Optimal Pointwise Approximation of Anticipating SDES

Peter Parczewski

Abstract. We derive the optimal rate of convergence for the mean squared error at the terminal point for anticipating linear stochastic differential equations, where the integral is interpreted in Skorohod sense. Although alternative proof techniques are needed, our results can be seen as generalizations of the corresponding results for Itô SDEs. As a key tool we extend optimal approximation results for vectors of correlated Wiener integrals to general random vectors, which contain the solutions of our Skorohod SDEs.

1. Introduction

We suppose a Brownian motion \((W_t)_{t \in [0,1]}\) on the probability space \((\Omega, \mathcal{F}, P)\), where the \(\sigma\)-field \(\mathcal{F}\) is generated by the Brownian motion and completed by null sets. Under the assumption that the Brownian motion is evaluated at an equidistant time grid, i.e. we have the information \(W_{1/n}, W_{2/n}, \ldots, W_1\), the investigation of optimal approximation with respect to the mean squared error (MSE) for Itô stochastic differential equations (SDEs) is well-studied. For SDEs interpreted in the classical Itô sense there is a rich literature on numerical results, approximation algorithms and error analysis. We mention the monographs [8, 11]. We also refer to the survey [13] and the comprehensive study of the pointwise optimal approximation of Itô SDEs given in [14].

Much less is known about numerical schemes for anticipating stochastic differential equations. Some upper bounds of the MSE are known from the investigation of Euler schemes for a class of Skorohod SDEs, see e.g. [19] and [18]. An Euler scheme for an anticipating Stratonovich SDE is analysed in [1]. A Wong-Zakai result for anticipating Stratonovich SDEs is given in [5] by rough path theory. None of these studies deal with optimal approximation or lower bounds of the MSE. Skorohod SDEs arise in several applications, e.g. the computation of derivative-free option price sensitivities [4, 7].

To the best of our knowledge, this is the first study of optimal approximation for anticipating SDEs, where the integral is interpreted in Skorohod sense.

There are very few existence results on Skorohod SDEs. However, all the difficulties arise already at linear Skorohod SDEs with a nonadapted initial value: Let \(a, \sigma, f \in C^1([0,1]; \mathbb{R})\) and consider the Wiener integral \(I(f) = \int_0^1 f(s) dW_s\) and some sufficiently smooth function \(F : \mathbb{R} \to \mathbb{R}\) (cf. Theorem 1 below). Then the unique solution \((X_t)_{t \in [0,1]}\) of the Skorohod SDE with the nonadapted initial value

\[
dx_t = a(t)X_t dt + \sigma(t)X_t dW_t, \quad X_0 = F(I(f)), \quad t \in [0,1],
\]

exists in \(L^2(\Omega \times [0,1])\) (see e.g. [3]). The solution of (1.1) has a simple representation in terms of a convolution operator, the Wick product, a basic tool in stochastic analysis, see e.g. [3, 9] and Section 2 as

\[
X_t = X_0 \diamond \exp \left( \int_0^t \sigma(s) dW_s + \int_0^t (a(s) - \sigma^2(s)/2) ds \right).
\]

The main difficulty is the handling of the convolution operator. We denote the norm and inner product on \(L^2 := L^2([0,1]; \mathbb{R})\) by \(\| \cdot \|\) and \((\cdot, \cdot)\). The stochastic calculus on \(L^2(\Omega) := L^2(\Omega, \mathcal{F}, P)\)
is based on the Gaussian Hilbert space \( \{ I(f) : f \in L^2 \} \subset L^2(\Omega) \). The Wick exponential, i.e. the stochastic exponential of a Wiener integral \( I(f) \), is defined by
\[
\exp(I(f)) := \exp(I(f) - \|f\|^2/2).
\] (1.3)

Given a random variable \( X \in L^2(\Omega) \), we are interested in the approximation
\[
\hat{X}^n \in L^2(\Omega, \sigma(W_{1/n}, \ldots, W_1), P),
\]
that minimizes the mean squared error (MSE) \( \mathbb{E}[(X - \hat{X}^n)^2] \). This is clearly given by
\[
\hat{X}^n := \mathbb{E}[X|W_{1/n}, W_{2/n}, \ldots, W_1].
\] (1.4)

In the following we mean by
\[
f' \in BV
\]
that \( f \in L^2 \) is differentiable and \( f' : [0, 1] \to \mathbb{R} \) is of bounded variation. As a generalization of Wick exponentials following \( \mathbb{W} \mathbb{A} \), we define the class of Wick-analytic functionals \( \mathbb{W} \mathbb{A} \) as
\[
F(I(f_1), \ldots, I(f_K)) = \left( \sum_{k=0}^{\infty} a_{1,k} I(f_1)^{\otimes k} \right) \cdots \left( \sum_{k=0}^{\infty} a_{K,k} I(f_K)^{\otimes k} \right),
\]
where \( K \in \mathbb{N} \), \( f_1, \ldots, f_K \in L^2 \) and \( \sup\{ \sqrt{k}!|a_{i,k}| : i \leq K, k \geq 1 \} < \infty \). The smoothness of Wick-analytic functionals \( F(\cdot) \) is explained in Section 2.

Our main result on optimal pointwise approximation is:

**Theorem 1.** Suppose \( f', \sigma' \in BV \), \( a : \mathbb{R} \to \mathbb{R} \) is integrable, \( X_0 = F(I(f)) \in \mathbb{W} \mathbb{A} \). Then for the solution of the Skorohod SDE \( (\mathbb{W} \mathbb{A}) \) we have
\[
\lim_{n \to \infty} n^2 \mathbb{E}[(X_1 - \hat{X}_1^n)^2] = e^{2 \int_0^1 a(s)ds} 12 \int_0^1 \mathbb{E} \left[ \left( (f'(s)F'(I(f)) + \sigma'(s)F(I(f))) \circ e^{I(\sigma)} \right)^2 \right] ds.
\]

A direct application of Theorem 1 on a Wick exponential type initial value gives:

