m) - \tilde{L}_{i,j}^n(l)| > (m-l)^\beta \Delta t_n \} \quad \text{for some} \quad 0 \leq l < m \leq 2^n, 1 \leq i, j \leq d', i \neq j\}
\]

as well as

\[
\Gamma_L \triangleq \max\{1, \max_{1 \leq i,j \leq d', i \neq j} \max_{n \leq N_2} \max_{0 \leq l < m < 2^n-1} \frac{|\tilde{L}_{i,j}^n(m) - \tilde{L}_{i,j}^n(l)|}{(m-l)^\beta \Delta t_n} \}.
\]

Then, for \( 1 \leq i, j \leq d', i \neq j, n \geq 1 \) and \( 0 \leq s < t \leq 1, s, t \in D_n \), we have

\[
\tilde{R}_{i,j}^n(s, t) - \tilde{R}_{i,j}^n(s, t) = \sum_{k=s2^n+1}^{t2^n} \tilde{A}_{i,j}^n(t_{k-1}) (t_{k} - (t_{k-1})) = \tilde{L}_{i,j}^n(t2^n) - \tilde{L}_{i,j}^n(s2^n),
\]

which implies \( \Gamma_{R-R} \leq \Gamma_L \). We can now proceed to show \( \Gamma_L \) has finite moments of every order in the similar fashion as we did for \( \Gamma_L \). This completes the proof.
Proof of Lemma 6.1.5. Let \( X^M_n(\cdot) \) be the following Milstein discretization scheme with step size \( 2^{-n} \):

\[
X^M_n(t^n_{k+1}) = X^M_n(t^n_k) + \mu_i(X^M_n(t^n_k)) \Delta t_n + \sum_{j=1}^{d'} \sigma_{ij}(X^M_n(t^n_k)) \Delta B^b_{j,k} \\
+ \sum_{j=1}^{d'} \sum_{l=1}^{d} \sum_{m=1}^{d'} \frac{\partial \sigma_{ij}}{\partial x_l}(X^M_n(t^n_k)) \sigma_{lm}(X^M_n(t^n_k)) A_{mj}(t^n_k, t^n_{k+1}).
\]

where we use \( A_{ij}(s, t) \) instead of \( \tilde{A}_{ij}(s, t) \) defined in [5, L.4]. (This distinguishes \( X^M_n(\cdot) \) from \( X_n(\cdot) \), our antithetic scheme.). Then, fixing \( \mu \in \mathcal{L}_1 \) and \( \mu^{(n)} \in \mathcal{L}_1 \) with bounding number \( L_1 \), we can compute constant \( C_1 \) explicitly in terms of \( L_1, \|B\|_\alpha \) and \( \|A\|_{2\alpha} \) (originally denoted as \( M, \|Z\|_\alpha \) and \( \|A\|_{2\alpha} \) in [4]) such that for \( n \) large enough and \( r, t \in D_n \),

\[
\|X^M_n(t) - X^M_n(r)\|_\infty \leq C_1 |t - r|^\alpha
\]

See page 305 of [4, Lemma 6.1]. To get the result for \( X_n(\cdot) \) instead of \( X^M_n(\cdot) \), we follow page 283 of [4, Lemma 2.1], replacing \( \|A\|_{2\alpha} \) by \( \|\tilde{A}\|_{2\alpha} \) in notation, we define

\[
\begin{align*}
C_1(\delta) &= \tilde{d}L_1\|B\|_\alpha + 1/2 \\
C_2(\delta) &= \tilde{d}^3L_1^2\|A\|_{2\alpha} + 1/2 \\
C_3(\delta) &= \frac{2}{1-2^{-2x}}(\tilde{d}L_1C_1(\delta)^2\|B\|_\alpha + \tilde{d}^2L_1C_2(\delta)\|B\|_\alpha + \tilde{d}^3L_1^2\|B\|_\alpha + 2\tilde{d}^3L_1^2C_1(\delta)\|A\|_{2\alpha})
\end{align*}
\]

and we then find some fixed polynomial \( P(x) > 1 \) for \( x > 1 \) so that if \( \delta = \left( P(L_1, \|B\|_\alpha, \|A\|_{2\alpha}) \right)^{-1} \), then

