On the Monte Carlo simulation of BSDE’s: an improvement on the Malliavin weights

D. Crisan*, K. Manolarakis†, N. Touzi ‡

Abstract

Monte Carlo simulation of backward stochastic differential equations offers a probabilistic numerical method for the approximation of solutions of semilinear PDEs. We propose a generic framework for the analysis of such numerical approximations. The general results are used to re-visit the convergence of the algorithm suggested by Bouchard and Touzi [6]. By dropping the higher order terms in the Skorohod integrals resulting from the Malliavin integration by parts in [6], we introduce a variant of the latter algorithm which allows for a significant reduction of the numerical complexity. We prove the convergence of this improved Malliavin-based algorithm, and we derive a bound on the induced error. In particular, we show that the price to pay for our simplification is to use a more accurate localizing function.

Key words: BSDEs, weak approximations, Monte Carlo methods, Malliavin calculus

Mathematics subject classification: 65C30; 65C05; 60H07; 62G08

*Department of Mathematics, Imperial College London, 180 Queen’s Gate, London, SW7 2AZ, UK.
†Department of Mathematics, Imperial College London, 180 Queen’s Gate, London, SW7 2AZ, UK.
‡Centre de Mathématiques Appliques Ecole Polytechnique UMR CNRS 7641 91128 Palaiseau Cedex FRANCE
1 Introduction

Let \((\Omega, \mathcal{F}, P)\) be a probability space on which we have defined the decoupled forward-backward stochastic differential equation (BSDE henceforth)

\[
\begin{align*}
    dX_t &= b(X_t)dt + \sigma(X_t)dW_t \\
    -dY_t &= f(t, X_t, Y_t, Z_t)dt - Z_t dW_t
\end{align*}
\] (1.1)

with the initial condition \(X_0 = x\) for the forward component and the final condition \(Y_1 = \Phi(X_1)\) for the backward component. In (1.1), \(W = \{W_t, t \geq 0\}\) is a \(d\)-dimensional Brownian motion. The coefficients of the system (1.1) are given by the function \(b : \mathbb{R}^d \to \mathbb{R}^d\), \(\sigma : \mathbb{R}^d \to \mathbb{R}^{d \times d}\) for the forward component and \(f : [0, 1] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}\) for the backward part. Also \(\Phi : \mathbb{R}^d \to \mathbb{R}\) is a real valued function. A solution to the forward backward system (1.1) is the triplet \(\Theta_t = (X_t, Y_t, Z_t)\). The approximation of the forward part of the solution \(X\) is well understood (see, for example [14] and the references therein).

The focus of this paper is on obtaining approximations for the process \(Y\). Discrete-time schemes based on the approximation of the Brownian motion have been analysed by Chevance [7], Coquet, Macquevicius and Mémin [11], Briand, Delyon and Mémin [17], Antonelli and Kohatsu-Higa [1], Ma, Protter, San Martin and Torres [13], Bally and Pagès [2].

In contrast with the above literature, we concentrate on approximations of solutions of backward stochastic differential equations based on the Monte Carlo simulation of the underlying Brownian motion, thus continuing the studies of Bouchard and Touzi [6], Zhang [18] and Gobet, Lemor and Warin [9]. See also Bouchard and Elie [5] for the case of jump-diffusions.

As a first contribution, we propose a generic framework for the analysis of such approximations. By isolating the one-step approximation operator of the algorithm, we provide a transparent set of sufficient conditions in order to study the rate of convergence of the approximation. The main tool is an expansion of the error as a Trotter product. This generic framework applies to various numerical methods. For instance, the above
methodology is applied in [8] in order to prove convergence of a probabilistic numerical algorithm where the regressions are estimated by using the cubature method of Lyons and Victoir [15] to approximate the distribution of \((X, W)\).

This methodology is applied to the Malliavin integration-by-parts based algorithm suggested in [6] whose main idea is to exploit the representation of the regression function

\[
r(x) := \mathbb{E}[Y|X = x] = \frac{\mathbb{E}[Y\varepsilon_{\{x\}}(X)]}{\mathbb{E}[\varepsilon_{\{x\}}(X)]},
\]

where \(\varepsilon_{\{x\}}\) denotes the Dirac mass at \(x\), as

\[
r(x) = \mathbb{E}\left[\frac{Y\mathbf{1}_{\mathbb{R}^d_+}(X-x)S^h(\phi(X-x))}{\mathbb{E}[\mathbf{1}_{\mathbb{R}^d_+}(X-x)S^h(\phi(X-x))]}\right]
\]

for smooth random variables \(X\) and \(Y\) valued respectively in \(\mathbb{R}^d\) and \(\mathbb{R}^k\), for some integers \(d, k \geq 1\). Here \(S^h\) is an iterated Skorohod integral, and \(\phi\) is a smooth function with \(\phi(0) = 1\), called localizing function.

The Malliavin integration-by-parts algorithm of [6] turns out to have high numerical complexity mainly due to the two following facts:

- the Heaviside function \(\mathbf{1}_{\mathbb{R}^d_+}\) obtained by integrating up the Dirac measure; this numerical difficulty can be dealt with by appealing to advanced sorting algorithms from the operations research literature, see e.g. [3].

- the weights obtained from the Malliavin integration-by-parts consist of iterated Skorohod integrals with exploding number of terms when converted into Itô stochastic integrals (which is necessary for the simulation step).

The second main contribution of this paper is to suggest an improvement of the Malliavin integration-by-parts based algorithm by reducing the numerical complexity of the above weights. To do this, we first re-visit completely the convergence results of [6]. In particular our methodology allows to clarify some points in the proofs of [6]. We next suggest to drop higher order terms of the weights by a judicious expansion. This leads to a reduction of the number of terms at least from \(4^d\) to \(2^d\), where \(d\) is the dimension of the forward
state variable. Hence the number of elementary calculations is at least cut by \(2^d\), for instance with \(d = 5\) we reduce the effort by more than 96%!

The price to pay for such a reduction in the numerical complexity is a change of scale in the localizing function \(\phi^h(x) := \phi(xh^{-\alpha})\) of [4] and [6], where \(h\) is the time step. While \(h = 1/2\) in [6] is the optimal scale, we find in the present setting that \(\alpha\) must be chosen strictly larger than 1/2.

The paper is organised as follows. After collecting some preliminaries in Section 2, we report our one-step operator approach for the discretisation of BSDEs in Section 3. We provide a review of the Monte Carlo method of [6] in Section 4. Section 5 introduces our main modification by truncating the weights, and provides the corresponding asymptotic results for the error analysis.

## 2 Preliminaries

We start first by giving some notation and the assumptions that we will be using throughout the paper. We are working on a complete probability space \((\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\})\) and on finite time interval \([0, 1]\) where a \(d\) dimensional Brownian motion is defined \(W_t, t \in [0, 1]\). The filtration of the space is the augmented Brownian filtration. We denote by \(\mathbb{M}^n\) the space of \(n \times n\) matrices equipped with the norm \(|A| := \left(\sum_{i,j=1}^n |a_{ij}|^2\right)^{1/2}\). We denote by \(A^*\) the transpose of an element of \(\mathbb{M}^n\). The Euclidean spaces are equipped with the usual Euclidean norm. The inner product of two elements in \(\mathbb{R}^d\) is denoted by \(x \cdot y\).

We will be using the following assumptions on the coefficients of system (1.1)

**(H1a)** The functions \(b : \mathbb{R}^d \to \mathbb{R}^d, \quad \sigma : \mathbb{R}^d \to \mathbb{M}^d\) are Lipschitz continuous

\[|b(x) - b(y)| + |\sigma(x) - \sigma(y)| \leq K|x - y|\]

and \(\sigma\) is uniformly positive definite.

**(H1b)** The functions \(b, \sigma, \sigma^{-1} \in C_b^\infty\).
(H2) The driver of the BSDE $f : [0, 1] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ is Lipschitz in the spatial variables

$$|f(t, x, y, z) - f(t, x', y', z')| \leq K(|x - x'| + |y - y'| + |z - z'|)$$

The terminal condition is a Lipschitz function of the state process $\xi = \Phi(X_1)$ with the same Lipschitz constant as above.

A solution to the system (1.1) is an adapted triplet of processes $(X, Y, Z)$ taking values in $\mathbb{R}^d$, $\mathbb{R}$, $\mathbb{R}^d$, respectively, such that

$$\mathbb{E} \left[ \sup_{0 \leq t \leq 1} |Y_t|^2 + \int_0^1 |Z_s|^2 ds \right] < \infty \quad (2.1)$$

Following for example, theorem 2.1 of [10], such a solution exists when conditions (H1a) and (H2) are satisfied. In what follows we will be primarily working with these two hypothesis. The enhanced condition (H1b) will be required to show additional smoothness, in the Malliavin sense, for the Euler scheme associated with the forward component of (3.2).

3 Discretizing the BSDE

Consider a partition of the time interval $\pi = \{0 = t_0 < t_1 < \ldots < t_n = 1\}$ along which, we aim to discretize the system (1.1). Denote by $\pi_i = t_i - t_{i-1}$ and $|\pi| := \max_{1 \leq i \leq n} \pi_i$ the partition mesh. For the discretization of the Brownian motion we write $\Delta W_{i+1} = W_{t_{i+1}} - W_{t_i}$ and $\Delta W^j_{i+1}$ for its $j$-th entry, where $j = 1, 2, \ldots, d$. The forward component of system (1.1) is approximated by an Euler scheme. In particular if

$$X^\pi_t(s, x) := x + b(x)(t - s) + \sigma(x)(W_t - W_s)$$

then the Euler scheme corresponding to the partition $\pi$ is defined as

$$X^\pi_{i+1} = X^\pi_i(t_{i-1}, X^\pi_{i-1}), \quad i = 1, \ldots, n.$$
The following estimate is standard (see theorem 10.2.2 [14])

$$\max_{0 \leq t \leq n-1} \mathbb{E} \left[ \sup_{t_i \leq t \leq t_{i+1}} |X_t - X_t^\pi|^2 \right] \leq C(1 + |x|^2)|\pi| \quad (3.1)$$

For the discretization of the backward components of the solution we define recursively on $\pi$ the random variables $Y_{t_i}^\pi, Z_{t_i}^\pi, i = 0, \ldots, n$, as follows:

$$Y_{t_n}^\pi = \Phi(X_{t_n}^\pi), \quad Z_{t_n}^\pi = 0$$

$$Z_{t_{i-1}}^\pi = \pi_{i-1}^{-1}\mathbb{E}_{t_{i-1}} [Y_{t_i}^\pi \Delta W_{t_i}]$$

$$Y_{t_{i-1}}^\pi = \mathbb{E}_{t_{i-1}} [Y_{t_i}^\pi] + \pi_i f(t_{i-1}, X_{t_{i-1}}^\pi, Y_{t_{i-1}}^\pi, Z_{t_{i-1}}^\pi)$$

Above $\mathbb{E}_t$ stands for conditional expectation with respect to the $\sigma$-field $\mathcal{F}_t^\pi$, the natural filtration of the Euler scheme, $\mathcal{F}_t^\pi = \sigma \left( X_{t_j}^\pi : j \leq i \right)$. When the terminal condition $\xi$ is not path dependent, the random variables $Y_{t_i}^\pi, Z_{t_i}^\pi$ are (deterministic) functions of the underlying $X_{t_i}^\pi$ and the above conditional expectations are actually regressions

$$\mathbb{E}_{t_{i-1}} [Y_{t_i}^\pi \Delta W_{t_i}] = \mathbb{E} \left[ Y_{t_i}^\pi \Delta W_{t_i} | X_{t_{i-1}}^\pi \right], \quad \mathbb{E}_{t_i} [Y_{t_i}^\pi] = \mathbb{E} \left[ Y_{t_i}^\pi | X_{t_{i-1}}^\pi \right]$$

The proof of the following theorem may be found in [6] or in a slight different formulation in [18] for the path dependent case.