**Example 2.** Suppose \( f', \sigma' \in BV \), \( a \) is integrable and the nonadapted initial value \( X_0 = e^{\sigma'(I(f))} \). The solution of the linear Skorohod SDE is then given by
\[
X_t = e^{\sigma(I(f))} \circ e^{\int_0^t \sigma(s)ds} \circ W_t \circ e^{\int_0^t a(s)ds} = e^{\sigma(I(f + 1_{[0,t]} \sigma))} \circ e^{\int_0^t a(s)ds},
\]
and the terminal value \( X_1 \) satisfies the asymptotic optimal approximation
\[
\lim_{n \to \infty} n^2 \mathbb{E}[(X_1 - \hat{X}_1^n)^2] = e^{2 \int_0^1 a(s)ds + \|f + \sigma\|^2/12}.
\]

However, here \( X_1 \) coincides with the solution of a linear Itô SDE and the constant above is contained in the optimal approximation results in \( \mathbb{I} \). Considering some \( X_0 = e^{\sigma(I(f))} \) with \( f \in C^1([0, 2]; \mathbb{R}), f \geq 1 \) and \( \text{supp}(f) = [0, 2] \) (in that case a larger \( L^2(\Omega) \) as well), this is not covered by \( \mathbb{I} \) anymore, as \( X_1 \) is now nonadapted. The extension of Theorem 1 to such extended time horizons and initial values is straightforward.

In contrast to Itô SDEs, concerning Skorohod integrals and anticipating SDEs, we have no martingale or Markov tools, in particular we have to handle the lack of Itô isometry and well-known bounds as martingale inequalities (e.g. the BDG inequality). Therefore we make use of subtle computations of the Wiener chaos expansion of all objects involved. As already observed on the optimal approximation of Skorohod integrals \( \mathbb{I} \), this necessary alternative approach leads surprisingly to natural generalizations of the corresponding results for Itô SDEs. Notice again that Theorem 1 extends the corresponding results for linear Itô SDEs in \( \mathbb{I} \), Theorem 1. The main tools for our considerations are optimal approximation results for functionals of Wiener integrals. This leads to nice compatibility relations for optimal approximations of all random elements (Section 3). Due to these general results, Theorem 1 extends easily to more general nonadapted initial values (Theorem 1).

The generalization to an asymptotically optimal approximation scheme for \( \mathbb{I} \) is part of subsequent work.
2. Skorohod SDEs and Wick-analytic functional

For a possibly nonadapted process \((u_s)_{s\in[0,1]}\), the Skorohod integral \(\int_0^1 u_s dW_s\) can be defined as a natural extension of the Itô integral, see e.g. [6][9][17]. We make use of a definition via Wick exponentials (1.3), which have many useful properties. In particular, \(\{\exp^a(I(f)) : f \in L^2\}\) is a total set in \(L^p(\Omega, \mathcal{F}, P)\), \(p > 0\) (see e.g. [10] Cor. 3.40) and allows the characterization of random variables, the S-transform definition of the Skorohod integral (cf. e.g. [10] Sec. 16.4):

**Definition 3.** Suppose \(u = (u_s)_{s\in[0,1]}\) is a (possibly nonadapted) square integrable process on \((\Omega, \mathcal{F}, P)\) and \(Y \in L^2(\Omega, \mathcal{F}, P)\) such that

\[
\forall g \in L^2 : \mathbb{E}[Y e^{\circ I(g)}] = \int_0^1 \mathbb{E}[u_s e^{\circ I(g)}] g(s) ds,
\]

then \(\int_0^1 u_s dW_s = Y\) defines the Skorohod integral of \(u\) with respect to \((W_t)_{t\in[0,1]}\).

For more information on the Skorohod integral we refer to [10][12][16]. The definition above closely relates to a convolution imitating the product of uncorrelated random variables as \(\mathbb{E}[X \circ Y] = \mathbb{E}[X] \mathbb{E}[Y]\), which is implicitly contained in the Skorohod integral and a fundamental tool in stochastic analysis. Due to the injectivity of \(X \mapsto \mathbb{E}[X e^{\circ I(g)}]\) the Wick product can be introduced via

\[
\forall g \in L^2 : \mathbb{E}[(X \circ Y)e^{\circ I(g)}] = \mathbb{E}[X e^{\circ I(g)}] \mathbb{E}[Y e^{\circ I(g)}]
\]

on a dense subset in \(L^2(\Omega) \times L^2(\Omega)\) (see e.g. [10] Chap. 16 for more details). For example, it is \(e^{\circ I(g)} \circ e^{\circ I(h)} = e^{\circ I(g+h)}\) for all \(f, g \in L^2\). In particular, for a Wiener integral \(I(f)\), the Hermite polynomials play the role of monomials in standard calculus as \((I(f))^k = h_k^{\|f\|^2}(I(f))\) and the notation Wick exponential is well justified by \(\exp^a(I(f)) = \sum_{k=0}^\infty \frac{1}{k!} I(f)^{\circ k}\). For more details on Wick exponentials we refer to [9][10][12]. We note that the derivative rule for Hermite polynomials as polynomials \(h_k^{\|f\|^2}(x) = p(x, \|f\|^2)\) gives for the ordinary derivative

\[
\frac{\partial}{\partial x} h_k^{\|f\|^2}(I(f)) = kh_k^{k-1}(I(f)).
\]

An example in \(\mathcal{W}A\), for the Wiener integral \(\int_0^1 W_s ds = \int_0^1 (1-s) dW_s\), is

\[
\sin \left( \int_0^1 W_s ds \right) = \sin \left( \int_0^1 W_s ds \right) e^{\int_0^1 (1-s)^2 ds/2} = \sum_{k=1}^\infty (-1)^{k-1} e^{k-1/6} \frac{(2k-1)!}{(2k-1)!} \left( \int_0^1 W_s ds \right)^{\circ (2k-1)},
\]

which cannot be simulated exactly.

We denote the linear span by \(\text{lin}(\mathcal{W}A)\). Notice that \(\text{lin}(\mathcal{W}A) \subset L^2(\Omega)\) (see e.g. [15] Proposition 9).