\[
C_3(\delta)^{2\alpha} + L_1\delta^{1-\alpha} + \tilde{d}^3L_1^2\|A\|_{2\alpha}\delta^{\alpha} < 1/2 \quad \text{also} \quad C_3(\delta)^{\alpha} < 1/2,
\]

so that Equation (6.4) in page 308 of [4, Lemma 6.1] is satisfied:

\[
\begin{align*}
C_1(\delta) &\geq \tilde{d}L_1\|B\|_\alpha + L_1\delta^{1-\alpha} + \tilde{d}L_1\|B\|_\alpha + \tilde{d}^3L_1^2\|\tilde{A}\|_{2\alpha}\delta^{\alpha} \\
C_2(\delta) &\geq \tilde{d}^3L_1^2\|A\|_{2\alpha} + \tilde{d}^3L_1^2\|\tilde{A}\|_{2\alpha} \\
C_3(\delta) &\geq \frac{2}{1-2^{-2x}}(\tilde{d}L_1C_1(\delta)^2\|B\|_\alpha + \tilde{d}^2L_1C_2(\delta)\|B\|_\alpha + \tilde{d}^3L_1^2\|B\|_\alpha + 2\tilde{d}^3L_1^2C_1(\delta)\|A\|_{2\alpha})
\end{align*}
\]

which gives, according to line 12 – 17 of page 308 of [4, Lemma 6.1], that

\[
\|X_n(t) - X_n(r)\|_\infty \leq \frac{2}{\delta} C_1(\delta)|t - r|^\alpha \quad \text{(A.9)}
\]

for all \( n \) large enough where \( \Delta t_n \leq \frac{1}{2}\delta \). Notice here we have changed the result to address \( X_n(\cdot) \) instead of \( X^M_n(\cdot) \), and so far it just follows from an easy modification of [4, Lemma 6.1].
Now, to extend the result for \( n \) where \( \Delta t_n > \frac{\delta}{2} \), notice the recursion step in (3.6) is carried out at most \( 2^n \) number of times where \( 2^n = (\Delta t_n)^{-1} < 2(\delta)^{-1} = 2\mathcal{P}(L_1, \|B\|_\alpha, \|A\|_{2\alpha}) \). By analyzing (3.6) term by term, we have

\[
\|X_n(t^n_{k+1}) - X_n(t^n_k)\|_\infty \leq d(C_1 \Delta t_n + \bar{d}L\|B\|_\alpha + \bar{d}^3 L^2\|A\|_{2\alpha} \Delta t^n_{2\alpha})
\]

\[
\leq d(C_1 + \bar{d}L\|B\|_\alpha + \bar{d}^3 L^2\|A\|_{2\alpha}) \Delta t^n_{2\alpha}
\]

for some \( C > 1 \). Since \( \Delta t_n < 1 \), thus, for \( \Delta t_n > \frac{\delta}{2} \),

\[
\|X_n(t) - X_n(r)\|_\infty \leq \left| \frac{t - r}{\Delta t_n} \right| \bar{d}(C_1 + \bar{d}L\|B\|_\alpha + \bar{d}^3 L^2\|A\|_{2\alpha}) \Delta t^n_{2\alpha}
\]

\[
\leq \bar{d}(C_1 + \bar{d}L\|B\|_\alpha + \bar{d}^3 L^2\|A\|_{2\alpha}) \Delta t^n_{2\alpha} \leq \bar{d}(C_1 + \bar{d}L\|B\|_\alpha + \bar{d}^3 L^2\|A\|_{2\alpha}) \mathcal{P}(L_1, \|B\|_\alpha, \|A\|_{2\alpha}) \cdot |t - (\bar{A})_{10}|
\]

where the last line follows from \( \mathcal{P}(x) > 1 \) for \( x > 1 \). The second to last line follows from \( \Delta t_n > \frac{\delta}{2} \). We now combine (A.9) and (A.10) and let

\[
\mathcal{P}'(L_1, \|B\|_\alpha, \|A\|_{2\alpha}) \triangleq 2\bar{d}(C_1 + \bar{d}L\|B\|_\alpha + \bar{d}^3 L^2\|A\|_{2\alpha}) \cdot \mathcal{P}(L_1, \|B\|_\alpha, \|A\|_{2\alpha}) \cdot \frac{2}{\delta} C_1(\delta)
\]

be the polynomial where

\[
\|X_n(t) - X_n(r)\|_\infty \leq \mathcal{P}'(L_1, \|B\|_\alpha, \|\tilde{A}\|_{2\alpha}) |t - r|^\alpha
\]

for all \( n \). This completes the proof. \( \square \)