**Theorem 3.1.** Associated with the backward scheme (3.2) we define the step processes $Y_t^\pi = \sum_{i=0}^{n-1} Y_{t_i}^\pi 1_{(t_i, t_{i+1})}(t) + Y_{t_n}^\pi 1_{(t_n)}(t)$ and respectively $Z_t^\pi = \sum_{i=0}^{n-1} Z_{t_i}^\pi 1_{(t_i, t_{i+1})}(t) + Z_{t_n}^\pi 1_{(t_n)}(t)$. Then

$$\mathbb{E} \left\{ \sup_{0 \leq t \leq T} |Y_t - Y_t^\pi|^2 + \int_0^1 |Z_t - Z_t^\pi|^2 dt \right\} \leq C|\pi| \quad (3.3)$$

for a constant $C$ independent of the partition.

The above scheme is not directly implementable. Special care needs to be taken for the computation of the involved conditional expectations. However it does highlight the important fact that, to obtain a fully implementable scheme we need to approximate the regression functions as well as the underlying process. In this spirit we start by defining the following operators
Definition 3.2. Let $C^1_b = \{ g : \mathbb{R}^d \to \mathbb{R}, \text{Lipschitz} \}$. Then define the non linear operators $R_i : C^1_b \to C^1_b$, $i = n, \ldots, 0$ as follows

\[
R_n g(x) = g(x) \\
R_i g(x) = \mathbb{E} \left[ g(X^\pi_{t_{i+1}}(t_i, x)) \right] + \pi_{i+1} f \left( t_i, x, R_i g(x), \pi_{i+1}^{-1} \mathbb{E} \left[ g(X^\pi_{t_{i+1}}(t_i, x)) \Delta W_{t+1} \right] \right)
\]  

(3.4)

Remark 3.3. In this paper we will work with forward backward systems where the terminal condition is a Lipschitz function of $X_1$. When an operator from the above family is applied to a function which is $C^1_b$ and the driver $f$ of the BSDE is one time continuously differentiable then, the function $R_i g$ is also differentiable with bounded first order derivatives. This is just an easy consequence of the chain rule and the implicit function theorem (see [18] for details). In fact, there exists a uniform bound on the first derivative of the iteration of the family \{R_i\},

\[
\max_{0 \leq i \leq n-1} \| \nabla R_i \ldots R_{n-1} \Phi(x) \|_\infty < \infty
\]

Note that the operator $R_i$ is well defined for a larger class of functions $g$, not necessarily in $C^1_b$. For instance, $R_i g$ could be defined for functions $g$ of polynomial growth.

Using a backward induction argument we have the following representation

\[
Y^\pi_{t_i} = R_i R_{i+1} \ldots R_{n-1} \Phi(X^\pi_{t_i})
\]  

(3.5)

for $i = 0, 1, \ldots, n$. The next result uses the notation:

\[
\| f \|_{L^p(P_{t_i, t_{i+1}})} := \left( \mathbb{E} \left[ |f(X^\pi_{t_i})|^p \right] \right)^{1/p}
\]

Lemma 3.4. For every $i = 0, \ldots, n-1$ the operator $R_i$ enjoys the following Lipschitz-type property :

\[
|R_i g_1 - R_i g_2| (x) \leq \frac{1 + C_{\pi_{i+1}}}{1 - K_{\pi_{i+1}}} \| g_1 - g_2 \|_{L^p(P_{t_i, t_{i+1}})}
\]  

(3.6)

for some $p > 1$, where $K$ is the Lipschitz constant of $f$ and $C$ some generic constant which depends on $d, p, \text{and } K$. 

7
Proof. Let us fix a value for \( i \). Starting from the definition of \( R_i \) we have that

\[
R_i g_1(x) - R_i g_2(x) = \mathbb{E} \left[ (g_1 - g_2)(X_{t_i+1}^1(t_i, x)) \right] \\
+ \pi_{i+1} \left\{ f \left( t_i, x, R_i g_1(x), \frac{1}{\pi_{i+1}} \mathbb{E} \left[ g_1(X_{t_i+1}^1(t_i, x)) \Delta W_{i+1} \right] \right) \right. \\
- f \left( t_i, x, R_i g_2(x), \frac{1}{\pi_{i+1}} \mathbb{E} \left[ g_2(X_{t_i+1}^1(t_i, x)) \Delta W_{i+1} \right] \right) \left\} \right.
\]

By the mean value theorem and the \( K \)-Lipschitz property of \( f \), we can find deterministic functions \( \nu(x) : \mathbb{R}^d \to \mathbb{R} \), \( \zeta(x) : \mathbb{R}^d \to \mathbb{R}^d \), which are bounded by \( K \), such that

\[
\pi_{i+1} \nu(x) (R_i g_1(x) - R_i g_2(x)) \\
+ \zeta(x) \cdot \mathbb{E} \left[ (g_1 - g_2)(X_{t_i+1}^1(t_i, x)) \Delta W_{i+1} \right]
\]

and we have that

\[
(1 - \pi_{i+1} \nu(x))(R_i g_1(x) - R_i g_2(x)) = \mathbb{E} \left[ (g_1 - g_2)(X_{t_i+1}^1(t_i, x)) \right] + \zeta(x) \cdot \mathbb{E} \left[ (g_1(X_{t_i+1}^1(t_i, x)) - g_2(X_{t_i+1}^1(t_i, x))) \Delta W_{i+1} \right]
\]

From (3.7) we deduce that

\[
(1 - \pi_{i+1} K) |(R_i g_1(x) - R_i g_2(x))| \\
\leq \left| \mathbb{E} \left[ (g_1 - g_2)(X_{t_i+1}^1(t_i, x)) (1 + \zeta(x) \cdot \Delta W_{i+1}) \right] \right| \\
\leq \|g_1 - g_2\|_{L^p(P_{t_i, t_i+1}^x)} \mathbb{E} \left[ (1 + \zeta(x) \cdot \Delta W_{i+1})^q \right]^\frac{1}{q} \\
\leq \|g_1 - g_2\|_{L^p(P_{t_i, t_i+1}^x)} \mathbb{E} \left[ (1 + \zeta(x) \cdot \Delta W_{i+1})^{2k} \right]^\frac{1}{2k}
\]

where \( p, q \) are conjugate and \( k \) an integer with \( k > q/2 \). Since \( \zeta(x) \) is deterministic and
bounded by the Lipschitz constant of \( f, K \), we have that
\[
\mathbb{E} \left[ (1 + \zeta(x) \cdot \Delta W_{i+1})^{2k} \right] = \sum_{j=0}^{2k} \binom{2k}{j} \mathbb{E} \left[ (\zeta(x) \cdot \Delta W_{i+1})^{2j} \right]
\]
\[
= \sum_{j=0}^{2k} \binom{2k}{j} \mathbb{E} \left[ (\zeta(x) \cdot \Delta W_{i+1})^{2j} \right]
\]
\[
\leq 1 + C_{\pi_{i+1}}
\]
which completes the proof of the lemma.

We need to clarify how the above result iterates on the family \( \{ R_i \} \), a point which will be used in the sequel. We apply Lemma 3.4 to \( |R_k R_{k+1} g_1 - R_k R_{k+1} g_2|(x) \) and to \( |R_{k+1} h(x) - R_{k+1} g(x)| \) to get
\[
|R_k R_{k+1} h(x) - R_k R_{k+1} (x)| \leq \frac{1 + C_{\pi_{k+1}}}{1 - K_{\pi_{k+1}}} \| R_{k+1} h - R_{k+1} g \|_{L^p(P_{x_0, t_{k+1}}, P_{x_1, t_{k+2}})}
\]
and
\[
|R_{k+1} h(x) - R_{k+1} g(x)| \leq \frac{1 + C_{\pi_{k+2}}}{1 - K_{\pi_{k+2}}} \| h - g \|_{L^p(P_{x_0, t_{k+1}}, P_{x_1, t_{k+2}})}.
\]
Then
\[
|R_k R_{k+1} h(x) - R_k R_{k+1} g(x)| \leq \left( \frac{1 + C_{\pi_{k+1}}}{1 - K_{\pi_{k+1}}} \right) \left( \frac{1 + C_{\pi_{k+2}}}{1 - K_{\pi_{k+2}}} \right) \| h - g \|_{L^p(P_{x_0, t_{k+1}}, P_{x_1, t_{k+2}})}
\]
by the semi group property. In general it holds that
\[
|R_0 \ldots R_i g(x) - R_0 \ldots R_i h(x)| \leq \prod_{j=1}^{i+1} \frac{1 + C_{\pi_j}}{1 - K_{\pi_j}} \| g - h \|_{L^p(P_{x_0, t_{k+1}}, P_{x_1, t_{k+2}})}
\]
for all \( i = 0, \ldots, n - 1 \).