**Remark 4.** Let a Wick-analytic functional \(F(I(f)) = \sum_{k=0}^\infty a_k I(f)^{\circ k}\). Thanks to the derivative rule for Hermite polynomials, it is \(F'(I(f)) = \sum_{k\geq 0} (k+1) a_k+1 I(f)^{\circ k}\). Due to \((k+1)/k \leq 2\), we have \(\sup\{ \sqrt{(k+1)!|a_k+1|} : k \geq 1 \} \leq \left( \sup\{ \sqrt{k!|a_k|} : k \geq 1 \} \right)^2 < \infty\). Therefore it clearly is \(F'(I(f)) \in \mathcal{W}A\).

An iteration of the conclusion in Remark 1 (cf. [15] Proposition 10) yields:

**Proposition 5.** All derivatives of elements in \(\text{lin}(\mathcal{W}A)\) as in Remark 4 are in \(\text{lin}(\mathcal{W}A)\) as well.

One can identify \(\text{lin}(\mathcal{W}A)\) as a class of smooth random variables in Malliavin calculus (cf. [16] p. 25]). However, for optimal approximation the Wick-analytic representation is more appropriate, see Section 3.
3. Approximation and Wiener Chaos

In this section we present general results on optimal approximation and on simple implementable approximations for the class $\text{lin}(W\mathcal{A})$. However, besides these essential tools for our main result (Theorem 1), these optimality results and optimality constants are interesting for its own. At the end we prove the main result Theorem 1 and give a further generalization.

The Wiener chaos expansion in terms of Wick analytic functionals has the advantage that the optimal approximation carries over to functionals in terms of Wick products. In fact, we have (see [10 Corollary 9.4] or [6 Lemma 6.20]):

**Proposition 6.** For $X, Y, X \circ Y \in L^2(\Omega)$ and the sub-$\sigma$-field $\mathcal{G} \subseteq \mathcal{F}$:

$$\mathbb{E}[X \circ Y | \mathcal{G}] = \mathbb{E}[X | \mathcal{G}] \circ \mathbb{E}[Y | \mathcal{G}].$$

This immediately implies that the computations on optimal approximation can be extended to Wick-analytic functionals on the underlying Wiener integrals:

**Corollary 7.** For $F = F(I(f_1), I(f_2), \ldots, I(f_K)) \in \text{lin}(W\mathcal{A})$ and the optimal approximation $\hat{F}^n$, we have

$$\hat{F}^n = F \left( \hat{I}(f_1)^n, \hat{I}(f_2)^n, \ldots, \hat{I}(f_K)^n \right).$$

However, we are interested in the implications for the MSE and the constants involved. This is done in three subsections: Finite chaos random variables, Infinite chaos random variables and the conclusion to the proof of our main result.

3.1. Finite chaos. We clearly have

$$\hat{W}^n_t = \mathbb{E}[W_t|W_{1/n}, W_{2/n}, \ldots, W_1] = W_t^{\text{lin}},$$

where

$$W_t^{\text{lin}} := W_{i/n} + n(t-i/n)(W_{(i+1)/n} - W_{i/n}), \quad t \in [i/n, (i+1)/n),$$

is the linear interpolation of $W$ with respect to the equidistant time grid. All further computations will be based on the following convergences:

**Proposition 8.** Suppose $f, g \in \text{BV}$. Then

$$\lim_{n \to \infty} n^2 \mathbb{E} \left[ (I(f) - \hat{I}(f)^n)(I(g) - \hat{I}(g)^n) \right] = \frac{1}{12} \langle f', g' \rangle.$$

**Proof.** Via integration by parts $I(f) = f(1)W_t - \int_0^1 f'(s)W_s \, ds$. Hence, Fubini’s theorem yields

$$I(f) - \hat{I}(f)^n = -\int_0^1 f'(s) \left( W_s - W_s^{\text{lin}} \right) \, ds = -\sum_{i=0}^{n-1} \int_{i/n}^{(i+1)/n} f'(s) \left( W_s - W_s^{\text{lin}} \right) \, ds. \quad (3.1)$$

We recall the well-known covariances of these Brownian bridges:

$$\int_{i/n}^{(i+1)/n} \int_{j/n}^{(j+1)/n} \mathbb{E}[W_s - W_s^{\text{lin}}][W_t - W_t^{\text{lin}}] \, ds \, dt = 1_{i=j} \frac{1}{12n^3}.$$

Hence, for $g, h \in C([0,1]; \mathbb{R})$, the mean value theorem then gives

$$\int_{i/n}^{(i+1)/n} \int_{j/n}^{(j+1)/n} g(s)h(t)\mathbb{E}[W_s - W_s^{\text{lin}}][W_t - W_t^{\text{lin}}] \, ds \, dt = g(\xi_i)h(\xi_i)1_{i=j} \frac{1}{12n^3}$$

for appropriate $\xi_i, \xi_i' \in [i/n, (i+1)/n)$. Thus, (3.1), gives the Riemann sum

$$\mathbb{E} \left[ (I(f) - \hat{I}(f)^n)(I(g) - \hat{I}(g)^n) \right] = \frac{1}{12n^2} \left( \frac{1}{n} \sum_{i=0}^{n-1} f'(\xi_i)g'(\xi_i') \right). \quad (3.2)$$
We denote $h_k := h(k/n)$ for $h \in C([0,1];\mathbb{R})$ and the total variation

$$T(h) := \sup \left\{ \sum_{i=0}^{m-1} |h(t_{i+1}) - h(t_i)| : m \in \mathbb{N}, 0 = t_0 < \ldots < t_m = 1 \right\}.$$  

Hence, via the mean value theorem for some $n \int_{i/n}^{(i+1)/n} h(s) \, ds = h(s_i), s_i \in [i/n, (i+1)/n]$, it is

$$\left| \int_0^1 h(s) \, ds - \frac{1}{n} \sum_{i=0}^{n-1} h_i \right| \leq \frac{1}{n} \sum_{i=0}^{n-1} |h(s_i) - h(i/n)| \leq \frac{T(h)}{n}. \quad \text{(3.3)}$$

Moreover, for all $i = 0, \ldots, n - 1$, we notice

$$|f'_i g'_i - f'(\xi_i) g'(\xi_i)| \leq |f'_i - f'(\xi_i)| \|g'||\infty + |g'_i - g'(\xi_i)| \|f'||\infty,$$

and therefore, via triangle inequality,

$$\frac{1}{12} \left| \frac{1}{n} \sum_{i=0}^{n-1} f'_i g'_i - \left( \frac{1}{n} \sum_{i=0}^{n-1} f'(\xi_i) g'(\xi_i) \right) \right| \leq \frac{1}{12n} \left( T(f') \|g'||\infty + T(g') \|f'||\infty \right). \quad \text{(3.4)}$$