**Proof of Lemma 5.1.6** The discretization scheme \( \hat{X}^n(\cdot) \) from Equation (2.4) on page 280 of [4] is defined as:

\[
\hat{X}_i^n(t^n_{k+1}) = \hat{X}_i^n(t^n_k) + \mu_i(\hat{X}_i^n(t^n_k)) \Delta t_n + \sum_{j=1}^{d'} \sigma_{ij}(\hat{X}_i^n(t^n_k)) \Delta B_{ij}^n
\]

\[
+ \sum_{j=1}^{d'} \sum_{l=1}^{d} \sum_{m=1}^{d'} \frac{\partial \sigma_{ij}(\hat{X}_i^n(t^n_k))}{\partial x_l} \sigma_{lm}(\hat{X}_i^n(t^n_k)) \hat{A}_{ij}(t^n_k, t^n_{k+1})
\]

where \( \hat{A}_{ij}(s, t) = 0 \) for \( i \neq j \) and \( \hat{A}_{i,i}(s, t) = A_{i,i}(s, t) \forall 1 \leq i \leq d \) as in Definition 5.1.2. Consequently, it is defined on page 280 of [4], as in Definition 5.1.2, that

\[
R_{i,j}^n(t^n_s, t^n_m) \triangleq \sum_{k=t+1}^{m} \{ A_{i,j}(t^n_{k-1}, t^n_k) - \hat{A}_{i,j}(t^n_{k-1}, t^n_k) \}
\]

and \( \Gamma_R \triangleq \sup_n \sup_{0 \leq s, t \leq 1} \max_{1 \leq i, j \leq d'} \frac{|R_{i,j}^n(s, t)|}{|t - s|^{\beta} \Delta t^n_{2\alpha - \beta}}. \)
With a slight change in notation, we replace $M$ with $L_1(\omega)$, $\|Z\|_{\alpha}$ with $\|B\|_{\alpha}$, then according to [4, Theorem 2.1], we can find constant $G$ (for notation consistency with [4]) explicitly in terms of $L_1, K_\alpha, K_{2\alpha}$ and $K_R$ such that

$$\|\hat{X}^n(t) - X_t\|_\infty \leq G \Delta t_n^{2\alpha - \beta}$$

where we may take $K_\alpha = \|B\|_{\alpha}, K_{2\alpha} = \|A\|_{2\alpha}$ and $K_R = \Gamma_R + 1$.

To prove a similar result for $\|X^n(t) - X_t\|_\infty$ instead of $\|\hat{X}^n(t) - X_t\|_\infty$, we replace $\Gamma_R$ with our $\tilde{\Gamma}_\alpha$ defined in Definition 5.1.2 the proof will follow exactly as in the proof of Theorem 2.1 in [4][Proposition 6.1 and 6.2]. Particularly, we are able to compute constant $G$ such that $\|X^n(t) - X_t\|_\infty \leq G \Delta t_n^{2\alpha - \beta}$.

Moreover, following Section 2.2 on pages 282 - 283 of [4] (part of which is shown in Lemma (5.1.5)), the construction of the constant $G$ only involves multiplication and addition among the variables $L_1, \|B\|_{\alpha}, \|A\|_{2\alpha}, \Gamma_R$ and constants. This suggests that we can find some fixed polynomial $P''(\cdot)$ such that $P''(x) > 1$ for $x > 1$

$$\|X^n(t) - X_t\|_\infty \leq P''(L_1, \|B\|_{\alpha}, \|A\|_{2\alpha}, \Gamma_R) \Delta t_n^{2\alpha - \beta}.$$  

\[\square\]