As mentioned above, the conditional expectations involved in (3.2) are not computable in most cases. Hence, the next step would be to substitute the expectation operator \( \mathbb{E}[\cdot] \) in (3.4), with a simulation-based approximating operator \( \hat{\mathbb{E}}[\cdot] \) that is explicitly computable.
The simulation could rely on a Monte-Carlo method as is the case in [6] and [12], or on an evaluation on a tree as in [8]. Given, such an operator \( \hat{E}[\cdot] \) we define for \( i = n, \ldots, 0 \)

\[
\hat{R}_n g(x) = g(x) \\
\hat{R}_i g(x) = \hat{E} \left[ g(X_{t_{i+1}}^\pi (t_i, x)) \right] + \pi_{i+1} f(t_i, x, \hat{R}_i g(x), \frac{1}{\pi_{i+1}} \hat{E} \left[ g(X_{t_{i+1}}^\pi (t_i, x)) \Delta W_{i+1} \right])
\]

(3.8)

Given the above family of empirical operators, we define as approximating value of \( Y_{t_i} \)

\[
\hat{Y}_{t_i}^\pi = \hat{R}_i \hat{R}_{i+1} \ldots \hat{R}_{n-1} \Phi(X_{t_i}^\pi)
\]

We are now able to express the error between the Backward Euler scheme \( Y^\pi \) and the empirical value \( \hat{Y}^\pi \), as an error between the operators \( E[\cdot], \hat{E}[\cdot] \). Let

\[
R_{i,n-1} g := R_i \ldots R_{n-1} g \quad \hat{R}_{i,n-1} g := \hat{R}_i \ldots \hat{R}_{n-1} g
\]

Then the global error at each step \( i = 0, \ldots, n-1 \) is given by

\[
Y_{t_i}^\pi - \hat{Y}_{t_i}^\pi = R_{i,n-1} \Phi(x) - \hat{R}_{i,n-1} \Phi(x).
\]

(3.9)

We also set by convention \( \Delta W^0_{i+1} := 1 \), and we introduce the error

\[
\mathcal{V}[g; l](x) := (E - \hat{E}) \left[ g(X_{t_{i+1}}^\pi (t_i, x)) \Delta W_{i+1}^l \right]
\]

and denote the local \( p \)- errors at time \( t_i \) as

\[
\mathcal{E}_i^p[g; l] := \left\| \mathcal{V}[g; l](X_{t_i}^\pi) \right\|_p
\]

(3.10)

where \( i = 0, 1 \ldots, n-1, \ l = 0, 1 \ldots, d \) and \( p > 1 \).

**Theorem 3.5.** Assume that the partition is equidistant \( (t_i = 1/n) \) and that \( (H1a) \) and \( (H2) \) are satisfied. Then, for any \( p > 1 \) and \( i = 0, \ldots, n-1 \), the following holds

\[
| Y_{t_i}^\pi - \hat{Y}_{t_i}^\pi | \leq \frac{C}{|\pi|} \max_{0 \leq i \leq n-1} \mathcal{E}_i^p \left[ \hat{R}_{i,n} \Phi; l \right]
\]

for some constant \( C \) which is independent of the partition.
Proof. In this proof $C$ will be a constant which may vary from line to line. It is clearly sufficient to prove the result for $i = 0$.

We start from the definition of the regression and the empirical regression operators. For any function $g$, such that the corresponding expectation is well defined, we have that

$$R_i g(x) - \hat{R}_i g(x) = (\mathbb{E} - \hat{\mathbb{E}})[g(X_{t_{i+1}}^\pi(t_i, x))]
+ \pi_{i+1} \left\{ f(t_i, x, R_i g(x), \frac{1}{\pi_{i+1}} \mathbb{E} \left[g(X_{t_{i+1}}^\pi(t_i, x)) \Delta W_{i+1} \right] \right.
- f(t_i, x, R_i g(x), \frac{1}{\pi_{i+1}} \hat{\mathbb{E}} \left[g(X_{t_{i+1}}^\pi(t_i, x)) \Delta W_{i+1} \right]) \}
$$

Using the Lipschitz property of $f$ we may argue as in the proof of Lemma 3.4 to deduce that

$$|R_i g(x) - \hat{R}_i g(x)| \leq \frac{K}{1 - K\pi_{i+1}} \left\{ \left| (\mathbb{E} - \hat{\mathbb{E}}) \left[g(X_{t_{i+1}}^\pi(t_i, x))\right] \right| 
+ \left| (\mathbb{E} - \hat{\mathbb{E}}) \left[g(X_{t_{i+1}}^\pi(t_i, x)) \Delta W_{i+1} \right] \right| \right\} \tag{3.11}$$

We now go back to (3.9) and develop the error as a Trotter product, that is we add and subtract $n$ terms where every term is a combination of $n - i$ empirical backward projections and $i$ perfect ones $R_{0,i} \hat{R}_{i+1,n-1} \Phi(X_{t_0}^\pi)$.

$$|Y_{t_0}^\pi - \hat{Y}_{t_0}^\pi| = |R_{0,n} \Phi(X_{t_0}^\pi) - \hat{R}_{0,n} \Phi(X_{t_0}^\pi)| 
\leq \sum_{i=0}^{n-1} |R_{0,i-1} \hat{R}_{i,n-1} \Phi(X_{t_0}^\pi) - R_{0,i} \hat{R}_{i+1,n} \Phi(X_{t_0}^\pi)| \tag{3.12}$$

We may apply the Remark following lemma 3.6, to every term in the sum to deduce that

$$|R_{0,i} \hat{R}_{i+1,n-1} \Phi(x) - R_{0,i-1} \hat{R}_{i,n-1} \Phi(x)| 
\leq \prod_{j=1}^{i} \left( \frac{1 + C\pi_{j+1}}{1 - K\pi_{j+1}} \right) \left\| (R_j - \hat{R}_j) \hat{R}_{i+1,n} \Phi(X_{t_{i+1}}^\pi(0, x)) \right\|_p \tag{3.13}$$

Using (3.11) we conclude that

$$\left\| (R_i - \hat{R}_i) \hat{R}_{i+1} \ldots \hat{R}_{n-1} \Phi(X_{t_i}^\pi(0, x)) \right\|_p 
\leq \frac{K(d + 1)}{1 - K\pi_{i+1}} \max_{0 \leq l \leq d} \mathcal{E}_l^{\pi} \left[ \hat{R}_{i,n} \Phi; l \right]$$
Plugging these estimates in (3.12) we have that
\[ |Y_{\pi_{0}} - \hat{Y}_{\pi_{0}}| \leq \frac{K(d + 1)}{1 - K\pi_{i+1}} \sum_{i=0}^{n-1} \prod_{j=1}^{i} \left( \frac{1 + C_{\pi_{j+1}}}{1 - K\pi_{j+1}} \right) \max_{0 \leq l \leq d} \mathcal{E}_{i} \left[ \hat{R}_{i,n}\Phi; l \right] \]
\[ \leq \frac{K(d + 1)}{1 - K\pi_{i+1}} \max_{0 \leq l \leq d} \mathcal{E}_{i} \left[ \hat{R}_{i,n}\Phi; l \right] \sum_{i=0}^{n-1} \prod_{j=1}^{i} \left( \frac{1 + C_{\pi_{j+1}}}{1 - K\pi_{j+1}} \right) \]
and the proof completes by noting that
\[ \sum_{i=0}^{n-1} \prod_{j=1}^{i} \left( \frac{1 + C_{\pi_{j+1}}}{1 - K\pi_{j+1}} \right) \leq e^{C + K}. \]

4 The Malliavin calculus approach

We will now depart from the abstract level of the previous section to analyze the method presented in [6]. Throughout this section the stronger condition (H1b) is assumed to hold. Also, the partition is assumed to be equidistant so that \( \pi_{i} = |\pi|, \forall i \). The choice for the simulation-based approximation \( \hat{\mathbb{E}}[\cdot] \) of \( \mathbb{E}[\cdot] \) is based on a representation of the conditional expectation obtained by means of the (Malliavin) integration by parts formula. The exact formula of this representation involves some weights, that are in fact iterated Skorohod integrals. So, we first introduce the related notation.

We denote by \( J_{k} \) the subset of \( \mathbb{N}^{k} \) with elements \( I = (i_{1}, \ldots, i_{k}) \) that satisfy \( 1 \leq i_{1} < \ldots < i_{k} \leq d \). Trivially, we set \( J_{0} = \emptyset \). Given two elements \( I \in J_{k}, J \in J_{q} \) we define their concatenation \( I * J : = (r_{1}, \ldots, r_{l}) \) with \( k \land q \leq l \leq d \land (k + q) \), \( r_{i} \in I \cup J \) for every \( i \leq l \), and \( 1 \leq r_{1} < \ldots < r_{p} \leq d \). Finally, for any \( J \in J_{k}, k = 0,1,\ldots,d \) we write \( J^{c} \) for its complementary set, that is, the unique set such that \( J * J^{c} = \{1,2,\ldots,d\} \).

Given a matrix valued process \( h \) with columns denoted by \( h^{i} \) and a random variable \( F \) we denote by
\[ S_{I}^{h}[F] : = \int_{0}^{\infty} F(h_{t}^{i})^{*} dW_{t} \]
and for a multi index \( I = (i_{1}, \ldots, i_{k}) \)
\[ S_{I}^{h}[F] : = S_{i_{1}}^{h} \circ \ldots \circ S_{i_{k}}^{h}[F] \]
where the integrals are understood in the Skorohod sense. We extend the definition to $I = \emptyset$ by setting $\mathcal{S}_0^h[F] = F$. Let $\phi$ be a bounded and continuous real valued function with $\phi(0) = 1$. We will say that $\phi$ is a smooth localizing function if

$$\partial_I \phi(x) \in C_0^0(\mathbb{R}^d), \quad \text{for any } k = 0, 1, \ldots, d, \quad I \in \mathcal{J}_k$$

trivially we set $\partial_{\emptyset} \phi = \phi$. We denote the collection of all smooth localizing functions by $\mathcal{L}$. For integrability reasons, we will follow [6] by requiring the localizing function to satisfy the condition:

$$\sum_{I \in \bigcup_{k=1}^d \mathcal{J}_k} \int_{\mathbb{R}^d} |x|^m \partial_I \phi(x)^2 dx < \infty. \quad (4.1)$$

By $\mathcal{L}_0$, we denote the collection of all localizing functions satisfying (4.1). For $i = 0, 1, \ldots, n$, we consider the matrix valued process

$$h_i(t) = \frac{1}{|\pi|} \sigma^{-1}(X_{\pi t_{i-1}}^\pi) 1_{(t_{i-1}, t_i)}(t) - \frac{1}{|\pi|} \sigma^{-1}(X_{\pi t_{i-1}}^\pi) \left(I_d + |\pi| \nabla b(X_{\pi t_{i-1}}^\pi) + \sum_{j=1}^d \nabla \sigma^j(X_{\pi t_{i-1}}^\pi) \Delta W_{i+1}^j \right) 1_{(t_{i-1}, t_i)}(t)$$