Thanks to (3.2) - (3.1), we conclude

$$\lim_{n \to \infty} n^2 \mathbb{E} \left[ (I(f) - \hat{I}(f)^n)(I(g) - \hat{I}(g)^n) \right] = \frac{1}{12} \langle f', g' \rangle. \quad \text{(3.5)}$$

\begin{remark}
In the simplest case, both converge towards $\frac{1}{12} \|f'||\infty$ as $n$ tends to infinity:

$$n^2 \mathbb{E} \left[ (I(f) - \hat{I}(f)^n)(I(f) - I^n(f)) \right], \quad n^2 \mathbb{E} \left[ (I(f) - \hat{I}(f)^n) I(f) \right].$$

As a direct consequence of (3.3), (3.4) and $T(f) g' \leq T(f') \|g'||\infty + T(g') \|f'||\infty$, we observe the error expansion

$$\left| n^2 \mathbb{E} \left[ (I(f) - \hat{I}(f)^n)(I(g) - \hat{I}(g)^n) \right] - \frac{1}{12} \langle f', g' \rangle \right| \leq \frac{1}{6n} \left( T(f') \|g'||\infty + T(g') \|f'||\infty \right),$$

which can be extended to interesting error expansions for all further results.
\end{remark}

The multiple chaos extension of the covariance limits in Proposition 8 is:

\begin{proposition}
Suppose $f', g' \in BV$ and $k \in \mathbb{N}$ is fixed. Then

$$\lim_{n \to \infty} n^2 \mathbb{E} \left[ ((I(f)^{\cdot \cdot \cdot k}) - (\hat{I}(f)^{\cdot \cdot \cdot k}))(I(g)^{\cdot \cdot \cdot k} - (\hat{I}(g)^{\cdot \cdot \cdot k})) \right] = \frac{1}{12} \langle f', g' \rangle k! \langle f, g \rangle^{(k-1)}. \quad \text{(3.6)}$$

Proof. For higher chaos terms we observe the standard expansion

$$I(f)^{\cdot \cdot \cdot k} - (\hat{I}(f)^{\cdot \cdot \cdot k}) = (I(f) - \hat{I}(f)^{\cdot \cdot \cdot}) \circ (I(f)^{\cdot \cdot \cdot k} - (\hat{I}(f)^{\cdot \cdot \cdot k})).$$

We show that the right hand side is close enough to the simplified variable

$$(I(f) - \hat{I}(f)^{\cdot \cdot \cdot}) \circ kI(f)^{\cdot \cdot \cdot (k-1)}.$$ 

Dealing with $L^2$-norms of Gaussian variables, we will frequently make use of Wick’s Theorem,

$$\mathbb{E} \left[ (I(f_1) \circ \cdots \circ I(f_n)) (I(g_1) \circ \cdots \circ I(g_m)) \right] = 1_{\{n=m\}} \sum_{\sigma \in S_n} \prod_{i=1}^n \langle f_i, g_{\sigma(i)} \rangle, \quad \text{(3.7)}$$

for all $n, m \in \mathbb{N}$, $f_1, \ldots, f_n, g_1, \ldots, g_m \in L^2$, where $S_n$ denotes the group of permutations on $\{1, \ldots, n\}$ (see e.g. [10] Theorem 3.9). Via (3.6) and the reformulation for general products

$$\sum_{j=2}^k a^{k-j} (a^{j-1} - b^{j-1}) = (a - b) \sum_{j=2}^k a^{k-j} \sum_{l=1}^{j-1} a^{l-1} b^{j-l-1} = (a - b) \sum_{j=2}^k \sum_{l=1}^{j-1} a^{k-1-l} b^{j-l-1}$$

OPTIMAL POINTWISE APPROXIMATION OF ANTICIPATING SDES 5
\[ I = (a - b) \sum_{l=1}^{k-1} (k - l)a^{k-1-l}b^{l-1}, \]

for the difference, we have
\[
( I(f) - \widehat{I(f)}^n ) \circ kI(f)^{\circ(k-1)} - \left( I(f)^{\circ k} - \left( \widehat{I(f)}^n \right)^{\circ k} \right) 
\]
\[
= \left( I(f) - \widehat{I(f)}^n \right) \circ \sum_{j=2}^{k} I(f)^{\circ k-j} \circ \left( I(f)^{\circ (j-1)} - \left( \widehat{I(f)}^n \right)^{\circ (j-1)} \right) 
\]
\[
= \left( I(f) - \widehat{I(f)}^n \right)^{\circ 2} \circ \sum_{l=1}^{k-1} (k - l)I(f)^{\circ (k-1-l)} \circ \left( \widehat{I(f)}^n \right)^{\circ (l-1)}. \quad (3.8) 
\]