**Remark.** A technical detail here is that the construction of the constant $G$ in Section 2.2 of [4] for Theorem 2.1 actually only makes the statement of Theorem 2.1 valid for $n$ “large” enough, see the proof of Theorem 2.1 in Section 6 of [4][Proposition 6.1 and 6.2]. However, we may extend the result to hold for all $n$ using the similar method in our (A.10) of Lemma 5.1.5. There we modified the proof to extend the result originally only valid for $n$ “large” enough, meaning $\Delta t_n \leq \frac{\delta}{2}$ for $\delta = (P_1(L_1, \|B\|_{\alpha}, \|A\|_{2\alpha}))^{-1}$, to all $n$ while still maintaining the bound $P'(\cdot)$ to be some polynomial of $L_1, \|B\|_{\alpha}$ and $\|A\|_{2\alpha}$. The situation is similar here, and by a similar but more lengthy argument, we can extend the result of Lemma (5.1.6) to hold for all $n$ while still making the upper bound of $G$ above, namely $P''(\cdot)$, to be a polynomial of $L_1, \|B\|_{\alpha}, \|A\|_{2\alpha}$.

**Proof of Lemma 5.1.7.** Denote $X(t; \mu, B), 0 \leq t \leq 1$ to be the solution of SDE under field $\mu(\cdot) \in \mathcal{L}_1$ and Brownian motion $B(t), 0 \leq t \leq 1$. Let $X_n(t; \mu^{(n)}, B)$ be our antithetic scheme under the field $\mu^{(n+1)} \in \mathcal{L}_1$ instead of $\mu^{(n)}$. Since the Brownian increments $\Delta B_k, 1 \leq k \leq 2^n$ are the same for $B(\cdot)$ and $B^{(n+1),a}(\cdot)$ by (3.9), we have that

$$X_n(1; \mu^{(n+1)}, B) = X_n(1; \mu^{(n+1)}, B^{n+1,a})$$
and thus,
\[
\|X_{n+1}(1) - X_{n+1}^a(1)\|_\infty \\
\leq \|X_{n+1}(1) - X_n(1; \mu^{(n+1)}, B)\|_\infty + \|X_n^a(1) - X_n(1; \mu^{(n+1)}, B^{n+1,a})\|_\infty \\
\leq \|X_{n+1}(1) - X(1; \mu^{(n+1)}, B)\|_\infty + \|X_n(1; \mu^{(n+1)}, B) - X(1; \mu^{(n+1)}, B)\|_\infty \\
+ \|X_n^a(1) - X(1; \mu^{(n+1)}, B^{n+1,a})\|_\infty + \|X_n(1; \mu^{(n+1)}, B^{n+1,a}) - X(1; \mu^{(n+1)}, B^{n+1,a})\|_\infty \\
\leq 2 \Big( P''(L_1, \|B\|_\alpha, \|A\|_{2\alpha}, \Gamma \tilde{R}) + P''(L_1, \|B^{n+1,a}\|_\alpha, \|A^{n+1,a}\|_{2\alpha}, \Gamma \tilde{R}^{n+1,a}) \Big) \Delta t_n^{2\alpha - \beta} \\
(A.11)
\]

The last line follows from Lemma 5.1.6 where quantity \(\|B^{n+1,a}\|_\alpha, \|A^{n+1,a}\|_{2\alpha}, \Gamma \tilde{R}^{n+1,a}\) is defined for \(B^{n+1,a}(\cdot)\) as for \(B(\cdot)\) in Definition 5.1.1. Now, raising inequality (A.11) to the eighth power and using Lemma 5.1.4, we can find polynomial \(P_4(x) > 1\) for \(x > 1\) such that
\[
\mathbb{E}_{\mu(\omega)}[\|X_{n+1}(1) - X_{n+1}^a(1)\|_\infty^8] \leq P_4(L_1) \Delta t_n^{8(2\alpha - \beta)}
\]
for all \(n\).

Proof of Lemma 7.4.7 Using elementary calculus, we have that \(log(1 - r) \geq -2r\) for all \(0 \leq r \leq \frac{1}{2}\). Thus, to show (5.33), we only need to show
\[
\frac{2t}{\sqrt{2\pi n^\epsilon \cdot \log(b)}} e^{-\frac{t \log(b)^2 n^{2\epsilon}}{2t^2}} \leq \frac{1}{2} \tag{A.12}
\]
\(\forall n \geq 1\) and for this we can choose \(b_{t,1} > 3\) large enough such that
\[
\frac{2t}{\sqrt{2\pi \cdot \log(b)}} e^{-\frac{t \log(b)^2}{2t^2}} \leq \frac{1}{2},
\]
thus (A.12) will hold \(\forall n \geq 1, b > b_{t,1}\).