Under (H1b), the Euler scheme is infinitely many times differentiable in the Malliavin sense, and the derivative is given by

$$D_t X_{\pi t}^\pi = \sigma(X_{\pi t}^\pi) 1_{t \leq t_1},$$

$$D_t X_{\pi t}^\pi = D_t X_{\pi t_{i-1}}^\pi + |\pi| \nabla b(X_{\pi t_{i-1}}^\pi) D_t X_{\pi t_{i-1}}^\pi$$

$$+ \sum_{j=1}^d \nabla \sigma^j(X_{\pi t_{i-1}}^\pi) D_t X_{\pi t_{i-1}}^\pi \Delta W_{i+1}^j + \sigma(X_{\pi t_{i-1}}^\pi) 1_{(t_{i-1}, t_i)}(t)$$

Observe that the process $h_i(t)$ satisfies the identities

$$\int_0^1 D_t X_{\pi t_i}^\pi h_i(t) dt = I_d, \quad \int_0^1 D_t X_{\pi t_{i+1}}^\pi h_i(t) dt = 0 \quad (4.2)$$

Also, for all $I \in \bigcup_{k=0}^d \mathcal{J}_k$, $l = 0, \ldots, d$, and $\phi \in \mathcal{L}_0$, the iterated integrals

$$S_I^{h_i} \left[ \Delta W_{i+1}^l \phi(X_{\pi t_i}^\pi - x) \right] = \sum_{J \subseteq I} (-1)^{|J|} \partial_J \phi(X_{\pi t_i}^\pi - x) S_J^{h_i} [a(\Delta W_{i+1})] \quad (4.3)$$
is well defined, and belong to $D^{1,2}$, the space of one time Malliavin differentiable random variables. The latter equality follows from the first property of (4.2).

The following representation is reported from [4] and [6].

**Theorem 4.1.** Let $\rho$ is a real valued function, $0 \leq l \leq d$, and assume that the random variable $F = \rho(X_{t_{i+1}}^\pi) \Delta W_{l+1}^{i} \in L^2$. Then for all localizing functions $\phi, \phi_F \in L_0$:

$$
\mathbb{E} [F|X_{t_i}^\pi = x] = \frac{\mathbb{E} \left[ Q_F[h_i, \phi^F](x) \right]}{\mathbb{E} \left[ Q^1[h_i, \phi](x) \right]},
$$

where

$$
Q_F[h_i, \phi](x) := H_x(X_{t_i}^\pi) \rho(X_{t_{i+1}}^\pi) S_{(i, \ldots, d)}^{h_i} [\Delta W_{l+1}^{i} \phi_R(X_{t_i}^\pi - x)],
$$

and $H_x(y) := \prod_{i=1}^d 1_{x_i \leq y_i}$ is the Heaviside function in $d$-dimensions.

Observe that, $q_i(x) = \mathbb{E} [Q^1[h_i, \phi](x)]$ is the density of $X_{t_i}^\pi$. The above representation paves the way for a Monte Carlo approach to the problem of conditioning. In particular it suggests to substitute the regression function $\mathbb{E} [F|X_{t_i}^\pi = x]$ with a ratio of empirical means. However we should note that in the simulation, there could be integrability issues. If we consider $N$ independent random variables $\{B_k(x)\}_{k=1}^N$ with $\mathbb{E}[B_k(x)] = q_i(x)$, $\forall k$ then the estimator of the denominator $\hat{q}_i(x) = N^{-1} \sum_{k=1}^N B_k(x)$ has a gaussian asymptotic distribution. For this reason, in [6] (Lemma 3.3) the authors take advantage of preliminary bounds available for $Y_{t_i}^\pi$, $\mathbb{E}_{i-1}[Y_{t_i}^\pi \Delta W_{l+1}^{i}]$. In particular, there exists functions $\Psi_i(x)$, $\overline{\Psi}_i(x)$, $\zeta_i(x)$, $\overline{\zeta}_i(x)$ of polynomial growth, such that

$$
\Psi_i(X_{t_i}^\pi) \leq Y_{t_i}^\pi \leq \overline{\Psi}_i(X_{t_i}^\pi)
$$

$$
\zeta_{i-1}(X_{t_{i-1}}^\pi) \leq \mathbb{E}_{i-1}[Y_{t_i}^\pi \Delta W_{l}^{i}] \leq \overline{\zeta}_{i-1}(X_{t_{i-1}}^\pi)
$$

for $i = 0, \ldots, n$ and $l = 0, \ldots, d$. With the above bounds at hand let us now build the backward algorithm. Let us denote by $({\Omega}^0, {\mathcal{F}}^0, {\mathbb{P}}^0)$ the original probability space. Consider $nN$ copies of the Euler scheme approximation of the forward diffusion, that is, will be making use of $N$ copies at every point $t_k$ on the grid. We define $n$ probability spaces, $\{(\Omega^i, {\mathcal{F}}^i, {\mathbb{P}}^i)\}_{i=1}^n$, where in each of these spaces $N$ copies live, i.e. $\left\{X_{j}^\pi\right\}_{j \in {\mathcal{N}}^i}$, $N_i =$
\{N_{i+1}, \ldots, N(i+1)\}, live on the space \(\Omega^i\). Further on, we consider the product probability space

\[
(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P}) = (\otimes_{i=0}^{n} \Omega^i, \otimes_{i=0}^{n} \mathcal{F}^i, \otimes_{i=0}^{n} \mathbb{P}^i) \tag{4.6}
\]

and we lift all processes in this space. Integration with respect to the measure \(\bar{P}\) is denoted by \(\bar{E}\), while integration with marginal measures, \(\mathbb{P}^i, \mathbb{P}^i \times \mathbb{P}^k\), is denoted by \(E^i, E^{i,k}\) etc.

The operator \(\hat{E}\) suggested from Theorem 4.4 is defined as follows for \(i = n - 1, \ldots, 1\):

\[
\hat{E}_i \left[ g(X_{t_{i+1}}^\pi (t_i, x)) \Delta W_{i+1}^l \right] = T^\xi_1 \left[ \hat{E}_i \left[ g(X_{t_{i+1}}^\pi (t_i, x)) \Delta W_{i+1}^l \right] \right] \tag{4.7}
\]

with \(T^\xi_i(x) = \zeta_i(X^\pi_{t_i}) \wedge x \vee \zeta_i(X^\pi_{t_i})\) for any \(x \in \mathbb{R}^d\) and

\[
\hat{E}_i \left[ g(X_{t_{i+1}}^\pi (t_i, x)) \Delta W_{i+1}^l \right] = \frac{\hat{Q}^F[h_i, \phi_F](x)}{\hat{Q}^1[h_i, \phi_1](x)},
\]

for possibility different localizing functions \(\phi_1\) and \(\phi_F \in L_0\), and

\[
\hat{Q}^F[h_i, \phi_F](x) = \frac{1}{N} \sum_{j \in N_{i+1}} H_x(X^\pi_{t_i}) g(X^\pi_{t_{i+1}}, S^h_{l_i} \left[ \phi_R \left( X^\pi_{t_i} - x \right) \Delta W_{i+1}^{l,k} \right])
\]

\[
\hat{Q}^1[h_i, \phi_1](x) = \frac{1}{N} \sum_{l \in N_{i+1}} H_x(X^\pi_{t_i}) S^h_{l_i} \left[ \phi_1 \left( X^\pi_{t_i} - x \right) \right].
\]

Here, \(\Delta W_{i+1}^{l,0} = 1\) and for \(1 \leq k \leq d\), \(\Delta W_{i+1}^{l,k}\) is the \(k\)-th component of the \(l\)-th copy in the \(N_{i+1}\) bundle of copies of the increment \(\Delta W_{i+1}\). Moreover, \(h^i_l\) stands for the evaluation of the function \(h_i\) using the \(l\)-th copy of the Euler scheme. At time \(t_0\), \(\bar{E}\) is just the empirical mean

\[
\hat{E}_0 \left[ g(X_{t_1}(0, x)) \Delta W_{1}^l \right] = \frac{1}{N} \sum_{j \in N_1} g(X^\pi_{t_0}) a(\Delta W_{1}^{j,l})
\]

The bounds in the values of \(Y^\pi\) are used in the definition of the operators \(\hat{R}_i, i = 0, \ldots, n-1\).
1. In the current set up (3.8) becomes
\[\tilde{R}_t g(x) = g(x)\]
\[\tilde{R}_r g(x) = \tilde{E} \left[ g(X_{i+1}^\pi(t, x)) \right] + |\pi| f \left( t, x, \tilde{R}_r g(x), |\pi|^{-1} \tilde{E} \left[ g(X_{i+1}^\pi(t, x)) \Delta W_{i+1} \right] \right)\]
\[\tilde{R}_r g(x) = T_i^\Psi (\tilde{R}_r g(x)) \quad i = 0, 1, \ldots, n\]
where \(T_i^\Psi(F) = \Psi_i \vee F \wedge \Psi_i\).

**Remark 4.2.** The conclusions of theorem 4.1 remain valid when applied to a random function \(\varrho\). In particular, if \(\varrho := \varrho(x, \xi)\) where \(\xi\) is independent of \(\mathcal{F}^\pi\) the representation (4.4) is true, only now both sides involve random functions instead of deterministic ones. A careful inspection of the algorithm shows that (4.4) needs to be applied to such functions.

The key point is Theorem 5.1 in [6]. This gives the description for the local error.

**Theorem 4.3.** For \(0 \leq i \leq n - 1\) and \(0 \leq l \leq d\), consider the random variable \(F = \varrho(X_{i+1}^\pi, \omega) \Delta W_{i+1}\), where \(\varrho(\cdot, \omega)\) is a random function with \(\omega \in \Omega^{i+1} \times \ldots \Omega^n\).

For \(\phi_F \in L_0\), let \(q_i^F(x) := \mathbb{E}^0 \left[ Q^F[h_i, \phi_F](x) \right]\) be defined as in Theorem 4.1 so that \(r_i(x) := q_i^F(x)/q_1^F(x) = \mathbb{E}^{i+1, \ldots, n}[F|X_i^\pi = x]\), and consider the regression estimator
\[\hat{r}_i(x) := T_i^\gamma (\tilde{r}_i(x)), \quad \text{with} \quad \hat{r}_i(x) := \frac{\hat{Q}_i^F[h_i, \phi](x)}{\hat{Q}_1^F[h_i, \psi](x)}\].