We give a sufficient upper bound on the L^2-norm. By the covariances
\[
\mathbb{E} \left[ (I(f) - \widehat{I(f)}^n)(I(g) - \widehat{I(g)}^n) \right] = \mathbb{E} \left[ I(f) - \widehat{I(f)}^n \right] \mathbb{E} \left[ I(g) \right] = 0, \\
\mathbb{E} \left[ I(f) \widehat{I(f)}^n \right] = \mathbb{E} \left[ I(f)^2 \right] \geq \mathbb{E} \left[ (\widehat{I(f)}^n)^2 \right] 
\]
and Wick’s Theorem \[(6.7), \text{ for all } 0 \leq m, m' < k, \text{ according to the scheme of numbers of factors} \]
\[
2 \times \{ I(f) - \widehat{I(f)}^n \} \times 2 \\
m \times \{ I(f) - \widehat{I(f)}^n \} \times m' \\
k - m \times \{ I(f) - \widehat{I(f)}^n \} \times k - m' 
\]
and the shorthand notation
\[
\epsilon_n := \mathbb{E} \left[ (I(f) - \widehat{I(f)}^n)^2 \right], 
\]
we obtain
\[
\mathbb{E} \left[ \left( (I(f) - \widehat{I(f)}^n)^{\circ2} \circ I(f)^{\circ m} \circ (\widehat{I(f)}^n)^{\circ k-m} \right) \left( (I(f) - \widehat{I(f)}^n)^{\circ2} \circ I(f)^{\circ m'} \circ (\widehat{I(f)}^n)^{\circ k-m'} \right) \right] 
\leq 2 \epsilon_n^2 k! ||f||^{2k} + 4mm' \epsilon_n^3 (k - 1)! ||f||^{2(k-1)} + m(m - 1)m'(m' - 1) \epsilon_n^4 (k - 2)! ||f||^{2(k-2)}. \quad (3.10) 
\]
Thus, via \[(6.8) - (3.10) \text{ and } \sum_{l=1}^{k-1-m} ((k - l) \cdots (k - l - m)) = \frac{1}{2+m} k(k - 1) \cdots (k - l - m) \text{ for } k - 1 - m \geq 0 \text{ (clear by induction)}, \]
we conclude
\[
\mathbb{E} \left[ \left( I(f)^{\circ k} - \left( \widehat{I(f)}^n \right)^{\circ k} - (I(f) - \widehat{I(f)}^n) \circ kI(f)^{\circ (k-1)} \right)^2 \right] 
\]
\[
= \mathbb{E} \left[ \left( \sum_{l=1}^{k-1} (k - l) \left( I(f) - \widehat{I(f)}^n \right)^{\circ2} \circ I(f)^{\circ (k-1-l)} \circ \left( \widehat{I(f)}^n \right)^{\circ (l-1)} \right)^2 \right] 
\]
\[
= \sum_{l,l' = 1}^{k-1} (k - l)(k - l') \mathbb{E} \left[ \left( (I(f) - \widehat{I(f)}^n)^{\circ2} \circ I(f)^{\circ (k-1-l)} \circ \left( \widehat{I(f)}^n \right)^{\circ (l-1)} \right) \times \left( I(f) - \widehat{I(f)}^n \right)^{\circ2} \circ I(f)^{\circ (k-1-l') \circ \left( \widehat{I(f)}^n \right)^{\circ (l'-1)} \right) \right] 
\leq 2 \sum_{l=1}^{k-1} (k - l)^2 \epsilon_n^2 (k - 2)! ||f||^{2(k-2)} + 4 \left( \sum_{l=1}^{k-2} (k - l)(k - l - 1) \right)^2 \epsilon_n^3 (k - 3)! ||f||^{2(k-3)} 
\]
\[
+ \sum_{l=1}^{k-3} (k - l)(k - l - 1)(k - l - 2)^2 \epsilon_n^4 (k - 4)! ||f||^{2(k-4)} 
\]
\[
= \frac{1}{2} (k(k - 1))^2 \epsilon_n^2 \mathbb{E} \left[ (I(f)^{\circ (k-2)})^2 \right] + \frac{4}{9} (k(k - 1)(k - 2))^2 \epsilon_n^3 \mathbb{E} \left[ (I(f)^{\circ (k-3)})^2 \right] 
\]
\[ = (a - b) \sum_{l=1}^{k-1} (k - l)a^{k-1-l}b^{l-1}, \]
Hence, due to (3.13), Wick’s Theorem (3.7), the covariances (3.9) and Proposition 8, we conclude
\[ j \text{ observe for all } e \]
\[ \text{In particular, via (3.11) and the shorthand notation} \]
\[ \text{for all} \]
\[ \text{Infinite chaos.} \]
\[ 3.2. \]
\[ 0 \]
Infinite chaos.
\[ 0 \]
Proof. We firstly present the proof for \( F = G \). The Wiener chaos expansion yields \( F(I(f)) = \sum_{k \geq 0} a_k \cdot I(f)^{\circ k} \) for some unique coefficients \( a_k \in \mathbb{R} \). Due to Proposition 5, \( \frac{\partial}{\partial x} I(f)^{\circ k} = k I(f)^{\circ (k-1)} \), we observe for all \( j \in \mathbb{N} \),
\[ \mathbb{E} \left[ \left( \frac{\partial^j}{\partial x^j} F(I(f)) \right)^2 \right] = \sum_{k \geq j} \frac{a_k^2}{(k-j)!^2} \mathbb{E} \left[ (I(f)^{\circ (k-j)})^2 \right]. \]
Hence, via (3.11) and the shorthand notation \( e_n = \mathbb{E} \left[ (I(f) - \widehat{I}(f))^n \right] \), we conclude
\[ \sum_{k \geq 1} \frac{a_k^2}{k!^2} \mathbb{E} \left[ \left( (I(f)^{\circ k} - (\widehat{I}(f)^{\circ k}) - (I(f) - \widehat{I}(f))^n \right) \circ k I(f)^{\circ (k-1)} \right]^2 \]
\[ \leq \frac{1}{2} e_n^2 \sum_{k \geq 2} \frac{a_k^2}{(k-2)!^2} \mathbb{E} \left[ (I(f)^{\circ (k-2)})^2 \right] + \frac{4}{9} e_n^3 \sum_{k \geq 3} \frac{a_k^2}{(k-3)!^2} \mathbb{E} \left[ (I(f)^{\circ (k-3)})^2 \right] \]
\[ + \frac{1}{16} e_n^4 \sum_{k \geq 4} \frac{a_k^2}{(k-4)!^2} \mathbb{E} \left[ (I(f)^{\circ (k-4)})^2 \right] \]
\[ = \frac{e_n^2}{2} \mathbb{E} \left[ (F''')^2 \right] + \frac{4e_n^3}{9} \mathbb{E} \left[ \left( \frac{\partial^3}{\partial x^3} F \right)^2 \right] + \frac{e_n^4}{16} \mathbb{E} \left[ \left( \frac{\partial^4}{\partial x^3} F \right)^2 \right] \in \mathcal{O}(n^{-4}). \]
Thus, via Corollary \[3.15\] and Proposition \[10\] (cf. \[3.14\]), we obtain

\[
\lim_{n \to \infty} n^2 \mathbb{E} \left[ \left( F - \hat{F}^n \right)^2 \right] = \lim_{n \to \infty} n^2 \sum_{k \geq 1} \frac{a_k^2}{k!^2} \mathbb{E} \left[ (I(f)^{\circ k} - (I(f)^n)^{\circ k})^2 \right]
\]

\[
= \lim_{n \to \infty} n^2 \sum_{k \geq 1} \frac{a_k^2}{k!^2} \mathbb{E} \left[ ((I(f) - \hat{I}(f)^n) \circ kI(f)^{(k-1)})^2 \right]
\]

\[
= \frac{1}{12} \|f'\|^2 \sum_{k \geq 1} \frac{a_k^2}{(k-1)!} \|f'\|^2(k-1) = \frac{1}{12} \|f'\|^2 \mathbb{E} \left[ (F'(I(f)))^2 \right].
\]