To check (5.34), if we pick \(b_{t,2} > 3\), we have
\[
\frac{1}{n^\epsilon} e^{-\frac{(t \log(b))^2 n^{2\epsilon}}{2t^2}} \geq \frac{1}{n^\epsilon \cdot \log(b)} e^{-\frac{(t \log(b))^2 n^{2\epsilon}}{2t^2}}.
\]

Thus, if we write \(\rho(b) \triangleq e^{-\frac{(t \log(b))^2}{2t^2}}\), for (5.34) to hold, we just need to make sure that we can find \(b_{t,2}\) large so that \(\rho(b_{t,2})\) is small enough and satisfies:
\[
\frac{1}{n^\epsilon} \rho(b_{t,2})^{n^{2\epsilon}} \leq \frac{1}{n^2} \rho(b_{t,2}) \iff n^{2-\epsilon} \rho(b_{t,2})^{n^{2\epsilon-1}} \leq 1 \iff (2-\epsilon) \log(n) + (n^{2\epsilon-1}) \log(\rho(b_{t,2})) \leq 0
\]
for \(n \geq 1\) where the last equivalent condition is easy to check using calculus by setting \(\rho(b_{t,2})\) small enough.

Finally, we take \(b_t \triangleq \max\{b_{t,1}, b_{t,2}\}\) to conclude the proof. \qed
Proof of Lemma 5.4.2. Using tail bound that \( \int_{\xi}^{\infty} e^{-\frac{t^2}{2}} dt \leq \frac{1}{\xi} e^{-\frac{\xi^2}{2}} \) for all \( \xi > 0 \), we obtain
\[
\mathbb{P}(M > b) = 1 - \mathbb{P}(M \leq b) = 1 - \prod_{n=1}^{\infty} \mathbb{P}(|Z_n| \leq bn^\varepsilon) \leq 1 - \prod_{n=1}^{\infty} \mathbb{P}(1 - \frac{2}{\sqrt{2\pi \cdot bn^\varepsilon}} e^{-\frac{b^2 n^2 \varepsilon^2}{2}}), \tag{A.13}
\]
Thus, we have
\[
\mathbb{E}[e^{tM}] = \int_0^\infty \mathbb{P}(e^{tM} > b) db \\
\leq b_t + \int_{b_t}^\infty \mathbb{P}(M > \frac{\log(b)}{t}) db \quad \text{(where } b_t \text{ is from Lemma 5.4.1)} \\
\leq b_t + \int_{b_t}^\infty (1 - \prod_{n=1}^{\infty} (1 - \frac{2t}{\sqrt{2\pi n^\varepsilon \cdot \log(b)}} e^{-\frac{\log(b)^2 n^2 \varepsilon^2}{2t^2}})) db \quad \text{(by using (A.13))} \\
\leq b_t + \int_{b_t}^\infty (1 - e^{-2 \sum_{n \geq 1} \frac{2t}{\sqrt{2\pi n^\varepsilon \cdot \log(b)}} e^{-\frac{\log(b)^2 n^2 \varepsilon^2}{2t^2}}}) db \quad \text{(using (5.33))} \\
\leq b_t + \int_{b_t}^\infty (2 \sum_{n \geq 1} \frac{2t}{\sqrt{2\pi n^\varepsilon \cdot \log(b)}} e^{-\frac{\log(b)^2 n^2 \varepsilon^2}{2t^2}}) db \quad \text{(using (5.34))} \\
\leq b_t + \int_{b_t}^\infty (2 \sum_{n \geq 1} \frac{2t}{\sqrt{2\pi n^2}} e^{-\frac{\log(b)^2}{2t^2}}) db \quad \text{(where } C \text{ is some fixed constant)} \\
\leq b_t + C \cdot \int_{b_t}^\infty \frac{\log(b)}{2t^2} db \quad \text{(change of variable by } \log(b) = s) \\
\leq b_t + C \cdot \int_{\log(b_t)}^{\infty} e^{-\frac{s^2}{2t^2}} ds < \infty
\]
\( \square \)