Then, for any \(p > 1\), we have the error estimate:
\[\|r_i(X_i^\pi) - \hat{r}_i(X_i^\pi)\|_p \leq \frac{1}{N^{1/2p}} (2\Gamma(F))^{1/p}\] (4.9)
where
\[\Gamma(F) := \int_{\mathbb{R}^d} \gamma(x) \left[ ||Q^F[h_i, \phi](x)||_{L^2(\mathbb{P})} + (|r(x)| + \gamma(x))||Q^1[h_i, \psi](x)||_{L^2(\mathbb{P})} \right] dx\]
and
\[\gamma(x) := |\tilde{\Psi}_i(x) - \Psi_i(x)| \vee |\tilde{\zeta}_i(x) - \zeta_i(x)|\].
For the sake of completeness, we include the proof of Theorem 4.3 in the Appendix. In particular, the factorisation (4.6) of the probability space \((\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})\) clarifies some points in the proof of [6].

The above result provides us with a control on the error of the simulation of the conditional expectation. The error is controlled by means of the parameter \(N\), which is the number of simulated paths used at time \(t_i\) for the Monte Carlo method. However, there is a further negative influence of the partition mesh \(|\pi|\) hidden in the functional \(\Gamma\). In particular, the second norm of the involved random variables explodes when the partition mesh shrinks. The exact rate of explosion is given in the following lemma, reported from [6].

**Lemma 4.4.** Let \(\mu : \mathbb{R}^d \to \mathbb{R}\) be a function of polynomial growth
\[
\sup_{x \in \mathbb{R}^d} \frac{|\mu(x)|}{1 + |x|^m} < \infty \quad \text{for some } m \geq 1
\]
For \(\phi \in \mathcal{L}_0\), set \(\phi^\pi(x) := \phi(|\pi|^{-1/2}x)\) and let \(F\) be defined as in Theorem 4.3. Assume further that \(F \in L^{2+\epsilon}\) for some \(\epsilon > 0\). Then
\[
\max_{1 \leq i \leq n} \int_{\mathbb{R}^d} \mu(x) \|Q^F[h_i, \phi^\pi](x)\|_{L^2(\bar{P})} dx \leq \frac{C}{|\pi|^{d/4}}.
\]

We are now able to conclude the estimate for the global error of the algorithm.

**Theorem 4.5.** Assume that (H1b) and (H2) hold true. Then for the Malliavin calculus algorithm the following error estimate holds
\[
|Y_{t_0}^\pi - \hat{Y}_{t_0}^\pi| \leq \frac{C}{|\pi|^{1+d/4} N^{1/2p}}.
\]

**Proof.** An easy induction argument shows that \(V := \hat{R}_{i,n-1}g(x) = g(x, \omega)\) for \(\omega \in \Omega^{i+2} \times \ldots \Omega^n\), and some appropriate random function \(g\). Hence we fall into the context of theorem
4.3 and this result gives us for every $0 \leq l \leq d$:

$$\| (E - \hat{E}) \left[ \varphi(X_{i+1}(t_i, x), \omega) \Delta W_{i+1}^l \right] \|_p \leq$$

$$\frac{1}{N^{1/2p}} \left( \int_{\mathbb{R}^d} \gamma(x) \left[ \| Q^V[h_i, \phi^\pi](x) \| _{L^2(p)} \right.$$

$$+ (|r(x)| + \gamma(x)) \| Q^1[h_i, \psi^\pi](x) \| _{L^2(p)} \right) dx \right)^{1/p}$$

The above combined with theorem 3.5 gives

$$|Y_{t_0}^\pi - \hat{Y}_{t_0}^\pi| \leq \frac{C}{|\pi|} \frac{1}{N^{1/4}} \left( \int_{\mathbb{R}^d} \gamma(x) \left[ \| Q^V[h_i, \phi^\pi](x) \| _{L^2(p)} \right.$$

$$+ (|r(x)| + \gamma(x)) \| Q^1[h_i, \psi^\pi](x) \| _{L^2(p)} \right) dx \right)^{1/p}$$

The functions $\gamma(x), r(x)$ are of linear growth (see [6]) so lemma 4.4 finishes the proof. □

5 An improvement on the Malliavin weights

The Malliavin calculus algorithm provides us with an efficient method for the numerical solution of a BSDE. However one serious drawback of this method is that it can become quite heavy from the implementation viewpoint when the dimension is high. Our aim in this section is to present a variation of the algorithm presented in [6] that reduces the computational effort. We will show that one can consider the function

$$\tilde{h}_i(t) = \frac{1}{|\pi|} \left( \sigma^{-1}(X_{t_{i-1}}^\pi) 1_{[t_{i-1}, t_i)}(t) - \sigma^{-1}(X_{t_i}^\pi) 1_{[t_i, t_{i+1})}(t) \right) \tag{5.1}$$

in place of $h_i$ and in effect $S^{h_i}$ in place of $S^{h_i}$, and form a backward induction scheme based on these new weights. Let us try to appreciate the gain from this truncation.

To simplify things, in the definition of $h$, we treat $\frac{1}{|\pi|} \sigma^{-1}(X_{t_i}^\pi) \sum_{j=1}^d \nabla \sigma^j(X_{t_i}^\pi) \Delta W_{i+1}^j$ as being one term (when in fact there are $d$). Under this assumption, we may deduce that the iterated Skorohod integral $S^{h_i}[\varphi(X_{t_i}^\pi - x) \Delta W_{i+1}^l]$ constitutes of $4d$ terms. The same integral with respect to $\tilde{h}$ requires the computation of $2d$ terms. It follows that for this new
algorithm, we have to put in only $(1/2)^d$ of the effort required for the original algorithm, when computing the weights. For example, in five dimensions, we make approximately 4% of the original effort. To add to this, observe that the above assumption is quite moderate. The following should also be considered:

- By excluding the last part of $h$ from our computations we are able to avoid $d + 1$ terms for every integral instead of two.
- Avoiding the last part of $h$, we avoid the computation of the derivatives of the matrix valued function $\sigma(\cdot)$.
- Lastly, the last part of $h$ is also the anticipating part of it. Hence, it is the one contributing most to the Malliavin correction terms in the Skorohod integrals, which are now excluded from the computations.

The price that has to be paid for the reduction in the computational effort, is that more copies of the Euler scheme need to be used, to obtain error estimates equivalent to the previous section. This will be made precise in Theorem 5.6.

**Remark 5.1.** The truncation of the $h$ process does not affect the denominators in the representation of Theorem 4.1. Indeed, a repetitive application of Lemma 3.2.1 of [16] shows that

\[
q^1(x) = \mathbb{E} \left[ H_x(X_{t_i}^\pi) \mathbb{E}_i \left[ S^{\bar{h}_i} \left[ \phi(X_{t_i}^\pi - x) \right] \right] \right]
= \mathbb{E} \left[ H_x(X_{t_i}^\pi) \mathbb{E}_i \left[ S^{\bar{h}_i} \left[ \phi(X_{t_i}^\pi - x) \right] \right] \right]
\frac{1}{|\pi|} (-\sigma(X_{t_i-1}^\pi))^{\dagger} \cdot \Delta W_i
\]

\[
= \mathbb{E} \left[ H_x(X_{t_i}^\pi) S_t^{\bar{h}_i} \left[ \mathbb{E}_i \left[ S^{\bar{h}_i} \left[ \phi(X_{t_i}^\pi - x) \right] \right] \right] \right]
= \ldots = \mathbb{E} \left[ H_x(X_{t_i}^\pi) S_t^{\bar{h}_i} \left[ \phi(X_{t_i}^\pi - x) \right] \right]
\]

In other words, in the denominator the truncation is exact and no precision is lost. Given this, our effort will focus on the truncation of the numerator.
Observing that \( \int D_t X_t \tilde{h}_i(t) dt = \tilde{I}_d \), we see that expansion (4.3) holds for \( \tilde{h}_i \) as well:

\[
S^\tilde{h}_i \left[ \Delta W_{i+1}^l \varphi (X_{t_i}^\pi - x) \right] = \sum_{J \subseteq I} (-1)^{|J|} \partial_J \varphi (X_{t_i}^\pi - x) S_{\tilde{h}_i} \left[ \Delta W_{i+1}^l \right] (5.3)
\]

for every \( 0 \leq l \leq d \). Hence, in order to analyze the error due to the truncation of \( h_i \), we will concentrate on the difference \( S^h_{\tilde{h}_i} \left[ \Delta W_{i+1}^l \right] - S^h_{\tilde{h}_i} \left[ \Delta W_{i+1}^l \right] \) for every \( J \subseteq I \) and \( 0 \leq l \leq d \).

In the rest of this section, we analyse the error due to the truncation in order to better understand the price to be paid for the above gain in complexity. For this purpose we need to redefine the regression function according to the new “truncated” weights. In the same spirit with Theorem 4.1 we introduce an approximating expectation operator \( \bar{E}[] \) as follows: For \( R = \varrho(X_{t_i}^\pi) \Delta W_{i+1}^l \), \( 1 \leq i \leq d \) and \( 0 \leq l \leq d \), we write

\[
\bar{q}^R(x) := E \left[ Q^R [\bar{h}_i, \phi^\pi_R](x) \right] , \quad \bar{q}^1(x) := E \left[ Q^1 [\bar{h}_i, \phi^\pi_1](x) \right] = q^1(x)
\]

where given any \( \phi_1, \phi_R \in L_0 \), we denote \( \phi^\pi(x) := \phi(x/|\pi|^\alpha) \) and set

\[
\bar{E} \left[ \varrho(X_{t_i+1}^\pi, t_i, x)) \Delta W_{i+1}^l \right] = \begin{cases} T^{\bar{q}^R}(\bar{q}^R(x)/\bar{q}^1(x)), & a(x) = x_l, l = 1, \ldots, d \\ T^{\bar{q}^1}(\bar{q}^R(x)/\bar{q}^1(x)), & a(x) = 1 
\end{cases} (5.4)
\]

Observe that, here we have introduced a normalization in the localizing function, dividing by \( |\pi|^\alpha \). In the previous section we had \( \alpha = 1/2 \), whereas here we need to consider \( \alpha > 1/2 \). This is necessary for the control of the error and it is exactly what will lead us to worse variance estimates in comparison to the previous section, or in other words, it will lead us to require the generation of more copies of the forward Euler scheme (larger \( N \)) to obtain a rate of convergence of the same order with the previous section.