The proof for \( F \neq G \) with the Wiener chaos expansion \( G(I(g)) = \sum_{k \geq 0} \frac{b_k}{k!} I(g)^{\circ k} \) proceeds analogously, as via \[3.13\], \[3.15\] and \[3.14\],

\[
\lim_{n \to \infty} n^2 \mathbb{E} \left[ \left( F - \hat{F}^n \right) \left( G - \hat{G}^n \right) \right]
\]

\[
= \lim_{n \to \infty} n^2 \sum_{k \geq 1} \frac{a_kb_k}{k!^2} \mathbb{E} \left[ (I(f)^{\circ k} - (I(f)^n)^{\circ k}) \left( I(g)^{\circ k} - (I(g)^n)^{\circ k} \right) \right]
\]

\[
= \lim_{n \to \infty} n^2 \sum_{k \geq 1} \frac{a_kb_k}{k!^2} \mathbb{E} \left[ ((I(f) - \hat{I}(f)^n) \circ kI(f)^{(k-1)}) \left( (I(g) - \hat{I}(g)^n) \circ kI(g)^{(k-1)} \right) \right]
\]

\[
= \frac{1}{12} \langle f', g' \rangle \sum_{k \geq 1} \frac{a_kb_k}{(k-1)!} (f,g)^{k-1} = \frac{1}{12} \langle f', g' \rangle \mathbb{E} \left[ F'(I(f))G'(I(g)) \right].
\]

Example 12. Suppose \( f' \in BV \). Then we have

\[
\lim_{n \to \infty} n^2 \mathbb{E} \left[ \left( e^{I(f)} - e^{I(f)^n} \right)^2 \right] = \frac{1}{12} \|f'\|^2 \mathbb{E} \left[ e^{2\|f\|} \right].
\]

Remark 13. Looking at \[3.10\], we observe that \( \mathbb{E}[(I(f)^n)^2] \leq \mathbb{E}[(I(f))^2] \) is the only reason for the inequality. Hence, asymptotically, we obtain equalities in \[3.10\], \[3.11\] and \[3.15\]. This yields the following expansion of the optimal approximation error with \( e_n := \mathbb{E} \left[ (I(f) - \hat{I}(f)^n)^2 \right] \),

\[
\mathbb{E} \left[ \left( F(I(f)) - \hat{F}(I(f))^n \right)^2 \right] \sim e_n \mathbb{E} \left[ (F'(I(f)))^2 \right] + \frac{e_n^2}{2} \mathbb{E} \left[ (F''(I(f)))^2 \right]
\]

\[
+ \frac{4e_n^3}{9} \mathbb{E} \left[ (F^{(3)}(I(f)))^2 \right] + \frac{e_n^4}{16} \mathbb{E} \left[ (F^{(4)}(I(f)))^2 \right].
\]

Theorem \[11\] can be extended in various directions to multivariate functionals. For the proofs of our main result we need:

**Theorem 14.** Suppose \( f', g', f_1', g_1', \ldots \in BV \) and define the abbreviations

\[
F = F(I(f_1), \ldots, I(f_m)), \quad G = G(I(g_1), \ldots, I(g_m)), \quad F_{x_k} := \frac{\partial}{\partial x_k} F.
\]

Then

\[
\lim_{n \to \infty} n^2 \mathbb{E} \left[ (F - \hat{F}^n)(G - \hat{G}^n) \right] = \frac{1}{12} \sum_{i,j=1}^m \langle f_i', g_j' \rangle \mathbb{E} \left[ F_{x_i} G_{x_j} \right].
\]

**Proof.** Due to various Wiener chaos expansions the notations for arbitrary functionals become easily elaborately. However, the proof is a straightforward extension of the arguments for Theorem \[11\]. We present the proof for the two-dimensional case for \( G = F = F(I(f), I(g)) \in WA \). All other cases are straightforward generalizations as before. Let the Wiener chaos expansion

\[
F(I(f), I(g)) = \sum_{k,l=0}^{\infty} a_{k,l} f^{\circ k} \circ I(g)^{\circ l}
\]
for some coefficients $a_{k,l} \in \mathbb{R}$. Via Corollary 11 it is

$$
\mathbb{E} \left[ \left( F(I(f), I(g)) - F(I(\hat{f}), I(\hat{g})) \right)^n \right] = 
\sum_{k,l,k',l' \geq 0, k+l=k'+l'=N} a_{k,l} a_{k',l'} \mathbb{E} \left[ \left( I(f)^{ok} \cdot I(g)^{ol} - \left( I(\hat{f})^{n} \right)^{ok} \cdot \left( I(\hat{g})^{n} \right)^{ol} \right) \times \left( I(f)^{ok'} \cdot I(g)^{ol'} - \left( I(\hat{f})^{n} \right)^{ok'} \cdot \left( I(\hat{g})^{n} \right)^{ol'} \right) \right].
$$

As

$$
I(f)^{ok} \cdot I(g)^{ol} - \left( I(\hat{f})^{n} \right)^{ok} \cdot \left( I(\hat{g})^{n} \right)^{ol} = (I(f)^{ok} - (I(\hat{f})^{n})^{ok}) \cdot I(g)^{ol} + (I(\hat{f})^{n})^{ok} \cdot (I(g)^{ol} - (I(\hat{g})^{n})^{ol}),
$$

the right hand side in (3.19) is reduced to covariances of the terms of the type

$$
(I(f)^{ok} - (I(\hat{f})^{n})^{ok}) \cdot I(g)^{ol}, \quad (I(f)^{ok} - (I(\hat{f})^{n})^{ok}) \cdot (I(\hat{g})^{n})^{ol}.
$$