The next step is, using this new regression function \( \bar{E}[] \) to redefine the operators \( R_i \) of (3.4) as

\[
\tilde{R}_i g(x) = \bar{E} \left[ g(X_{t_i+1}^\pi, t_i, x)) \right] + |\pi| \int f \left( t_i, x, \tilde{R}_i g(x), |\pi|^{-1} \bar{E} \left[ g(X_{t_i+1}^\pi, t_i, x)) \Delta W_{i+1} \right] \right) \\
\tilde{R}_i g(x) = T^{\bar{q}^R}(\tilde{R}_i g(x)) (5.5)
\]
for $i = n - 1, \ldots, 0$ and any function $g$ in a suitable class. The iteration of the above family will give us the values

$$\bar{Y}^n_{t_i} := \bar{R}_i \ldots \bar{R}_{n-1} \Phi(x)$$

for $i = n - 1, \ldots, 0$ where $\Phi(\cdot)$ is the function of the terminal condition of the BSDE.

Finally we introduce the family of operators $\{\hat{\bar{R}}_i\}_{i=1}^n$ that corresponds to the simulation-based estimation of the family $\{\bar{R}_i\}_{i=1}^n$ in a Monte Carlo fashion similar to the previous section. That is, the definition for the family $\{\hat{\bar{R}}_i\}_{i=1}^n$ is similar to (4.8), with the difference that we use $\hat{h}_i(t)$ in place of $h_i(t)$ for every step $i$.

To produce meaningful error estimates, we first need to look at the $L^p$ norms of the iterated Skorohod integrals. We first introduce some additional notations. Let

$$\tilde{S}_{\gamma}^h_{I,\gamma} [\Delta W_{i+1}] := \delta(F\hat{h}_i)$$

and

$$\tilde{S}_{\gamma}^{\bar{h}}_{I,\gamma} [\Delta W_{i+1}] := \delta(F\bar{h}_i).$$

Then, by an easy induction argument, we see that for any $I_m \subseteq I$:

$$S_{I_m}^{\bar{h}} [\Delta W_{i+1}] - S_{I_m}^h [\Delta W_{i+1}] = \sum_{\gamma \in \{\pm 1\} \cap \{-1, -1, \ldots, -1\}} S_{I_m,\gamma} [\Delta W_{i+1}]. \quad (5.6)$$

When these integrals are of length one, the following estimate follows directly from the properties of Skorohod integration:

$$S_{I_i}^{h_i} [\Delta W_{i+1}] = \Delta W_{i+1} \bar{h}_i(t_{i-1}) \Delta W_i + \Delta W_{i+1} \tilde{h}_i(t_i) \Delta W_{i+1} - \int_{t_i}^{t_{i+1}} \text{Tr}[D_s(\Delta W_{i+1}) \tilde{h}_i(t_i)] ds,$$
which implies that
\[ \| S_{j}^{\hat{h}_{s_{0}}} \|_p \leq \frac{C}{\sqrt{|\pi|}}, \quad \| D_{s_{1}, \ldots, s_{\rho}} S_{j}^{\hat{h}_{s_{0}}} \|_p \leq \frac{C}{|\pi|}. \]

We also note that,
\[ S_{j}^{\hat{h}_{s_{0}}} \Delta W_{l+1}^{l} = \Delta W_{l+1}^{l} \hat{h}_{j}(t_{l-1}) \Delta W_{l} + \Delta W_{l+1}^{l} \hat{h}_{j}(t_{l}) \Delta W_{l+1} - \int_{t_{l}}^{t_{l+1}} \text{Tr}[D_{s}(\Delta W_{l+1}^{l}) \hat{h}_{j}(t_{l})] ds, \]
which implies that
\[ \| S_{j}^{\hat{h}_{s_{0}}} \|_p \leq C, \quad \| D_{s_{1}, \ldots, s_{\rho}} S_{j}^{\hat{h}_{s_{0}}} \|_p \leq \frac{C}{\sqrt{|\pi|}}, \]
but
\[ \| D_{s_{1}, \ldots, s_{\rho}} S_{j}^{\hat{h}_{s_{0}}} \|_p \leq \frac{C}{|\pi|}. \]

For the general case, we have the following estimate on the explosion rate. The proof is in the spirit of Lemma 6.1 of [6].

**Lemma 5.2.** For \( 1 \leq k, m \leq d \), consider a vector of times \( \tau = (s_{1}, s_{2}, \ldots, s_{k}) \in [t_{i-1}, t_{i+1}]^{k} \), a multi index \( I_{m} \in J_{m} \). Then for every \( \gamma \in \{ \pm 1 \}^{m} \setminus \{-1, -1, \ldots, -1\} \):
\[ \| D_{\tau} S_{I_{m}, \gamma} \|_p \leq C|\pi|^{(1-k-m)/2}, \quad 0 \leq l \leq d. \tag{5.7} \]

Moreover
\[ \| S_{j}^{\hat{h}_{s_{0}}} \|_p \leq \frac{C}{|\pi|^{(d-1)/2}} \]
and, for \( p > 1 \),
\[ \| S_{j}^{\hat{h}_{s_{0}}} \|_p, \quad \| S_{j}^{\hat{h}_{s_{0}}} \|_p = O(|\pi|^{-d/2}). \]

**Proof.** We only prove (5.7), as the remaining estimates follows the same line of argument as in [6]. For simplicity we write \( S_{I, \gamma} \Delta W_{l+1}^{l} = S_{I, \gamma} \). The proof is done with induction on the length of \( I_{m} \). The case \( m = 1 \) is discussed above. Assume that it holds for \( k - 1 \).
Consider $I_k = (j_1, \ldots, j_k)$ and $\gamma \in \{\pm 1\}^k$ and we assume for simplicity that $\gamma_1 = -1$. Then
\[
\|D_\tau S_{I_k, \gamma}\|_p = \left\| D_\tau \left\{ S_{I_{k-1}, \gamma_1} - \int_{t_{i-1}}^{t_{i+1}} D_s S_{I_{k-1}, \gamma_1} \bar{h}^{j_1}(s) \, ds \right\} \right\|_p 
\leq \sum_{\Lambda \subseteq \{t_1, \ldots, t_l\}} \|D_\Lambda S_{I_{k-1}, \gamma}\|_{2p} \|D_\Lambda \delta(\bar{h}^{j_1})\|_{2p} 
+ \int_{t_{i-1}}^{t_{i+1}} \|D_\Lambda s S_{I_{k-1}, \gamma_1} - \bar{h}^{j_1}(s)\|_{2p} \|D_\Lambda \bar{h}^{j_1}(s)\|_{2p} \, ds,
\]
where we have used the elementary rule $\delta(F u_t) = F \delta(u_t) - \int_0^1 D_t F u_t \, dt$ and Exercise 1.2.14 pg 34 of [16]. The term involving the Riemann integral is
\[
\int_{t_{i-1}}^{t_{i+1}} \|D_\Lambda s S_{I_{k-1}, \gamma_1} - \bar{h}^{j_1}(s)\|_{2p} \|D_\Lambda \bar{h}^{j_1}(s)\|_{2p} \, ds 
\leq 2C|\pi|^{\left(1-|\Lambda|^{-1-k+1}\right)/\pi-1} = |\pi|^{\left(1-|\Lambda|-k\right)/2}.
\]
As for the other terms, if $\Lambda = \tau$ by the induction hypothesis and the remark
\[
\|D_\tau S_{I_{k-1}, \gamma_1}\|_{2p} \|\delta(\bar{h}^{j_1})\|_{2p} \leq C|\pi|^{\left(1-|\Lambda|-k\right)/\pi-1/2} = C|\pi|^{\left(1-|\Lambda|-k\right)/2},
\]
and if $\Lambda \subset \tau$
\[
\|D_\Lambda s S_{I_{k-1}, \gamma_1}\|_{2p} \|D_\Lambda \delta(\bar{h}^{j_1})\|_{2p} \leq C|\pi|^{\left(1-|\Lambda|-k\right)/\pi-1} 
= |\pi|^{-\left(1-|\Lambda|-k\right)/2} 
\leq |\pi|^{\left(1-|\Lambda|\right)/2}
\]
as $|\Lambda| \leq l - 1$. Putting all the estimations together we have the result.

**Remark 5.3.** As in the previous section, it is an easy induction argument to show that for any $i = 0, 1, \ldots, n-1$, the iteration of the family $\{\hat{R}_i\}_{i=1}^n$ produces random functions that may be written as $\hat{R}_i \ldots \hat{R}_n \Phi(x, \chi)$ with $\chi$ a random variable. Observe that, with the notation of the previous section, $\chi$ can be shown to be a functional of the copies of the Euler scheme $\{X^\pi_l, l = i + 1, \ldots, n\}$.
To obtain a rate of convergence for this variation of the algorithm presented in [6], we need to compare $Y_{t_0} = R_0 \ldots R_n \Phi(x)$ with $\hat{Y}_{t_0} = \hat{R}_0 \ldots \hat{R}_n \Phi(x)$ when $X_0 = x$. The first step is to look at the local error between $R_i$, $\bar{R}_i$, $i = 0, \ldots, n - 1$, when applied to random functions of the form described in Remark 5.3.