Analogously to the proof of Proposition 11 (see e.g. 3.12–3.13), these terms behave in covariance computations like

$$
(I(f) - I(\hat{f})^n) \cdot (I(f)^{ok} \cdot I(g)^{ol}).
$$

Suppose $f', g', \tilde{f}', \tilde{g}' \in BV$. We recall that $I(f)^{ok} \cdot I(g)^{ol}$ are polynomials $p(x_1, x_2)$ (of $x_1, x_2 = I(f), I(g)$) and therefore differentiable. Then, analogously to 3.14 and due to the derivative rule for Hermite polynomials, for $k + l = k' + l'$ we obtain

$$
\mathbb{E} \left[ \left( (I(f)^{ok} - (I(\hat{f})^{n})^{ok}) \cdot I(g)^{ol} \right) \left( (I(f)^{ok'} - (I(\hat{f})^{n})^{ok'}) \cdot I(\tilde{g})^{ol'} \right) \right] 
\sim \mathbb{E} \left[ \left( (I(f) - I(\hat{f})^n) \cdot kI(f)^{ok-1} \cdot I(\tilde{g})^{ol} \right) \left( \left( (I(f) - I(\hat{f})^n) \cdot kI(\tilde{f})^{ok'-1} \cdot I(\tilde{g})^{ol'} \right) \right) \right] 
\sim \frac{1}{12n^2} \langle f', \tilde{f}' \rangle \mathbb{E} \left[ \left( kI(f)^{ok-1} \cdot I(g)^{ol} \right) \left( kI(\tilde{f})^{ok'-1} \cdot I(\tilde{g})^{ol'} \right) \right] 
= \frac{1}{12n^2} \langle f', \tilde{f}' \rangle \mathbb{E} \left[ \frac{\partial}{\partial x_1} I(f)^{ok} \cdot I(g)^{ol} \left( \frac{\partial}{\partial x_1} I(\tilde{f})^{ok'} \cdot I(\tilde{g})^{ol'} \right) \right].
$$

Thus, as in Proposition 10 for a fixed chaos (of order $N \in \mathbb{N}$), we conclude

$$
\sum_{k,l,k',l' \geq 0, k+l=k'+l'=N} a_{k,l} a_{k',l'} \mathbb{E} \left[ \left( I(f)^{ok} \cdot I(g)^{ol} - \left( I(\hat{f})^{n} \right)^{ok} \cdot \left( I(\hat{g})^{n} \right)^{ol} \right) \times \left( I(f)^{ok'} \cdot I(g)^{ol'} - \left( I(\hat{f})^{n} \right)^{ok'} \cdot \left( I(\hat{g})^{n} \right)^{ol'} \right) \right] 
\sim \frac{1}{12n^2} \|f''\|^2 \sum_{k,l,k',l' \geq 0, k+l=k'+l'=N} a_{k,l} a_{k',l'} \mathbb{E} \left[ \frac{\partial}{\partial x_1} I(f)^{ok} \cdot I(g)^{ol} \left( \frac{\partial}{\partial x_1} I(f)^{ok'} \cdot I(g)^{ol'} \right) \right] 
+ \frac{1}{12n^2} \|g''\|^2 \sum_{k,l,k',l' \geq 0, k+l=k'+l'=N} a_{k,l} a_{k',l'} \mathbb{E} \left[ \frac{\partial}{\partial x_2} I(f)^{ok} \cdot I(g)^{ol} \left( \frac{\partial}{\partial x_2} I(f)^{ok'} \cdot I(g)^{ol'} \right) \right] 
+ \frac{2}{12n^2} \langle f', g' \rangle \sum_{k,l,k',l' \geq 0, k+l=k'+l'=N} a_{k,l} a_{k',l'} \mathbb{E} \left[ \frac{\partial}{\partial x_1} I(f)^{ok} \cdot I(g)^{ol} \left( \frac{\partial}{\partial x_2} I(f)^{ok'} \cdot I(g)^{ol'} \right) \right].
$$

Hence, analogously to Theorem 11 summing up (Fubini’s theorem applies as we have uniform bounds via Proposition 5) leads to the asserted asymptotics

$$
\mathbb{E} \left[ \left( F(I(f), I(g)) - F(I(\hat{f}), I(\hat{g})) \right)^n \right] 
\sim \frac{1}{12n^2} \|f''\|^2 \mathbb{E} \left[ F_{x_1}^2 \right] + \frac{1}{12n^2} \|g''\|^2 \mathbb{E} \left[ F_{x_2}^2 \right] + \frac{2}{12n^2} \langle f', g' \rangle \mathbb{E} \left[ F_{x_1} F_{x_2} \right].
$$

□
Remark 15. In particular, due to

\[ E \left[ (F + G - \hat{F} + \hat{G}^n)^2 \right] = E \left[ (F - \hat{F})^2 + (G - \hat{G}^n)^2 + 2(F - \hat{F})(G - \hat{G}^n) \right], \]

or polarization, we conclude Theorem 1 for \( F, G \in \text{lin}(WA) \) as well.

3.3. Proof of the main result. Our optimal approximation result is now an easy application:

Proof of Theorem 1. Due to \( 1.2 \), Proposition 6 and the deterministic term \( \int_0^1 a(s)ds \), we have

\[ E[(X_1 - \hat{X}_1)^n] = E \left[ (F(I(f)) \circ e^{\sigma I(\sigma)} I^{1/n} a(s)ds - F(I(f))^n \circ e^{\sigma I(\sigma)^n} I^{1/n} a(s)ds)^2 \right] \]

\[ = e^{2n} I^{1/n} a(s)ds E \left[ (F(I(f)) \circ e^{\sigma I(\sigma)} - F(I(f))^n \circ e^{\sigma I(\sigma)^n})^2 \right]. \]

For the random variable

\[ G(I(f), I(\sigma)) := F(I(f)) \circ e^{\sigma I(\sigma)}, \]

by \( F(I(f)) \in WA \) and Proposition 5 these derivatives exist in \( L^2(\Omega) \) (for the function \( G(x_1, x_2) \)):

\[ G_{x_1} \frac{\partial}{\partial x_1} G(I(f), I(\sigma)) = F'(I(f)) \circ e^{\sigma I(\sigma)}, \quad G_{x_2} \frac{\partial}{\partial x_2} G(I(f), I(\sigma)) = F(I(f)) \circ e^{\sigma I(\sigma)}. \]