**Theorem 5.4.** Given a localizing function $\phi \in L_0$, let $\phi^\pi(x) := \phi(|\pi|^{-\alpha} x)$, and consider the definition of the regression (4.4) and the truncated regression functions (5.4) with localising function $\phi^\pi$. Then, there exists a constant $C$ independent of the partition such that:

1. If $\varrho$ is a Lipschitz real valued function, then for any $i = 0, \ldots, n - 1$ and $p > 1$:
   \[
   \|R_i \varrho(X^\pi_{t_i}) - \bar{R}_i \varrho(X^\pi_{t_i})\|_p \leq C|\pi|^\frac{1}{2}(\alpha + \frac{1}{2}).
   \]

2. Let $\chi$ be a random variable independent of $\mathcal{F}^\pi_{t_i}$ and assume that $\varrho(X^\pi_{t_i+1}, \chi) \Delta W_{t_i+1}^l \in L^q$, $q > 1$, with $\varrho$ a real valued random function and $l = 0, 1, \ldots, d$. Then
   \[
   \|R_i \varrho(X^\pi_{t_i}) - \bar{R}_i \varrho(X^\pi_{t_i})\|_p \leq C|\pi|^\frac{1}{2} \|\varrho(X^\pi_{t_i+1}, \chi)\|_{1/p}^{1/p}.
   \]

**Proof.** Let us first assume that we are working with a Lipschitz deterministic function $\varrho$. Using the notation and arguments presented in Theorem 3.5, where here we consider $\tilde{\mathbb{E}} \equiv \mathbb{E}$, we know that there exists a constant $C$ such that

\[
\|(R_i - \bar{R}_i) \varrho(X^\pi_{t_i})\|_p \leq \frac{C}{1 - K|\pi|} \max_{0 \leq l \leq d} \mathcal{E}_p^p [\varrho; l].
\]

Hence we may focus on the right hand side and for that, we fix a value for $l \in \{0, \ldots, d\}$. Let

\[
\nu_i(x) := \frac{\varrho^R(x)}{\varrho^1(x)}.
\]

Using the truncation, we have that

\[
\mathcal{E}_p^p [\varrho; l] = \left\| \left( \mathbb{E} - \mathbb{E}_\pi \right) \left[ \varrho(X^\pi_{t_{i+1}}, X^\pi_{t_i}) \Delta W_{t_i+1}^l \right] \right\|_p^p
\]

\[
\leq \mathbb{E} \left[ |\nu_i(X^\pi_{t_i}) - r_i(X^\pi_{t_i})|^p \wedge \gamma(X^\pi_{t_i})^p \right]
\]

\[
\leq \mathbb{E} \left[ |\nu_i(X^\pi_{t_i}) - r_i(X^\pi_{t_i})| \gamma(X^\pi_{t_i})^{p-1} \right],
\]
Using the representation (4.3) we get

\[ E \left[ |\nu_t(X_i^n) - r_i(X_i^n)| \gamma(X_i^n)^{p-1} \right] = E \left[ \left| \frac{q^R(X_i^n) - q^R(X_{i-1}^n)}{q^l(X_i^n)} \right| \gamma(X_i^n)^{p-1} \right] \]

\[ = \int_{\mathbb{R}^d} \left| E \left\{ H_x(X_i^n) \varrho(X_{i+1}^n)(S^{h_i} - S^{h_{i-1}})[\Delta W_{i+1}^l \phi^{*}(X_i^n - x)] \right\} \right| \gamma(x)^{p-1} dx. \]

Using the representation (4.3) we get

\[ E \left[ |\nu_t(X_i^n) - r_i(X_i^n)| \gamma(X_i^n)^{p-1} \right] = \int_{\mathbb{R}^d} \left| E \left\{ H_x(X_i^n) \varrho(X_{i+1}^n) \sum_{J \in \cup_{k \leq d-1} \mathcal{J}_k} (-1)^{|J|} \partial_J \phi^{*}(X_i^n - x) \left( S^{h_i}_{j_e} - S^{h_{i-1}}_{j_e} \right)[\Delta W_{i+1}^l] \right\} \right| \gamma(x)^{p-1} dx. \]

Arguing as in (5.2), with the help of Lemma 3.2.1 of [16], we have

\[ E \left[ \left( S^{h_i}_{j_e} - S^{h_{i-1}}_{j_e} \right)[\Delta W_{i+1}^l] | F_{t_i} \right] = 0. \]

So we may condition in (5.8) the integrand with respect to $F_{t_i}$ to get

\[ E \left[ |\nu_t(X_i^n) - r_i(X_i^n)| \gamma(X_i^n)^{p-1} \right] = \int_{\mathbb{R}^d} \left| E \left\{ H_x(X_i^n) \left( \varrho(X_{i+1}^n) - \varrho(X_i^n) \right) \right\} \times \right. \]

\[ \left. \sum_{J \in \cup_{k \leq d-1} \mathcal{J}_k} (-1)^{|J|} \partial_J \phi^{*}(X_i^n - x) \left( S^{h_i}_{j_e} - S^{h_{i-1}}_{j_e} \right)[\Delta W_{i+1}^l] \right\} \gamma(x)^{p-1} dx \]

Our next step is to perform an $\omega$-by- $\omega$ change of variables for the Riemann integral by setting $u = (X_i^n - x)/|\pi|^\alpha$. Observe that the definition of the localizing function $\phi^{*}$ yields $\partial_J \phi^{*}(x) = |\pi|^{-\alpha|J|} \partial_J \phi(\pi x/|\pi|)$. Hence we have that

\[ E \left[ |\nu_t(X_i^n) - r_i(X_i^n)| \gamma(X_i^n)^{p-1} \right] \leq \sum_{J \in \cup_{k \leq d-1} \mathcal{J}_k} \int_{\mathbb{R}^d} \left| E \left\{ \left( \varrho(X_{i+1}^n) - \varrho(X_i^n) \right)(S^{h_i}_{j_e} - S^{h_{i-1}}_{j_e})[\Delta W_{i+1}^l] \gamma(X_i^n - |\pi|^\alpha u)^{p-1} \right\} \right| \partial_J \phi(u) |du \]
We now apply a Hölder inequality and using the assertion of Lemma 5.2, the Lipschitz assumption on \( \varrho \) and the obvious estimate \( \| X_{\pi t_i}^\pi - X_{\pi t_i}^\pi \|_k \leq C|\pi|^{1/2} \) for any \( k \geq 1 \), we have

\[
\mathbb{E}[|\nu_i(X_{\pi t_i}^\pi) - r_i(X_{\pi t_i}^\pi)|^\gamma(X_{\pi t_i}^\pi)^{p-1}] \\
\leq C \sum_{J \in \mathcal{U}_{k \leq d-1} \mathcal{J}_k} |\pi|^{\frac{1}{2} + (\alpha - \frac{1}{2})(d-|J|)} \| \varrho(X_{\pi t_i}^\pi) - \varrho(X_{\pi t_i}^\pi) \|_p \| \gamma(X_{\pi t_i}^\pi - |\pi|^{\alpha} u)^{p-1} \|_{q_2} |\partial_J \phi(u)| du \\
\leq C \sum_{J \in \mathcal{U}_{k \leq d-1} \mathcal{J}_k} |\pi|^{1 + (\alpha - \frac{1}{2})(d-|J|)} \int_{\mathbb{R}_+^d} \| \gamma(X_{\pi t_i}^\pi - |\pi|^{\alpha} u)^{p-1} \|_{q_2} |\partial_J \phi(u)| du
\]

(5.9)

for some \( q_2 > 1 \) and \( C \), a constant independent of the partition. Now, since the function \( \gamma(\cdot) \) is of polynomial growth we have, with \( r \) an integer, that

\[
\int_{\mathbb{R}_+^d} \mathbb{E}[\gamma(X_{\pi t_i}^\pi - |\pi|^{\alpha} u)^{q_2(p-1)}]^{1/q_2} |\partial_J \phi(u)| du
\]

\[
\leq C \int_{\mathbb{R}_+^d} \mathbb{E}\left[ \left( 1 + \sum_{k=0}^r \binom{r}{k} |X_{\pi t_i}^\pi|^{k} |\pi|^{\alpha} u|^{r-k} \right)^{q_2(p-1)} \right]^{1/q_2} |\partial_J \phi(u)| du,
\]

and since \( |\partial_J \phi(u)| \) integrates against polynomials, by the assumptions on the localizing function, we have the first result. The second part of the theorem follows from (5.9) without using the last step estimate.

\[ \square \]

As already mentioned, here we need to consider an \( \alpha > 1/2 \) and this will have a negative impact, relative to the results of the previous section, on the integrated variance that controls the error of the simulation-based estimation of the family \( \{\hat{R}_i\}_{i=0}^{n-1} \). In particular using the assumptions and the notation of Lemma 4.4, we may show with identical arguments that

\[
\limsup_{|\pi| \to 0} |\pi|^{\alpha d/2} \max_{1 \leq \ell \leq n} \int_{\mathbb{R}^d} \mu(x) \| Q^R[\hat{R}_i, \phi^\pi](x) \|_2 dx < \infty, \quad (5.10)
\]

where \( \phi^\pi(x) := \phi\left( \frac{x}{|\pi|^{\alpha}} \right) \) and \( \mu(\cdot) \) is a function of polynomial growth.

In our main result we will make use of the following version of Gronwall’s whose proof follows from a straightforward induction. lemma.
Lemma 5.5. Let \( a_i, b_i, i = 0, 1, \ldots, n \) be two sequences of positive numbers, such that
\[
a_i \leq b_i + \epsilon \sum_{j=i+1}^{n} a_j, \quad i = 0, \ldots, n-1,
\]
with \( \epsilon > 0 \). Then we have that
\[
\sum_{i=0}^{n} a_i \leq \sum_{i=0}^{n} (1 + \epsilon)^i b_i + (1 + \epsilon)^n a_n.
\]

Theorem 5.6. Let \( \phi, \phi^\pi \) be as in Theorem 5.4 and consider the three families of operators
\[
\{ R_i \}_{i=0}^{n-1}, \{ \hat{R}_i \}_{i=0}^{n-1}, \{ \hat{\hat{R}}_i \}_{i=0}^{n-1}
\]
defined with the function \( \phi^\pi \), on the partition \( \pi \). As before, we have that \( Y^\pi_{t_i} = R_i \ldots R_n \Phi(X^\pi_{t_i}(0, x)), \forall i = 0, \ldots, n-1 \) and we set
\[
\hat{Y}^\pi_{t_i} := \hat{R}_i \ldots \hat{R}_n \Phi(X^\pi_{t_i}(0, x)).
\]

Then, for any \( p > 1 \) and \( \alpha \geq (3p - 1)/2 \), we have
\[
\max_{0 \leq i \leq n-1} \| Y^\pi_{t_i} - \hat{Y}^\pi_{t_i} \|_p \leq C \left( \frac{|\pi|^{1/2(1+\alpha)-1}}{|\pi|^{1+\alpha d/2p} N^{1/2p}} + 1 \right)
\]
for a constant \( C \) independent of the partition.

Remark 5.7. The result above reports on the rate of convergence for the global error. The restriction on \( \alpha \) appears because the use of Lemma 5.5, at the end of our proof calls for an \( \epsilon = O(1/n) \). Observe that with an equidistant partition, say \( |\pi| = 1/n \), this value of \( \alpha \) gives us an error accuracy of order \( 1/\sqrt{n} \), as long as we make use of \( N = n^{3p+ (3p-1)d} \) copies of the Euler scheme at every point of the partition.