Thanks to Theorem 14, we therefore conclude

\[ \lim_{n \to \infty} n^2 E \left[ (F(I(f)) \circ e^{\sigma I(\sigma)} - F(I(f))^n \circ e^{\sigma I(\sigma)^n})^2 \right] = \frac{1}{12} \left( \|f\|^2 E[(F'(I(f)) \circ e^{\sigma I(\sigma)})^2] + \|\sigma\|^2 E[(F'(I(f)) \circ e^{\sigma I(\sigma)})^2] + 2 \langle f', \sigma' \rangle E[(F'(I(f)) \circ e^{\sigma I(\sigma)})(F(I(f)) \circ e^{\sigma I(\sigma)})] \right) \]

\[ = \int_0^1 E \left[ \left( (f'(s)F'(I(f))) + \sigma'(s)F(I(f)) \right) \circ e^{\sigma I(\sigma)} \right]^2 ds. \]

This yields the asserted optimal convergence in Theorem 1. \( \square \)

We obtain easily further multivariate generalizations of the nonadapted initial value:

Theorem 16. Suppose \( f_1, \ldots, f_m, \sigma' \in BV, a : \mathbb{R} \to \mathbb{R} \) is integrable and

\[ X_0 := F = F(I(f_1), \ldots, I(f_m)) \in \text{lin}(WA). \]

Then for the solution of the Skorohod SDE \( 1.1 \) we have

\[ \lim_{n \to \infty} n^2 E[(X_1 - \hat{X}_1)^n] = \frac{2e^{1/n} a(s)ds}{12} \int_0^1 E \left[ \left( \sum_{k=1}^m f_k(s) F_{x_k} + \sigma'(s) F \right) \circ e^{\sigma I(\sigma)} \right]^2 ds. \]

Proof. Concerning a multivariate initial value \( F(I(f_1), \ldots, I(f_m)) \) in Theorem 16 we conclude analogously via Theorem 14, Remark 15 and the function

\[ G(I(f_1), \ldots, I(f_m), I(\sigma)) := F(I(f_1), \ldots, I(f_m)) \circ e^{\sigma I(\sigma)} \in \text{lin}(WA). \]

\( \square \)

Remark 17. The investigation of an asymptotically optimal approximation scheme for the linear Skorohod SDEs in Theorems 14 15 will follow in a subsequent work.

Acknowledgement 1. The author thanks Andreas Neuenkirch for many fruitful discussions.
REFERENCES

[1] Ahn, H. and Kohatsu-Higa, A. The Euler scheme for anticipating stochastic differential equations. *Stochastics Stochastics Rep.* **54** (3-4), (1995) 247–269.

[2] Buckdahn, R. Skorohod stochastic differential equations of diffusion type. *Probab. Theory Related Fields* **93** (3), (1992) 297–323.

[3] Buckdahn, R. and Nualart, D. Linear stochastic differential equations and Wick products. *Probab. Theory Related Fields* **99** (4), (1994) 501–526.

[4] Chen, N. and Glasserman, P. Malliavin Greeks without Malliavin calculus. *Stochastic Process. Appl.* **117** (11), 1689-1723, (2007).

[5] Coutin, L. and Friz, P. and Victoir, N. Good rough path sequences and applications to anticipating stochastic calculus. *Ann. Probab.* **35** (3), (2007) 1172–1193.

[6] Di Nunno, G. and Øksendal, B. and Proske, F. *Malliavin calculus for Lévy processes with applications to finance* Universitext. Springer, Berlin, 2009.

[7] Fournié, E. and Lasry, J.-M. and Lebuchoux, J. and Lions, P.-L. and Touzi, N. Applications of Malliavin calculus to Monte Carlo methods in finance. *Finance Stoch.* **3** (4), (1999) 391–412.

[8] Graham, C. and Talay, D. *Stochastic Simulation and Monte Carlo Methods*, Springer, Berlin, 2013.

[9] Holden H. and Øksendal, B. and Ubøe, J. and Zhang, T. *Stochastic Partial Differential Equations. A Modeling, White Noise Functional Approach. Second Edition* Springer, New York, 2010.

[10] Janson, S. *Gaussian Hilbert Spaces.*, Cambridge University Press, Cambridge, 1997.

[11] Kloeden, P. and Platen, E. *Numerical solution of stochastic differential equations*. Applications of Mathematics, 23. Springer-Verlag, Berlin, 1992.

[12] Kuo, H.-H. *White Noise Distribution Theory*. Probability and Stochastics Series. CRC Press, Boca Raton, 1996.

[13] Müller-Gronbach, T. and Ritter, K. Minimal errors for strong and weak approximation of stochastic differential equations. In *Monte Carlo and Quasi-Monte Carlo Methods 2006*, Springer, (2008) 53–82.

[14] Müller-Gronbach, T. Optimal pointwise approximation of SDEs based on Brownian motion at discrete points. *Ann. Appl. Probab.* **14** (4), (2004) 1605–1642.

[15] Neuenkirch, A. and Parczewski, P. Optimal approximation of Skorohod integrals. *J. Theor. Probab.* **31** (1), (2018) 206–231.

[16] Nualart, D. *The Malliavin Calculus and Related Topics*. Second Edition. Probability and its Applications. Springer, New York, 2006.

[17] Pardoux, É. Applications of anticipating stochastic calculus to stochastic differential equations. In *Lecture Notes in Math.* **1444**, Springer, (1990) 63–105.

[18] Shevchenko, G. Euler approximation for anticipating stochastic quasilinear differential equations. *Theory Probab. Math. Statist.* **72**, (2006) 167–175.

[19] Torres, S. and Tudor, C. The Euler scheme for a class of anticipating stochastic differential equations. *Random Oper. Stochastic Equations* **12** (3), (2004) 211–224.

Peter Parczewski: Institute of Mathematics B6, University of Mannheim, D-68131 Mannheim, Germany

Email address: parczewski@math.uni-mannheim.de