Proof of Theorem 5.6: We will estimate \( |Y^\pi_{t_0} - \hat{Y}^\pi_{t_0}| \) and remark that the estimate for \( i > 0 \) follows identical arguments. For ease of presentation we make the conventions that \( R_{-1} = R_{i,j} = 1 \), when \( i > j \), where \( 1 \) is the identity mapping. Once again, we use a Trotter
product expansion for the error:
\[
|Y_{t_0}^\pi - \hat{Y}_{t_0}^\pi| = |R_0 \cdots R_{n-1} \Phi(x) - \hat{R}_0 \cdots \hat{R}_{n-1} \Phi(x)|
\leq \sum_{i=0}^{n-1} |R_{0,i} \hat{R}_{i+1,n-1} \Phi(x) - R_{0,i-1} \hat{R}_{i,n-1} \Phi(x)|
\leq \sum_{i=0}^{n-1} \left\| (R_i - \hat{R}_i) \hat{R}_{i+1,n-1} \Phi(X^\pi_{t_i}) \right\|_p,
\]
where we have used Lemma 3.4, with \( p > 1 \). For every \( i \) we clearly have
\[
\left\| (R_i - \hat{R}_i) \hat{R}_{i+1,n-1} \Phi(X^\pi_{t_i}) \right\|_p \leq \left\| (R_i - \hat{R}_i) R_{i+1,n-1} \Phi(X^\pi_{t_i}) \right\|_p + \left\| (R_i - \hat{R}_i) (\hat{R}_{i+1,n-1} - R_{i+1,n-1}) \Phi(X^\pi_{t_i}) \right\|_p.
\]
To estimate the second term above, we may appeal to Theorem 4.3. The proof of the latter comes through in this case to give us,
\[
\left\| (R_i - \hat{R}_i) \hat{R}_{i+1,n-1} \Phi(X^\pi_{t_i}) \right\|_p \leq \left( \frac{2}{N^{1/2}} \right) \int_{\mathbb{R}^d} \gamma(x) \left[ \|Q^R[\bar{h}_i, \phi](x)\|_p + (|r(x)| + \gamma(x)) \|Q^1[\bar{h}_i, \psi](x)\|_p \right] dx \right)^{1/p}.
\]
Plugging the estimate (5.10) in the above we get
\[
\left\| (R_i - \hat{R}_i) \hat{R}_{i+1,n-1} \Phi(X^\pi_{t_i}) \right\|_p \leq \frac{C}{|\pi| \alpha d/2p N^{1/2}.}
\]
As for the first term in (5.11) we have
\[
\left\| (R_i - \hat{R}_i) \hat{R}_{i+1,n-1} \Phi(X^\pi_{t_i}) \right\|_p \leq \left\| (R_i - \hat{R}_i) R_{i+1,n-1} \Phi(X^\pi_{t_i}) \right\|_p + \left\| (R_i - \hat{R}_i) (\hat{R}_{i+1,n-1} - R_{i+1,n-1}) \Phi(X^\pi_{t_i}) \right\|_p.
\]
We know from Remark 3.3 that the functions \( R_{i+1,n-1} \Phi \) are Lipschitz uniformly in \( i \). Hence we apply the first assertion of Theorem 5.4 to the first term above, while the second assertion is applied to the other one, to obtain
\[
\left\| (R_i - \hat{R}_i) \hat{R}_{i+1,n-1} \Phi(X^\pi_{t_i}) \right\|_p \leq C \left( \left( |\pi| \frac{a+1/2}{\mathcal{R}} + |\pi|^{1/2} \right) \left\| (\hat{R}_{i+1,n-1} - R_{i+1,n-1}) \Phi(X^\pi_{t_i+1}) \right\|_p \right)^{1/p}.
\]
(5.13)
We would now like to eliminate the $1/p$-root above. To this end, observe that for any $a > 0$ and $p > 1$, it holds that $a^{1/p} \leq a + 1$. Hence

$$
\| (\hat{R}_{i+1,n-1} - R_{i+1,n-1}) \Phi(X_{i+1}^\pi) \|_{L^p}^{1/p} \leq |\pi|^{1/p - 1/2} \| (\hat{R}_{i+1,n-1} - R_{i+1,n-1}) \Phi(X_{i+1}^\pi) \|_p + |\pi|^{1/2p}.
$$

(5.14)

Putting together (5.12), (5.13), (5.14) we have that

$$
\| (R_i - \hat{R}_i) \hat{R}_{i+1,n-1} \Phi(X_i^\pi) \|_p \leq C \left( |\pi|^{\alpha/2} + \frac{1}{|\pi|^{\alpha/2p} N^{1/2p}} + |\pi|^{\alpha/2} \sum_{j=i+1}^{n-1} \| (\hat{\hat{R}}_{j+1,n-1} - R_{j+1,n-1}) \Phi(X_{j}^\pi) \|_p \right),
$$

which provides, by performing a Trotter product expansion to the last term,

$$
\| (R_i - \hat{R}_i) \hat{R}_{i+1,n-1} \Phi(X_i^\pi) \|_p \leq C \left( |\pi|^{\alpha/2} + \frac{1}{|\pi|^{\alpha/2p} N^{1/2p}} + |\pi|^{\alpha/2} \sum_{j=i+1}^{n-1} \| (\hat{R}_j - R_j) \hat{R}_{j+1,n-1} \Phi(X_{j}^\pi) \|_p \right),
$$

and the proof is completed by appealing to Lemma 5.5.

\[ \square \]

**Appendix**

**Proof of Theorem 4.3.** Using the bounds on $r_i$ we have

$$
E^{0,i+1} \left[ (r_i(X_i^\pi) - \hat{r}_i(X_i^\pi))^p \right] 
\leq E^{0,i+1} \left[ |\hat{r}_i(X_i^\pi) - r_i(X_i^\pi)|^p \wedge \gamma(X_i^\pi)^p \right] 
= E^{0,i+1} \left[ \left| \frac{\epsilon^F(X_i^\pi) - r_i(X_i^\pi)}{\hat{Q}^1(X_i^\pi)} \right|^p \wedge \gamma(X_i^\pi) \right],
$$

(15)

where

$$
\epsilon^F(x) := \hat{Q}^F(x) - q^F_i(x), \text{ and } \epsilon^1(x) := \hat{Q}^1(x) - q^1_i(x).
$$

We will be using later the fact that

$$
E^{i+1} |\epsilon^F(x)| \leq \|\epsilon^F(x)\|_{L^2(P^{i+1})} \leq \frac{1}{N^{1/2}} V_0^0[Q^F](x)^{1/2},
$$

29
where $V^0[Q^F]$ is the variance of the random variable $Q^F[h_i; \phi]$ with respect to the measure $\mathbb{P}^0$. Notice that the last inequality holds in the marginal measure $\mathbb{P}^{i+1}$. A similar result is true for $\varepsilon^1$.

Next, for $x \in \mathbb{R}^d$ let us consider the event

$$\mathcal{M}(x) = \left\{ \omega : |\hat{Q}^1(x, \omega) - q^1_i(x)| \leq \frac{1}{2} q^1_i(x) \right\}.$$ 

Using this set we may split the expectation above

$$\mathbb{E}^{0, i+1}\left[ \frac{|\varepsilon^F(X^\pi_{t_i}) - r_i(X^\pi_{t_i})\varepsilon^1(X^\pi_{t_i})|^{p} \wedge \gamma(X^\pi_{t_i})^p}{q^1_i(X^\pi_{t_i})} \right] \leq \mathbb{E}^{0, i+1}\left[ 2 \left| \frac{\varepsilon^F(X^\pi_{t_i}) - r_i(X^\pi_{t_i})\varepsilon^1(X^\pi_{t_i})}{q^1_i(X^\pi_{t_i})} \right|^p \gamma(X^\pi_{t_i})1_{\mathcal{M}(X^\pi_{t_i})} \right] + \mathbb{E}^{0, i+1}\left[ \gamma(X^\pi_{t_i})^p \ 1_{\mathcal{M}^c(X^\pi_{t_i})} \right],$$

where we have used the inequality $a^p \wedge b^p \leq ab^{p-1}$. For the first term on the right hand side we compute

$$2 \int_{\mathbb{R}^d} \mathbb{E}^{i+1}\left[ |\varepsilon^F(x) - r_i(x)\varepsilon^1(x)\gamma(x)| \right] dx \leq 2 \int_{\mathbb{R}^d} \|\varepsilon^F(x)\|_{L^2(\mathbb{P}^{i+1})} + |r_i(x)|\|\varepsilon^1(x)\|_{L^2(\mathbb{P}^{i+1})}\gamma(x)^{p-1} dx$$

$$= \frac{2}{N^{1/2}} \int_{\mathbb{R}^d} (V^0_F(x)^{1/2} + |r_i(x)|V^0_1(x)^{1/2}) \gamma(x)^{p-1} dx$$

$$\leq \frac{2}{N^{1/2}} \int_{\mathbb{R}^d} (\|Q^F\|_{L^2(\mathbb{P}^0)} + |r_i(x)|\|Q^1\|_{L^2(\mathbb{P}^0)}) \gamma(x)^{p-1} dx.$$
As for the second term we estimate it by means of the Chebychev inequality

\[
\mathbb{E}^{0,i+1} \left[ \gamma(X_{t_i}^\pi)^p \ 1_{\mathcal{M}^c(X_{t_i}^\pi)} \right] = \mathbb{E}^0 \mathbb{E}^{i+1} \left[ \gamma(X_{t_i}^\pi)^p \ 1_{\mathcal{M}^c(X_{t_i}^\pi)} \right] \\
= \mathbb{E}^0 \left[ \gamma(X_{t_i}^\pi)^p \mathbb{P}^{i+1} \left[ \mathcal{M}^c(X_{t_i}^\pi) \right] \right] \\
\leq \mathbb{E}^0 \left[ \gamma(X_{t_i}^\pi)^p \frac{1}{q_i^1(X_{t_i}^\pi)} \mathbb{E}^{i+1} \left[ 2|\hat{Q}^1(X_{t_i}^\pi) - q_i^1(X_{t_i}^\pi)| \right] \right] \\
= \mathbb{E}^{i+1} \left[ \gamma(x)^p \int_{\mathbb{R}^d} 2|\hat{Q}^1(x) - q_i^1(x)| \, dx \right] \\
\leq \frac{1}{N^{1/2}} \int_{\mathbb{R}^d} \gamma(x)^p V_{Q^1}(x)^{1/2} \, dx.
\]

We now have an estimate of the error with respect to the measure \(d\mathbb{P}^0 \times d\mathbb{P}^{i+1}\). The claimed result follows from an application of Fubini’s theorem.

\[ \square \]

References


